

A Consensus Model for Seasonal Hurricane Prediction

THOMAS H. JAGGER AND JAMES B. ELSNER

Department of Geography, The Florida State University, Tallahassee, Florida

(Manuscript received 18 February 2010, in final form 22 June 2010)

ABSTRACT

The authors apply a procedure called Bayesian model averaging (BMA) for examining the utility of a set of covariates for predicting the distribution of U.S. hurricane counts and demonstrating a consensus model for seasonal prediction. Hurricane counts are derived from near-coastal tropical cyclones over the period 1866–2008. The covariate set consists of the May–October monthly averages of the Atlantic SST, North Atlantic Oscillation (NAO) index, Southern Oscillation index (SOI), and sunspot number (SSN). BMA produces posterior probabilities indicating the likelihood of the model given the set of annual hurricane counts and covariates. The September SSN covariate appears most often in the higher-probability models. The sign of the September SSN parameter is negative indicating that the probability of a U.S. hurricane decreases with more sunspots. A consensus hindcast for the 2007 and 2008 season is made by averaging forecasts from a large subset of models weighted by their corresponding posterior probability. A cross-validation exercise confirms that BMA can provide more accurate forecasts compared to methods that select a single “best” model. More importantly, the BMA procedure incorporates more of the uncertainty associated with making a prediction of this year’s hurricane activity from data.

1. Introduction

Standard practice in modeling climate data is to ignore model uncertainty. Model uncertainty refers to the ambiguity associated with choosing a single model from a suite of similarly good models. The researcher employs a procedure that selects a set of covariates for a given class of models (model selection) and then uses an estimation method to determine parameter values for the model. Subsequently, the model is used to make inferences and predictions as if the selected model had generated the data. Unfortunately, this approach ignores the uncertainty in the model selection procedure resulting in overconfident inferences and predictions, which can lead to unwarranted decisions.

For example, given a pool of covariates for hurricane activity a stepwise regression procedure is frequently employed to search through the hundreds of combinations to arrive at a final reduced set of predictors (Klotzbach 2008). This final set is usually subjected to a hold-one-out cross-validation exercise to obtain an estimate of how well the set of covariates will predict future data.

However, as noted in Elsner and Schmertmann (1994), this exercise does not result in a “full” cross validation as the procedure for selecting the reduced set of covariates is not itself cross validated. Cross validation is a procedure for assessing how well an algorithm for choosing a particular model (including the predictor selection phase) will do in forecasting the unknown future (Michaelsen 1987; DelSole and Shukla 2009).

Bayesian model averaging (BMA) is an alternative to selecting a single “best” model. It works by assigning a probability to each model (combination of covariates), then averaging over all models weighted by their probability. That is, the BMA procedure assumes that all models under consideration are capable of explaining the data to some degree or another, so it is better to use them all. In this way, the uncertainty associated with model selection is incorporated into the procedure. Combining statistical climate models to improve various forecast characteristics (Colman and Davey 2003; Zou and Yang 2004) and to use BMA for combining predictive densities (Raftery et al. 2005) are relatively new topics. The purpose of this paper is to introduce the BMA procedure for producing a consensus seasonal hurricane forecast. In doing so, we show how the approach can facilitate our physical interpretation of the modeled relationships.

Corresponding author address: Thomas H. Jagger, Dept. of Geography, The Florida State University, Tallahassee, FL 32306.
E-mail: tjagger@blarg.net

The paper is outlined as follows. In section 2, the underlying statistical theory of BMA is introduced in the context of seasonal hurricane counts along with some related ideas. In section 3, the data describing seasonal hurricane activity along the U.S. coast and the corresponding covariates are described. In section 4, the results from the BMA procedure are discussed in the context of our physical understanding of seasonal hurricane activity. In section 5, the consensus model arising from the BMA is used to hindcast the 2007 and 2008 hurricane seasons. This section includes a comparison of the BMA with single-model selection procedures using a cross-validation exercise. The paper ends with conclusions and a summary along with some closing remarks.

2. Bayesian model averaging

We use the BMA procedure to develop a statistical consensus model for seasonal hurricane forecasting. We begin with some underlying theory of BMA in the context of seasonal hurricane counts. Let H_i , with $i = 1, \dots, N$, be the observed hurricane counts, one for each observation year. Assume that our model has k covariates, and let \mathbf{X} be the covariate matrix with components $X[i, j]$, with $i = 1, \dots, N$ and $j = 1, \dots, k$, associated with the i th observation of the $(j + 1)$ th covariate and with the intercept term $X[i, 1] = 1$ for all i . Associated with the intercept and k covariates are $k + 1$ parameters β_j , with $j = 1, \dots, k + 1$.

Because we have count data, we assume that the observations have a Poisson distribution, and the logarithm of the mean is regressed onto the covariates—that is,

$$H_i \sim \text{pois}(\lambda_i)$$

$$\log(\lambda_i) = \sum_{j=1}^{k+1} X[i, j]\beta_j.$$

This is a generalized linear model (GLM), and maximum likelihood estimation can be used to estimate the parameters. From these parameter estimations and the values of the corresponding covariates we can estimate λ from the regression equation. The future hurricane count conditional on these covariate values has a Poisson distribution with a mean of λ . Thus, the Poisson distribution provides a probabilistic forecast whose mean can be taken as a point forecast.

A full model uses all k covariates. However, it is usually the case that some of the covariates do not contribute information to the model. We can choose a smaller model by setting some of the k parameters to zero and estimating the rest of the parameters. Thus,

with k covariates there are a total of $m = 2^k$ models. The principal idea in BMA is that none of the m models are discarded, rather a probability is assigned to each and predictions averaged based on this probability. Models with greater probability are assigned proportionally more weight in the averaging.

Consider a simple case in which our observations of Y results from either 1 of 2 regression models $Y_1 = \alpha_1 + \epsilon_1$ (constant mean) and $Y_2 = \alpha_2 + \beta x + \epsilon_2$ (simple regression), where x is a covariate, and ϵ_1 and ϵ_2 are independent and normally distributed with means of zero and variances of σ_1^2 and σ_2^2 , respectively. Suppose we can assign a probability p that the constant mean model generated the observed data and a probability $1 - p$ that the simple regression model instead generated the data. Then under BMA, the posterior predictive expectation (mean) of Y is $\mu = p\mu_1 + (1 - p)\mu_2 = p\alpha_1 + (1 - p)(\alpha_2 + \beta x)$. This represents a consensus model that combines information from both models.

However, the posterior predictive distribution of Y given the data is not necessarily normal, rather it is a mixture of normal distributions with a posterior predicted variance of $p\sigma_1^2 + (1 - p)\sigma_2^2 + p(1 - p)(\alpha_2 + \beta x - \alpha_1)^2$. This variance under BMA is larger than a simple weighted sum of the individual model variances by an amount $p(1 - p)(\alpha_2 + \beta x - \alpha_1)^2$ that represents the uncertainty associated with model choice. Thus, the predictive distribution under BMA is wider than the distribution under any given model.

Over a set of competing models we need a method to assign a probability to each. We start with a collection of models M_i , with $i = 1, \dots, m$, where each model is a unique description of our data. For example, in the example above we need to assign a probability to the constant mean model and a probability for the simple regression model, with the constraint that the probability over all models must sum to one. In this study, a model represents the generalized linear regression structure and a set of covariates with a corresponding set of parameters including an intercept term.

Now for our dataset D and model M_i we can determine $P(D|M_i)$, which is the probability of our data given the model, and we assign to each model a prior probability $P(M_i)$. This is our prior belief on a given model. Under the situation of complete ambiguity we assigned $1/m$ to each model's prior probability. For example, in the above case if we believe both models are equally likely then we assign $P(M_1) = P(M_2) = 0.5$. Using Bayes Law we find the probability of the model given the data $P(M_i|D) = P(D|M_i) \times P(M_i)/P(D)$ because $P(D)$ is fixed for all models we can let $W_i = P(D|M_i) \times P(M_i)$ be the model weights with probabilities $P(M_i|D) = W_i / \sum_{i=1}^m W_i$.

Let the random variable H represent a future count prediction. The posterior distribution of H at h under each model is given by $f(h|D, M_i)$. The marginal posterior probability over all models is given by $f(h|D) = \sum_{i=1}^m f(h|D, M_i)P(M_i|D)$. We can find a single-point estimate of the future count—that is, the posterior mean of H over all models—by taking the expectation of H given the data, $\mathbb{E}(H|D) = \sum_{h=0}^{\infty} hf(h|D)$. Then expand $f(h|D)$ and switch the summation order $\mathbb{E}(H|D) = \sum_{i=1}^m P(M_i|D) \sum_{h=0}^{\infty} hf(h|D, M_i)$, which is $\sum_{i=1}^m P(M_i|D) \mathbb{E}(H|D, M_i)$ as in our example, where $\mathbb{E}(H|D, M_i) = \mu_i$. For a given model, $P(D|M_i)$ is the marginal likelihood over the parameter space. In other words, $P(D|M_i) = \int P(D|M_i, \theta) f(\theta|M_i) d\theta$, where $f(\theta|M_i)$ is the prior distribution of the parameters for model M_i , and $P(D|M_i, \theta)$ is the likelihood of the data given the model $L(\theta; M_i, D)$. In many cases this integral cannot be evaluated analytically or is infinite as when an improper prior is put on θ . Several approximation methods may be used to evaluate this integral as in Hoeting et al. (1999). In our approach, we use the Bayesian information criterion (BIC) approximation, which is based on a Laplace expansion of the integral about the maximum likelihood estimator (MLE) parameter estimates (Madigan and Raftery 1994).

Model selection attempts to find a single best model for the data. Selection criteria may include minimization of the Akaike information criterion (AIC; Akaike 1974), the BIC, or a cross-validation score. In a sense, BMA is the opposite of model selection; it keeps all models but assigns a probability based on how likely it would be for the data to have come from the model. A consensus model, representing a weighted average of all models, is then used to make predictions. If values for the prior parameters come from reasonably well-behaved distributions, then a consensus model from a BMA procedure yields the lowest mean square error (MSE) of any single best model (Raftery and Zheng 2003). BMA is described in more detail in Hoeting et al. (1999), whose work is based on Raftery (1996) for generalized linear models.

BMA provides better coverage probabilities on the predictions than any single model. Consider a dataset split into a training set and a testing set. Using the training set we can create $1 - \alpha$ credible intervals on the predictions. Then using the testing set we can calculate the proportion of observations that lie within the credible intervals (coverage probability). In standard practice with a single best model, the credible intervals are too small resulting in coverage probabilities less than $1 - \alpha$. Since BMA provides a larger variance than any model individually, the coverage probabilities on the predictions are greater or equal to $1 - \alpha$ (Raftery and Zheng 2003). For a comparison of BMA against other model

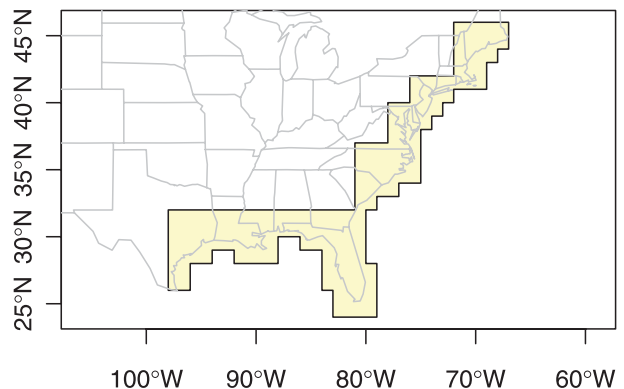


FIG. 1. Study area. The number of hurricanes affecting the U.S. coast is based on counting hurricanes crossing into the near-coastal polygon (soft yellow).

selection procedures see Castle et al. (2009). Here, we demonstrate BMA in the context of seasonal hurricane forecasting. We begin by describing the data.

3. Data

a. Hurricane frequency in the vicinity of the United States

Hurricane wind speed estimates are derived from the Hurricane Database (HURDAT or best track) maintained by the U.S. National Oceanic and Atmospheric Administration (NOAA) National Hurricane Center (NHC). Of interest is the fact that HURDAT is the official NOAA record of hurricane information for the Atlantic Ocean, Gulf of Mexico, and Caribbean Sea, including those storms that have made landfall in the United States. HURDAT consists of 6-h position, central pressure, and maximum sustained wind estimates for tropical cyclones back to 1851. The raw wind speed values in HURDAT are given in 5-kt (2.5 m s^{-1}) increments and knots (kt) are the operational unit used for reporting tropical cyclone intensity to the public in the United States. For cyclones prior to 1931, the 6-h positions and intensities are interpolated from once daily (1200 UTC) estimates. Here, we use the latest version of HURDAT as of November 2009.

A natural spline interpolation is used to obtain positions and wind speeds at 1-h intervals from the 6-h values. Because a complete dataset of all land-falling hurricanes is not available, and to be consistent, we use the near-coast region (see Fig. 1) of Jagger and Elsner (2006) and keep only the hurricane's single highest wind speed in the region. We use the term "intensity" as shorthand for "wind speed." Thus, the hurricane dataset we analyze and model in this study contains $N = 286$ hurricanes over the 143-yr period (1866–2008). A time

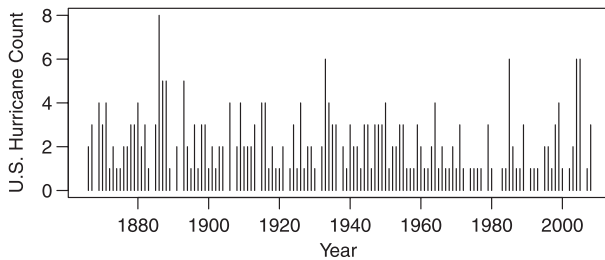


FIG. 2. Annual U.S. hurricane counts. The counts are based on tropical cyclones at hurricane intensity within a near-coastal region of the United States (see Fig. 1) over the period 1866–2008, inclusive.

series of the counts is shown in Fig. 2. The counts range from zero in 18 of the 143 yr to eight in 1886. There is no long-term trend in the annual counts.

b. Environmental variables related to U.S. hurricanes

On the annual time scale, and to a first order, it is known that a high ocean heat content and cold upper-air temperature provide the fuel for hurricanes, a calm atmosphere (low values of wind shear) allows a hurricane to intensify, and the position and strength of the subtropical high pressure region steers a hurricane that does form. Thus, U.S. hurricane activity responds to changes in large-scale climate conditions that affect or index these factors including SST as an indicator of oceanic heat content, sunspot number (SSN) as an indicator of upper-air temperature, El Niño–Southern Oscillation (ENSO) as an indicator of vertical wind shear, and the North Atlantic Oscillation (NAO) as an indicator of steering flow.

ENSO is characterized by basin-scale fluctuations in sea level pressure (SLP) across the equatorial Pacific Ocean. The Southern Oscillation index (SOI) is defined as the normalized sea level pressure difference between Tahiti and Darwin, and values are available back through the middle nineteenth century. The SOI is strongly anticorrelated with equatorial Pacific SSTs so that an El Niño warming event is associated with negative SOI values. The units are standard deviations. ENSO is an indicator of vertical wind shear and subsidence in the Atlantic region where tropical cyclones develop and negative SOI values imply greater shear and subsidence. The monthly SOI values (Ropelewski and Jones 1987) are obtained from the Climatic Research Unit (CRU).

The NAO is characterized by fluctuations in SLP differences. Index values for the NAO are calculated as the difference in SLP between Gibraltar and a station over southwestern Iceland and are obtained from the CRU (Jones et al. 1997). Monthly values can be considered an indicator of the strength and/or position of the subtropical Bermuda high (Elsner et al. 2001). We

speculate that the relationship might result from a teleconnection between the midlatitudes and tropics, whereby a below normal NAO during the spring leads to dry conditions over the continents and to a tendency for greater summer/fall middle-tropospheric ridging (enhancing the dry conditions). Ridging over the eastern and western sides of the North Atlantic basin tends to keep the middle tropospheric trough, responsible for hurricane recurvature, farther to the north during the peak of the season (Elsner and Jagger 2006).

The Atlantic SST covariate is an area-weighted average based on monthly SST values in 5° latitude–longitude boxes (Enfield et al. 2001) from the equator to 70°N latitude. Monthly values of Atlantic SST are the standard Atlantic multidecadal oscillation (AMO) index.

We consider also the influence variations in the sun might have on near-coastal hurricane activity. Interest is motivated by a recent work that speculates an increase in solar ultraviolet (UV) radiation during periods of strong solar activity will have a suppressing effect on tropical cyclone intensity as the temperature near the tropopause will warm through absorption of radiation by ozone and be modulated by dynamic effects in the stratosphere (Elsner and Jagger 2008). The sunspot numbers produced by the Solar Influences Data Analysis Center (SIDC), World Data Center for the sunspot index at the Royal Observatory of Belgium are obtained from NOAA.

The monthly covariate values are shown in Fig. 3 as image plots. The monthly values for May–October on the vertical axis are plotted as a function of year on the horizontal axis. The values are shown using a color ramp from blue (low) to yellow (high). The SST and SSN covariates are characterized by high month-to-month correlation as can be seen by the vertical striations.

Next, the annual frequency of hurricanes in the vicinity of the United States is modeled using count and monthly environmental data over the period 1866–2008. The environmental data include Atlantic SST, the SOI, an index for the NAO, and SSN. All four covariates have been statistically linked to U.S. coastal hurricane activity (Elsner and Jagger 2006, 2008). The hurricane counts represent the total number of near-U.S. hurricanes for the given season (predictand), and the covariates are monthly values from May through October.

4. Results

Assuming that hurricane occurrences follow a Poisson process with an unknown rate, annual hurricane counts follow a Poisson distribution with an unknown rate parameter. We assume that the logarithm of the rate parameter is a linear combination of some fixed but unknown

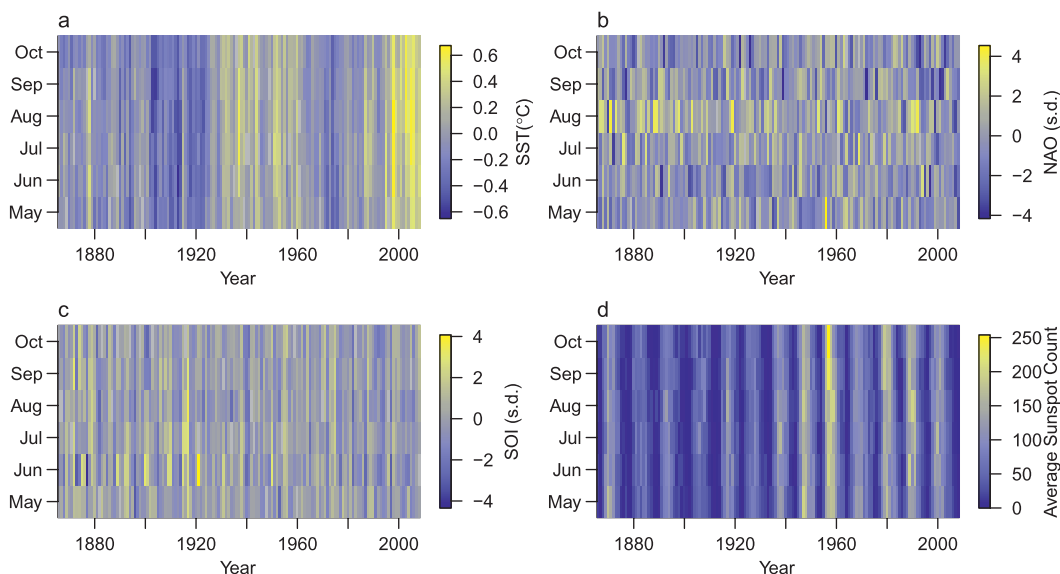


FIG. 3. Monthly values of the covariates. The covariates include (a) SST, (b) NAO, (c) SOI, and (d) sunspot number. The image plot shows the monthly values for May–October on the vertical axis as a function of year on the horizontal axis. The values are shown using a color ramp from blue (low) to yellow (high).

subset of our covariate set. Since the predictand is the set of annual hurricane counts, the specification is known as a Poisson GLM. The set of parameters (one for each covariate and the intercept) in a GLM can be found using the maximum likelihood method.

With 6 months and 4 environmental variables per month we have 24 covariates, so the total number of models is 2^{24} or more than 16.5 million. We reduced this number to 432 by first considering only the top 150 models of each size. From this subset of $150 \times 24 = 3600$ models, we reduce the set further by comparing the difference in BIC between each model and the model with the lowest BIC and keeping those models whose difference is less than 20. The calculations are carried out using the BMA package as discussed by Raftery et al. (2009) within the R programming language (R Development Core Team 2009). Specifically, we use the “bic.glm” function for the BMA procedure and “imageplot.bma” for displaying the results.

BIC is used as an approximation to the logarithm of $P(D|M)$ (Raftery 1996). The models are ordered by BIC so that the first model has the lowest BIC value, the second model has the second lowest BIC, and so on. The value of BIC for a given model is

$$-2\ln L + k \ln(n),$$

where L is the likelihood evaluated at the parameter estimates, k is the number of parameters to be estimated, and n is the number of years. BIC includes a penalty term $[k \ln(n)]$, which makes it useful for comparing

models with different sizes. If the penalty term was removed, $-2\ln L$ could be reduced just by increasing the number of model covariates. The BIC as a selection criterion results in choosing models that are parsimonious and asymptotically consistent, meaning that the model with the lowest BIC converges to the “true” model as the number of years of data increases.

The top five models having the lowest BIC values are given in Table 1. The table lists the intercept and covariates in the first column. The second column gives the posterior probability that the model parameter is not equal to zero across all 432 models. For example, the June SST covariate has a probability of 10.1% of being included in a model. The third and fourth columns are the posterior expected value and standard deviation across all models. The subsequent five columns are the five most probable models as indicated by values in rows corresponding to a covariate. For instance, the covariates included in the most probable model (model 1) are July SST, July SOI, June SSN, and September SSN. The number of variables in the model, the model BIC, and the posterior probability are also given in the table.

All 432 models are shown in Fig. 4 ordered left to right by decreasing posterior probability on the model. The first model is on the far left with the included covariates (July SST, July SOI, and SSN for June and September) shown with a color bar and the color corresponding to the model parameter sign (red for positive and blue for negative). The signs indicate that U.S. hurricane probability increases with July SST, July SOI, and June SSN and decreases with September SSN. The width of the bar

TABLE 1. Posterior statistics of the output from a BMA procedure. The EV and SD are the expected value and standard deviation of the posterior parameters, respectively. The values under the model numbers are the regression coefficients with positive values indicating a positive relationship between the covariate and the probability of a hurricane.

Term	$P(\text{Term} \neq 0)$	EV	SD	Model				
				1	2	3	4	5
Intercept	100.0	0.784	0.1104	0.803	0.830	0.708	0.722	0.702
sst.Oct	4.7	0.013	0.1191	—	—	—	—	—
sst.Sep	9.5	-0.067	0.2964	—	—	—	—	—
sst.Aug	10.1	0.026	0.2219	—	—	—	—	—
sst.Jul	58.2	0.476	0.5494	0.637	0.677	—	—	—
sst.Jun	10.1	0.030	0.2246	—	—	—	—	—
sst.May	4.4	-0.017	0.1592	—	—	—	—	—
nao.Oct	0.7	0.000	0.0035	—	—	—	—	—
nao.Sep	1.8	0.001	0.0069	—	—	—	—	—
nao.Aug	1.0	0.000	0.0043	—	—	—	—	—
nao.Jul	4.5	0.002	0.0149	—	—	—	—	—
nao.Jun	39.6	-0.040	0.0570	—	—	-0.116	-0.117	-0.108
nao.May	6.4	-0.004	0.0191	—	—	—	—	—
soi.Oct	34.0	0.052	0.0820	—	—	0.163	—	—
soi.Sep	10.8	0.015	0.0496	—	—	—	—	—
soi.Aug	25.3	0.038	0.0731	—	0.164	—	0.154	—
soi.Jul	36.1	0.063	0.0967	0.168	—	—	—	0.156
soi.Jun	8.1	-0.008	0.0317	—	—	—	—	—
soi.May	0.5	-0.000	0.0054	—	—	—	—	—
ssn.Oct	8.8	-0.000	0.0015	—	—	—	—	—
ssn.Sep	89.6	-0.009	0.0049	-0.011	-0.011	-0.011	-0.012	-0.012
ssn.Aug	1.7	-0.000	0.0005	—	—	—	—	—
ssn.Jul	1.8	-0.0000	0.0005	—	—	—	—	—
ssn.Jun	66.8	0.006	0.0050	0.008	0.008	0.010	0.010	0.010
ssn.May	10.0	0.001	0.0019	—	—	—	—	—
BIC				160.9	161.0	161.2	161.6	161.7
$P(D M)$				0.019	0.017	0.016	0.013	0.013

is proportional to the posterior probability so the bars become narrow with increasing model number.

The most important covariates for explaining annual variation in U.S. hurricane probability are easy to pick out on the image plot. They are the ones with the most consistent coloring from left to right across the image. The fewer the gaps, the more often the covariate is chosen in a model. These include September and June SSN, June NAO, July SST, and any of the months of July–September for the SOI. The results provide insight into the environmental factors important in hurricane activity. For instance, considering that on average August has many more hurricanes than July, why is July SST selected as a model covariate more often than August SST? The answer lies in the fact that when the hurricanes arrive in August and September, they lower the SST so the correlation between hurricane activity and SST weakens. That is, the thermodynamics of hurricane genesis and intensification works against the correlation. July SST better portends an active hurricane season, not because a warm ocean in July causes tropical cyclone intensification in August and September, but because hurricanes in August and September cool the ocean.

September SSN is the most consistently chosen covariate followed by June SSN. The sign on the September SSN parameter is negative indicating that the probability of a U.S. hurricane decreases with the increasing number of sunspots. This result accords with the hypothesis that increases in UV radiation from an active sun (greater number of sunspots) warms the upper troposphere resulting in greater thermodynamic stability and a lower probability of a hurricane over the western Caribbean and Gulf of Mexico (Elsner and Jagger 2008; Elsner et al. 2010). The positive relationship between hurricane probability and June SSN is explained by the direct influence the sun has on ocean temperature. Other alternative explanations are possible, especially in light of role the solar cycle likely plays in modulating the NAO (Kodera 2002; Ogi et al. 2003). Additional discussion on this topic is given in the conclusions and summary section.

The SOI covariates get chosen with a mixture across the months of July–October. The posterior probability is somewhat higher for the months of June and October and smallest for August and September. With El Niño conditions, convection over the eastern equatorial Pacific

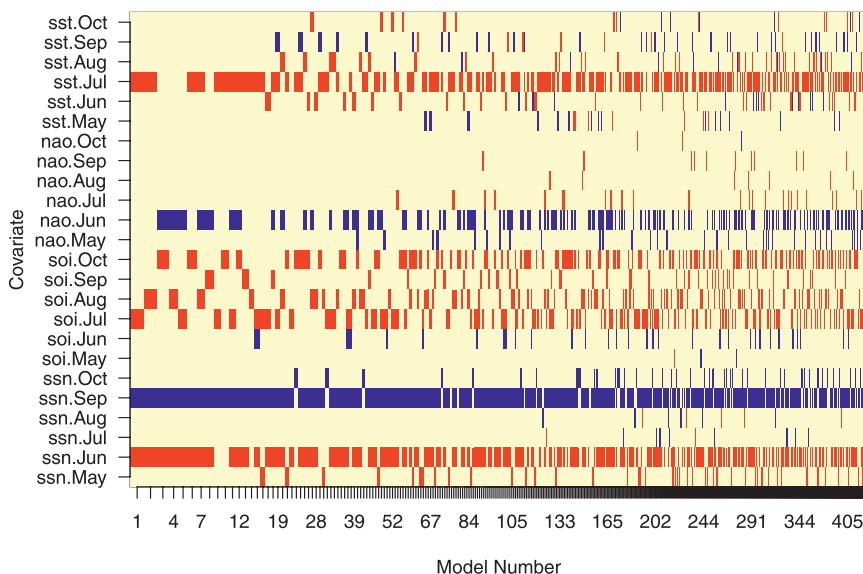


FIG. 4. Model covariates vs model number. The BIC is used to select the model. If a covariate is included in the model it is indicated by a red (positive parameter) or blue (negative parameter) bar. The bar height is constant, and the bar width is determined by the posterior probability of the model. The probabilities decrease with increasing model number.

produces increased shear and subsidence across the Atlantic (Gray 1984), but especially over the western Caribbean, where during the months of July and October a relatively large percentage of the North Atlantic hurricane activity occurs. Moreover, the inhibiting influence of El Niño might be less effective during the core months of August and September when, on average, other conditions tend to be favorable.

We determine the probability that a covariate by type irrespective of month will be chosen for at least 1 month by calculating the total posterior probability over all models that include at least 1 month of a given covariate. We find that SOI has the highest probability of being chosen at 98.1%. The SSN is close with a 97.6% chance, followed by SST at 81%, and the NAO at 48%. The lower probability of choosing the NAO reflects the large intraseasonal variability in this covariate as seen in Fig. 3.

For comparison we generate a random series of U.S. hurricane counts by sampling the original series. The random series has the same counts, but the counts are placed on random years over the period. The random series together with the covariates are used in the BMA procedure as before, and the results are mapped in Fig. 5. Here, the covariates are ordered by the model posterior probability. Results using the actual data are shown in the upper-left panel, and results using three randomized series are shown in the other three panels.

The comparison clearly demonstrates that, as a whole, our set of covariates has a meaningful relationship with U.S. hurricane activity. There are fewer models chosen

with the randomized datasets and the most probable model in each randomized set is the model with no covariates. In fact, the intercept-only model is between 2.2 and 5.2 times more likely than the model with at least one covariate among these three randomizations. Moreover, most of the models selected have only a single covariate and there is little consistency in the variable selected from one model to the next. This result demonstrates the efficacy of the procedure and the importance of these covariates in modulating the occurrence rate of hurricanes near the United States.

5. Consensus forecasts

In contrast to selecting a single model, the previous section demonstrates the results of assigning a posterior probability to a subset of possible models given the observed set of hurricanes. Each model can be used to make a prediction for this year's count. But which model should we believe? Fortunately, no choice is necessary. Instead, each model is used to make a prediction with the predictions subsequently averaged. The averaging procedures give greater weight to predictions from models with the higher posterior probability. The procedure, known as BMA, effectively produces a consensus forecast.

We assume for this discussion that full knowledge is available for all covariates over all months May–December. For any real preseason forecast, one must predict all the covariates from June through December

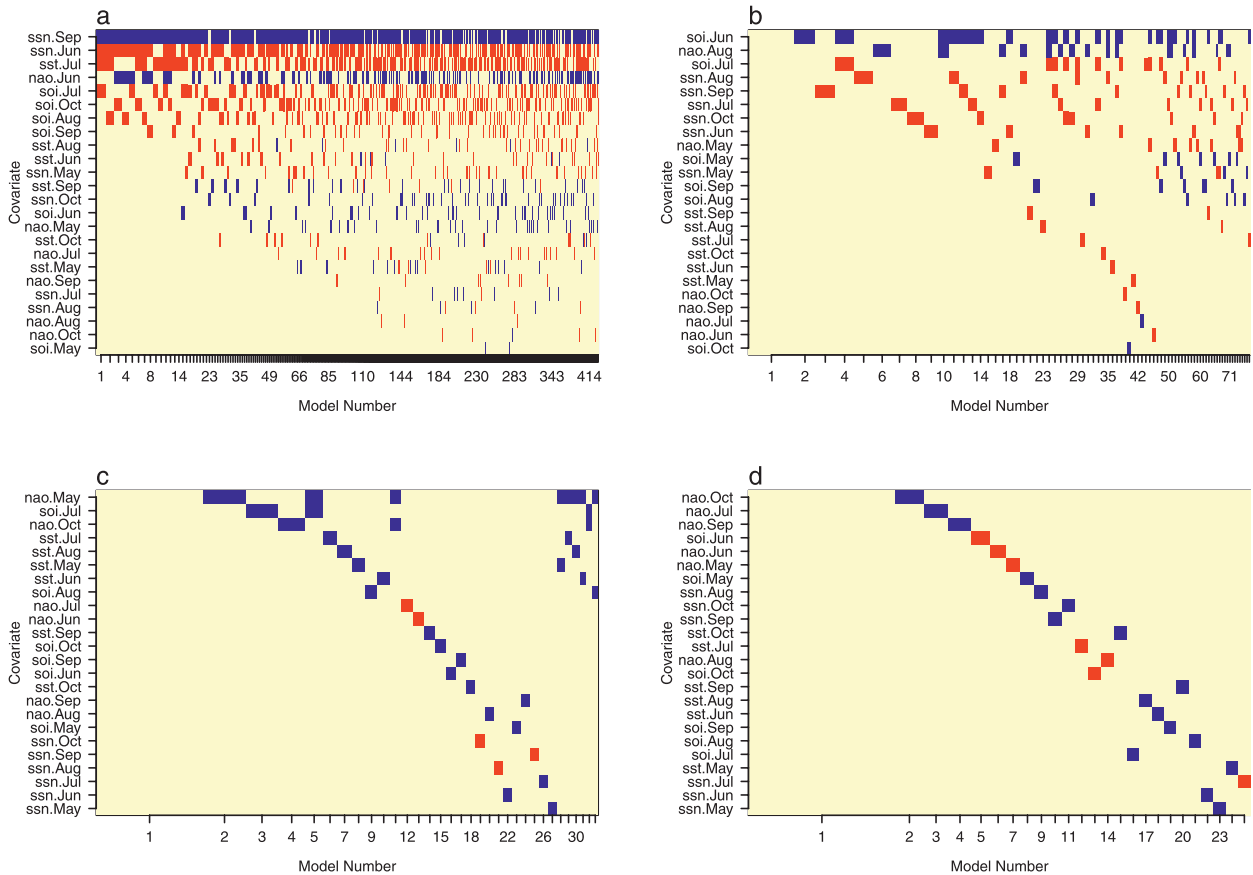


FIG. 5. Model covariates vs model number as in Fig. 4. The covariates are ordered by the model posterior probability. (a) The original time series of U.S. hurricane counts is used in the BMA procedure, and (b)–(d) three randomly sampled series of U.S. hurricanes are used.

in advance, and this prediction should be in the form of the joint distribution of all covariates over all months.

Figure 6 shows the posterior prediction probabilities for the 2007 and 2008 hurricane seasons using the consensus model from the BMA procedure. The predictions indicate a higher probability of at least one U.S. hurricane in 2008 compared with 2007. The posterior mode for the 2007 (2008) season is two (three) hurricanes. The model predicts a 58% chance of three or more hurricanes for 2007 and a 64% chance of three or more hurricanes for 2008. There was one hurricane in 2007 and three in 2008.

The consensus model predicts larger probabilities of an extreme year given the rate than would be expected from a Poisson process. That is, the consensus model is overdispersed with respect to a Poisson distribution. This makes sense as model uncertainty is incorporated in the consensus prediction. In other words, the consensus model achieved through BMA provides better coverage probabilities on the predictions.

As mentioned in the introduction, cross validation is a procedure for assessing how well an algorithm for

choosing a single best model will do in forecasting the unknown future. While BMA does not produce a single best model, here we test BMA against other model selection procedures using 11-fold cross validation. With

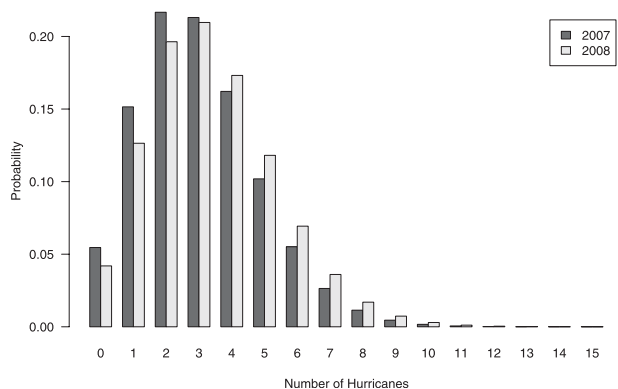


FIG. 6. Posterior predictions from the consensus model. The vertical axis is the probability of observing h number of hurricanes. Predictions are shown for the 2007 and 2008 hurricane season. There was one hurricane in 2007 and three hurricanes in 2008.

TABLE 2. Cross-validation skill scores. The skill scores include MSE, RPS, QS, and LogS. Methods include BMA and the single best Poisson GLM, where the best is defined by the smallest BIC and smallest AIC. A separate model with a constant intercept term (climatology) is also included. The values are based on an 11-fold cross-validation exercise that produces 143 separate hindcasts. A lower skill score corresponds to a procedure that results in a more accurate prediction.

Method	MSE	RPS	QS	LogS
BMA	2.125	0.786	-0.209	1.712
Smallest BIC	2.254	0.816	-0.198	1.745
Smallest AIC	2.303	0.812	-0.203	1.739
Climatology	2.261	0.805	-0.207	1.738

143 yr, each subset of data from the cross validation consists of 130 yr for training and 13 yr for testing. We compare the skill of the BMA procedure with the skill from two other selection procedures and with climatology. Both selection procedures involve the single best Poisson GLM, where best is defined in the first case as the model with the smallest BIC and in the second case as the model with the smallest AIC. Climatology involves a preselection of a fixed landfall rate in a Poisson specification.

For the BMA and the single-model selection procedures we calculate the following scoring rules as described in Czado et al. (2009): MSE, the ranked probability score (RPS), the quadratic (Brier) score (QS), and the logarithmic score (LogS). For each of these rules smaller scores indicate more accurate forecasts. Results are listed in Table 2. We find that for all scoring rules BMA provides a somewhat more accurate prediction than a procedure that selects a single best model using either BIC or AIC. Note that, since climatology as a model has been preselected, comparing its performance against the skill of the other procedures and BMA is not strictly valid.

6. Conclusions and summary

A model selection procedure (covariate screening) is typically employed to build a prediction model from data. An alternative approach advocated here is to assign a posterior probability to all models, where posterior probability reflects the likelihood of the model producing the observed data. A consensus forecast can then be issued that represents a weighted average of forecasts from all models. The weights are simply the posterior probabilities assigned to each model. The procedure is called BMA.

Here, BMA is applied to the set of hurricane counts along the U.S. coast from the period 1866–2008. Covariates include the June, July, August, September, and October monthly averages of the four covariates of SST,

NAO, SOI, and SSN. The model posterior probabilities provide insight into the physical processes connecting the covariates to near-coastal hurricane activity. For instance, July SST is selected more often than September SST as a model covariate because hurricanes in August and September cool the ocean, resulting in a lower contemporaneous correlation between ocean temperature and hurricane frequency.

The covariate chosen most often is September SSN followed by June SSN. We speculate the sun–hurricane relationship results from changes in upper-level temperatures due to changes in UV radiation. In short, increased solar activity—associated with sunspots—means more UV radiation reaching the earth’s upper atmosphere. But increased solar activity also increases the shortwave energy that helps warm the ocean that fuels hurricanes. The sun has a low-frequency period on the order of 11 yr (Schwabe cycle) that is positively correlated with Atlantic Ocean temperature and a high-frequency period on the order of 27 days (because of solar rotation) that is positively correlated with upper-atmosphere temperature. Thus, we would expect the June sunspot number to reflect the positive influence on hurricanes through the extra shortwave energy boost to the ocean and the September sunspot number to reflect the negative influence on hurricanes through the extra UV energy boost to the upper atmosphere.

Our thermodynamic hypothesis could be incomplete. For example, under an active sun the stratosphere warms unevenly, with the most-pronounced warming occurring at lower latitudes. This alters stratospheric winds (Meehl et al. 2009; van Loon et al. 2004, 2007), which could end up changing the strength of tropical cyclones. Our hunch is that circulation changes could influence weaker tropical cyclones as suggested by this work, but that the intensity of the stronger hurricanes are influenced more by thermodynamics near the surface and aloft. More work is needed on this topic.

A consensus forecast for the 2007 and 2008 season is made by averaging forecasts from all models weighted by their posterior probability. In a cross-validation exercise, the BMA procedure out performs, using various scoring rules, other model selection procedures in producing a more accurate forecast. This improvement in skill with a BMA approach notwithstanding, in a more philosophical vein we echo the sentiments of Montgomery and Nyhan (2010) that the enterprise of searching for the best model might be misguided for most climate data, which are often ambiguous as to the true model.

The consensus forecast will not necessarily give the smallest forecast error every year, but it will always provide a better assessment of forecast uncertainty compared to a forecast from a single model. The BMA

procedure provides a natural way to incorporate competing models in the forecast process, and it offers a risk management company a more rational approach to achieving consensus than has been used to date.

Acknowledgments. We thank the anonymous reviewers for their careful reviews of an earlier draft. The work was supported by Climatek Inc.

REFERENCES

- Akaike, H., 1974: New look at statistical-model identification. *IEEE Trans. Autom. Control*, **19**, 716–723.
- Castle, J. L., X. Qin, and W. R. Reed, cited 2009: How to pick the best regression equation: A review and comparison of model selection algorithms. [Available online <http://ideas.repec.org/p/cbt/econwp/09-13.html>.]
- Colman, A., and M. Davey, 2003: Statistical prediction of global sea-surface temperature anomalies. *Int. J. Climatol.*, **23**, 1677–1697.
- Czado, C., T. Gneiting, and L. Held, 2009: Predictive model assessment for count data. *Biometrics*, **65**, 1254–1261.
- DelSole, T., and J. Shukla, 2009: Artificial skill due to predictor screening. *J. Climate*, **22**, 331–345.
- Elsner, J. B., and C. P. Schertmann, 1994: Assessing forecast skill through cross validation. *Wea. Forecasting*, **9**, 619–624.
- , and T. H. Jagger, 2006: Prediction models for annual U.S. hurricane counts. *J. Climate*, **19**, 2935–2952.
- , and —, 2008: United States and Caribbean tropical cyclone activity related to the solar cycle. *Geophys. Res. Lett.*, **35**, L18705, doi:10.1029/2008GL034431.
- , B. H. Bossak, and X. Niu, 2001: Secular changes to the ENSO–U.S. hurricane relationship. *Geophys. Res. Lett.*, **28**, 4123–4126.
- , T. H. Jagger, and R. E. Hodges, 2010: Daily tropical cyclone intensity response to solar ultraviolet radiation. *Geophys. Res. Lett.*, **37**, L09701, doi:10.1029/2010GL043091.
- Enfield, D., A. Mestas-Nunez, and P. Trimble, 2001: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. *Geophys. Res. Lett.*, **28**, 2077–2080.
- Gray, W., 1984: Atlantic seasonal hurricane frequency: Part I: El Niño and 30-mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668.
- Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky, 1999: Bayesian model averaging: A tutorial. *Stat. Sci.*, **15**, 193–195.
- Jagger, T. H., and J. B. Elsner, 2006: Climatology models for extreme hurricane winds near the United States. *J. Climate*, **19**, 3220–3236.
- Jones, P., T. Jonsson, and D. Wheeler, 1997: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and southwest Iceland. *Int. J. Climatol.*, **17**, 1433–1450.
- Klotzbach, P. J., 2008: Refinements to Atlantic basin seasonal hurricane prediction from 1 December. *J. Geophys. Res.*, **113**, D17109, doi:10.1029/2008JD010047.
- Kodera, K., 2002: Solar cycle modulation of the North Atlantic Oscillation: Implication in the spatial structure of the NAO. *Geophys. Res. Lett.*, **29**, 1218, doi:10.1029/2001GL014557.
- Madigan, D., and A. Raftery, 1994: Model selection and accounting for model uncertainty in graphical models using Occams window. *J. Amer. Stat. Assoc.*, **89**, 1535–1546.
- Meehl, G. A., J. M. Arblaster, K. Matthes, F. Sassi, and H. van Loon, 2009: Amplifying the Pacific climate system response to a small 11-year solar cycle forcing. *Science*, **325**, 1114–1118.
- Michaelsen, J., 1987: Cross validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.
- Montgomery, J. M., and B. Nyhan, 2010: Bayesian model averaging: Theoretical developments and practical applications. *Polit. Anal.*, **18**, 245–270.
- Ogi, M., K. Yamazaki, and Y. Tachibana, 2003: Solar cycle modulation of the seasonal linkage of the North Atlantic Oscillation (NAO). *Geophys. Res. Lett.*, **30**, 2170, doi:10.1029/2003GL018545.
- Raftery, A., 1996: Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika*, **83**, 251–266.
- , and Y. Zheng, 2003: Discussion: Performance of Bayesian model averaging. *J. Amer. Stat. Assoc.*, **98**, 931–938.
- , T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.
- , J. Hoeting, C. Volinsky, I. Painter, and K. Y. Yeung, cited 2009: BMA: Bayesian model averaging. [Available online at <http://cran.r-project.org/package=bma>.]
- R Development Core Team, cited 2009: R: A language and environment for statistical computing. [Available online at <http://www.r-project.org>.]
- Ropelewski, C., and P. Jones, 1987: An extension of the Tahiti–Darwin Southern Oscillation index. *Mon. Wea. Rev.*, **115**, 2161–2165.
- van Loon, H., G. A. Meehl, and J. Arblaster, 2004: A decadal solar effect in the tropics in July–August. *J. Atmos. Sol. Terr. Phys.*, **66**, 1767–1778.
- , —, and D. J. Shea, 2007: Coupled air–sea response to solar forcing in the Pacific region during northern winter. *J. Geophys. Res.*, **112**, D02108, doi:10.1029/2006JD007378.
- Zou, H., and Y. Yang, 2004: Combining time series models for forecasting. *Int. J. Forecasting*, **20**, 69–84.