J4.3 AN INFORMATION-THEORETIC APPROACH TO QUANTIFYING THE UNCERTAINTY IN OPERATIONAL TROPICAL CYCLONE INTENSITY PREDICTIONS, WITH APPLICATION TO FORECAST VERIFICATION

Jonathan R. Moskaitis*
Massachusetts Institute of Technology, Cambridge, Massachusetts

1. INTRODUCTION

Quantification of the uncertainty inherent in predictions of tropical cyclone (TC) intensity is not only of scientific interest, but also is of relevance to users who must involve TC intensity forecasts in decision-making processes. Here, we describe an approach to quantify the uncertainty in (deterministic) operational TC intensity forecasts, based solely on a set of such forecasts and the corresponding set of observed intensity values. The aforementioned data sample is used to estimate an unconditional probability distribution of the observations, as well as a set of conditional probability distributions of the observations given the forecast. Qualitatively, if these conditional distributions are sharper than the unconditional distribution, then knowledge of the forecast value serves to reduce uncertainty about the value of the observation, relative to knowledge of the unconditional distribution alone. The average reduction in uncertainty about the observation due to knowledge of the forecast is quantified via calculation of the *mutual information* between the forecasts and observations, a concept borrowed from information theory.

Further details concerning mutual information and its interpretation are contained in Sec. 2. The data samples used in calculating the mutual information between various operational TC intensity forecasts and the observations are described in Sec. 3, followed by a demonstration of the calculation process for a particular data sample in Sec. 4. Sec. 5 then shows the results. It is seen that the mutual information between the various operational TC intensity forecasts and the observations is positive for all lead times (0 to 5 days), meaning that even the longer lead forecasts reduce uncertainty about the value of the observation relative to that inherent in the unconditional distribution of the observations. The ensuing discussion in Sec. 6 examines the results from the perspective of forecast verification, as mutual information can be readily interpreted as a summary measure of deterministic forecast quality. In this context, mutual information has a number of appealing properties, foremost of which is the ability to seamlessly include nominal forecasts (e.g. "dissipated") with ordinal forecasts (e.g. 70 kt) in the verification process.

*Corresponding author address: Jonathan R. Moskaitis, Rm. 54-1611, MIT, 77 Massachusetts Ave., Cambridge, MA 02139; e-mail: jonmosk@mit.edu.

2. MUTUAL INFORMATION

Fundamentally, mutual information quantifies the amount of information one random variable contains about another random variable (Cover and Thomas 2006). Suppose the two (discrete, scalar) random variables are the forecast, F, and the observation, X, such that the mutual information between the two variables is denoted $I(F;X)$. As a consequence of the mutual information between the two random variables, obtaining forecast value $F = f$ results in the average gain of $I(F;X)$ bits of information about the observed value $X = x$, relative to no knowledge of the forecast value. This average gain of information about the observation due to knowledge of the forecast, the mutual information, can be equivalently interpreted as the average reduction in the uncertainty about the observation due to knowledge of the forecast.

The aforementioned interpretations of mutual information as a difference in information/uncertainty are reflected in the formula

$$I(F;X) = H(X) - H(X|F), \qquad (1)$$

which expresses mutual information as the difference of the *entropy*, $H(X)$, and the *conditional entropy*, $H(X|F)$. The entropy pertains to the unconditional distribution of the observations, $t(x)$. It is the average amount of information needed to describe realization $X = x$, assuming $X \sim t(x)$. Entropy is given by the formula

$$H(X) = -\sum_x t(x)\log_2 t(x). \quad (2)$$

Entropy is minimized (at a value of $H(X) = 0$) for $t(x)$ such that X falls in a particular category with probability one, and maximized for $t(x)$ such that all categories have equal probability (Cover and Thomas 2006). In this sense, entropy measures the uncertainty in X: the least uncertainty is in a situation in which X definitely falls in a particular category, and the most uncertainty is in a situation in which X has equal probability of falling in each of the categories.

The conditional entropy is similar in nature to the entropy, but pertains instead to the set of conditional distributions of the observations given the forecast, $q(x|f)$. It is the average amount of information needed to describe realization $X = x$, assuming $X|F \sim q(x|f)$. The formula for conditional entropy is

$$H(X|F) = -\sum_f s(f)\sum_x q(x|f)\log_2 q(x|f), \quad (3)$$

where $s(f)$ is the unconditional probability distribution of the forecasts. The interpretation of conditional entropy in terms of uncertainty is analogous to that of entropy described previously.

It follows from Eq. 1, coupled with the interpretation of entropy in terms of uncertainty, that the mutual information, $I(F;X)$, is the average reduction in uncertainty about the observation due to knowledge of the forecast, relative to the uncertainty inherent in the unconditional distribution of the observations. The minimum value of $I(F;X)$ is zero, in the case that F and X are independent, such that the conditional entropy is exactly the same as the entropy (i.e. conditioning on an independent variable does not reduce uncertainty in X). The maximum value of $I(F;X)$ is $H(X)$, in the case that F and X are absolutely dependent, such that the conditional entropy is zero (i.e. conditioning on an absolutely dependent variable eliminates uncertainty in X). Note that absolute dependence does not necessarily imply $F = X$, but rather that a given value of F is always paired with a particular value of X in the data sample. The maximum and minimum values of $I(F;X)$ are such as to make a normalized version of mutual information convenient,

$$I_N(F;X) = \frac{I(F;X)}{H(X)} = 1 - \frac{H(X|F)}{H(X)}. \quad (4)$$

The maximum value of *normalized mutual information*, $I_N(F;X)$, is one and the minimum value is zero.

3. DATA SAMPLES

Sec. 2 shows that to calculate the mutual information between forecasts and observations, estimates of the probability distributions $t(x)$, $q(x|f)$, and $s(f)$ are necessary. These distributions can be obtained through manipulation of an estimate of the joint distribution of the forecasts and observations, $p(f,x)$. Here, a joint distribution is estimated according to the relative frequencies in a data sample consisting of operational TC intensity forecasts and the corresponding best track observations, denoted $\{(f_k, x_k); k = 1 \ldots N\}$. A data sample pertains to a particular lead time and TC intensity prediction system. Eight lead times (ranging from 0 to 120 h) and four TC intensity prediction agencies/systems (NHC, GFDL, D-SHIPS, SHIFOR) are considered here, making for a total of 32 data samples. The characteristics of these data samples are now briefly described.

The data samples pertain to operational intensity forecasts and observations of Atlantic basin TCs from the 2001 through 2005 seasons. Every data sample consists of $N = 1965$ (f,x) pairs, encompassing all realizations in which a forecast was initialized from each of the four forecast systems. Forecast and observed intensities are categorized into 1 of 30 categories: 29 ordinal categories of 5 kt width (centered every 5 kt from 20 kt to 160 kt), and 1 nominal category for the dissipated TC. A prediction of “dissipated” is assumed for lead times beyond the last lead time an ordinal forecast is issued, or if the ordinal forecast value is less than 17.5 kt. An observation of “dissipated” occurs if there is no intensity value in the best track corresponding to a time a forecast is valid, or if an ordinal best track value is less than 17.5 kt.

4. EXAMPLE CALCULATION

Here, the process by which mutual information is calculated is demonstrated for a particular data sample, consisting of 12 h lead time NHC TC intensity forecasts and the corresponding observations. Select probability distributions relevant to the calculation are displayed in Fig. 1. Fig. 1a shows the unconditional distribution of the observations, $t(x)$. Using Eq. 2, the entropy is calculated to be 4.30

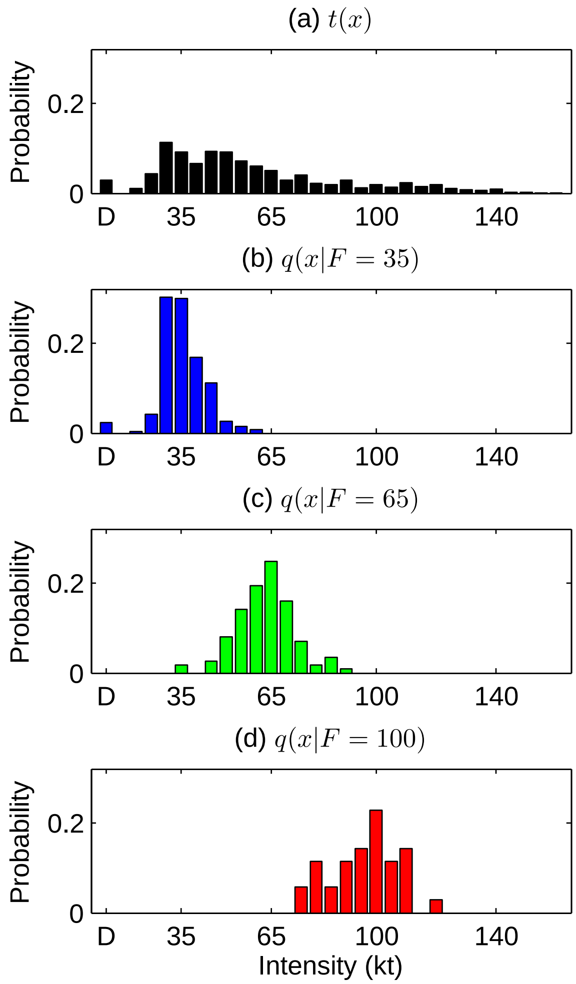

Figure 1: Select probability distributions pertaining to the calculation of the mutual information between the 12 h NHC TC intensity forecasts and observations. (a) Unconditional distribution of the observations, $t(x)$. (b) Conditional distribution of the observations given a forecast of 35 kt, $q(x|F = 35)$. (c-d) As in (b), but conditioning on a forecast of 65 kt and 100 kt, respectively. Note that in all panels, "D" denotes the nominal category of "dissipated", to the left of the ordinal intensity categories.

bits. This value of entropy is quite large (for reference, the maximum possible value is 4.95 bits), since the probability is spread out amongst many categories of observation, such that there is considerable uncertainty present. Fig. 1b-d shows three components of the set of conditional probability distributions of the observations given the forecast, $q(x|f)$; the distribution in Fig. 1b is conditioned on

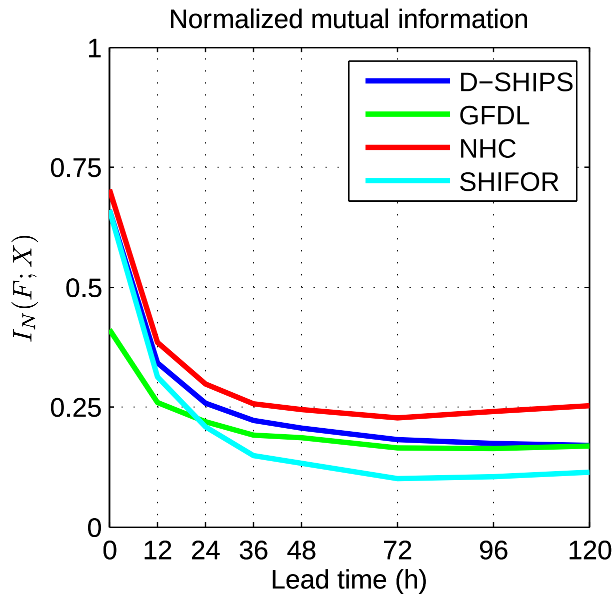

Figure 2: Normalized mutual information, as a function of lead time, for four operational TC intensity prediction systems.

a forecast of 35 kt, the distribution in Fig. 1c on a forecast of 65 kt, and the distribution in Fig. 1d on a forecast of 100 kt. These conditional distributions are clearly sharper in nature than the unconditional distribution in Fig. 1a, as probability is concentrated in a relatively limited number of categories. Thus, conditioning on the NHC 12 h forecast qualitatively appears to reduce the uncertainty in the value of the observed TC intensity. This inference is borne out in the calculation of the conditional entropy (using Eq. 3), which is 2.64 bits. Subtracting the conditional entropy from the entropy yields $I(F;X) =$ 1.66 bits for the mutual information between the 12 h NHC TC intensity forecasts and the observations (see Eq. 1), or $I_N(F;X) = 0.39$ after normalization. Hence, on average, knowledge of the 12 h NHC forecast reduces the uncertainty in the observation by 39 percent, relative to the uncertainty inherent in the unconditional distribution of the observations.

5. RESULTS

Fig. 2 shows the normalized mutual information, $I_N(F;X)$, for TC intensity forecasts from the NHC, GFDL, D-SHIPS, and SHIFOR prediction systems as a function of lead time. Normalized mutual information calculations were performed using the data samples described in Sec. 3 to estimate the probability distributions $t(x)$, $q(x|f)$, and $s(f)$ for each forecast system/lead time combination. Broadly speaking, $I_N(F;X)$ decreases rapidly for the first 24 hours of lead time, then decreases at a much slower rate through the 72 h lead time, before remaining roughly constant out to 120 h. Since $I_N(F;X)$ does not asymptote to zero with lead time, it is the case that the long-lead TC intensity forecasts reduce uncertainty about the value of the observation (on average), relative to the uncertainty inherent in the unconditional distribution of the observations. However, beyond 36 h the forecasts only reduce the uncertainty about the observation by 10–25 percent.

6. APPLICATION TO VERIFICATION

In addition to its role in the quantification of uncertainty, the mutual information between forecasts and observations can be utilized as a verification measure (as has been suggested by Leung and North 1990; DelSole 2005). Specifically, mutual information is a summary measure of forecast information content, a fundamental attribute of forecast quality. Mutual information has a number of characteristics that distinguish it from traditional summary measures of forecast accuracy, such as mean absolute error (MAE). As described previously, the calculation of $I(F;X)$ depends on an estimate of $p(f,x)$, the joint distribution of the forecasts and observations. Thus, calculation of mutual information is a natural extension to the distributions-oriented approach to verification (Murphy and Winkler 1987), which is based on analysis of $p(f,x)$. It is unnecessary to construct an estimate of $p(f,x)$ to calculate MAE, however, as MAE can be obtained through direct operation on the (f,x) pairs of the data sample.

Of particular pertinence to the verification of deterministic TC intensity forecasts is the difference between mutual information and MAE in the ability to handle nominal designations of the forecast and observed categories. Mutual information can accommodate a nominal, ordinal, or mixed categorization of the forecasts and observations because $I(F;X)$ is a function of only the probabilities of the categories (see Eqs. 1–3). MAE, which can be expressed as

$$\mathrm{MAE} = \sum_f \sum_x p(f,x)\,|f-x|\,, \qquad (5)$$

is a function of the probabilities of the categories

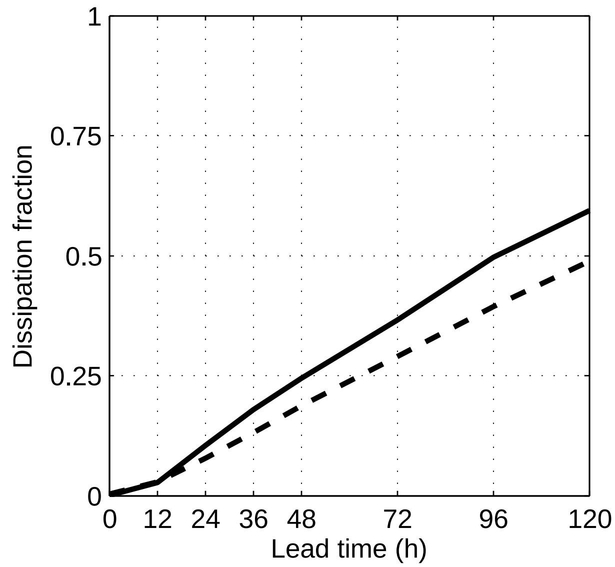

Figure 3: Dashed line shows the fraction of realizations in the data samples, as described in Sec. 3, in which the observation is classified as “dissipated” rather than as an ordinal intensity value. Similarly, the solid line shows the fraction of realizations in which the observation or any of the four operational TC intensity forecasts is classified as “dissipated”. Such realizations cannot be included in mean absolute error verification.

(elements of $p(f,x)$) and the designations of the categories (x and f). As such, nominal categories must be excluded from consideration in the calculation of MAE, to avoid taking the difference of 50 kt and “dissipated”, for instance. Consequentially, MAE verification of operational TC intensity forecasts is forced to ignore (f,x) realizations involving dissipation. Such realizations constitute a significant fraction of the total realizations in the data samples utilized here, as shown in Fig. 3. These realizations can be seamlessly included in the operational TC intensity forecast verification process by utilizing mutual information as a verification measure.

Given the interpretation of mutual information as a summary verification measure, a cautionary note concerning the relative performance of the operational TC intensity forecast systems in Fig. 2 is warranted. Over the five years of realizations encompassed in the data samples, the forecast systems themselves evolved substantially, such that the results may not necessarily be representative of the performance of the 2005 (or current) versions of the forecast systems. The application of mutual information verification to smaller data samples (tenable, perhaps, by reducing the number of ordinal intensity categories or statistically modeling the joint distribution) is a subject for future research.

Acknowledgement

The author wishes to thank Kerry Emanuel and James A. Hansen for their guidance, and acknowledge the support of ONR Grant N00014-05-1-0323.

References

Cover, T. M. and J. A. Thomas, 2006: *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ.

DelSole, T., 2005: Predictability and information theory. Part II: Imperfect forecasts. *Journal of the Atmospheric Sciences*, **62**, 3368–3381.

Leung, L. and G. R. North, 1990: Information theory and climate prediction. *Journal of Climate*, **3**, 5–14.

Murphy, A. H. and R. L. Winkler, 1987: A general framework for forecast verification. *Monthly Weather Review*, **115**, 1330–1338.