An information-theoretic approach to reconciling historical climate observations and impacts on agriculture

An Information-Theoretic Approach to Reconciling Historical Climate Observations and Impacts on Agriculture

MAX MAUERMAN,[a] EMILY BLACK,[b,c] VICTORIA L. BOULT,[b,c] RAHEL DIRO,[d] DAN OSGOOD,[a] HELEN GREATREX,[e,f,g] AND THABBIE CHILLONGO[h]

[a] *International Research Institute for Climate and Society, Columbia University, New York, New York*
[b] *National Centre for Atmospheric Science, Leeds, United Kingdom*
[c] *Department of Meteorology, University of Reading, Reading, United Kingdom*
[d] *Tetra Tech, Pasadena, California*
[e] *Department of Geography, The Pennsylvania State University, University Park, Pennsylvania*
[f] *Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania*
[g] *Institute for Computational and Data Sciences, The Pennsylvania State University, University Park, Pennsylvania*
[h] *Centre for Agricultural Research and Development, Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi*

(Manuscript received 10 February 2022, in final form 14 September 2022)

ABSTRACT: Decision-makers in climate risk management often face problems of how to reconcile diverse and conflicting sources of information about weather and its impact on human activity, such as when they are determining a quantitative threshold for when to act on satellite data. For this class of problems, it is important to quantitatively assess how severe a year was relative to other years, accounting for both the level of uncertainty among weather indicators and those indicators' relationship to humanitarian consequences. We frame this assessment as the task of constructing a probability distribution for the relative severity of each year, incorporating both observational data—such as satellite measurements—and prior information on human impact—such as farmers' reports—the latter of which may be incompletely measured or partially ordered. We present a simple, extensible statistical method to fit a probability distribution of relative severity to any ordinal data, using the principle of maximum entropy. We demonstrate the utility of the method through application to a weather index insurance project in Malawi, in which the model allows us to quantify the likelihood that satellites would correctly identify damaging drought events as reported by farmers, while accounting for uncertainty both within a set of commonly used satellite indicators and between those indicators and farmers' ranking of the worst drought years. This approach has immediate utility in the design of weather-index insurance schemes and forecast-based action programs, such as assessing their degree of basis risk or determining the probable needs for postseason food assistance.

SIGNIFICANCE STATEMENT: We present a novel statistical method for synthesizing many indicators of drought into a probability distribution of how bad an agricultural season was likely to have been. This is important because climate risk analysts face problems of how to reconcile diverse and conflicting sources of information about drought—such as determining a quantitative threshold for when to act on satellite data, having only limited, ordinal information on past droughts to validate it. Our new method allows us to construct a probability distribution for the relative severity of a year, incorporating both kinds of data. This allows us to quantify the likelihood that satellites would have missed major humanitarian droughts due to, for example, mistimed observations or unobserved heterogeneity in impacts.

KEYWORDS: Bayesian methods; Decision support; Indigenous knowledge; Ranking methods; Risk assessment; Societal impacts

1. Motivation and literature review

A common class of problems in climate risk management concerns how to effectively summarize a number of imperfect indicators of damaging weather for the purpose of making a policy decision. For instance, in determining a quantitative threshold to distinguish "drought" from "nondrought" years for the purposes of farmer assistance, researchers often consider many data sources, including both instrumental observations of environmental conditions—such as rainfall and vegetation indices—as well as information on droughts' human impact, such as nutrition indicators, yields and farmers' own reports (Benami et al. 2021). The latter type of information is vital for decision-makers, as the relationship between measured weather and its humanitarian consequences is a function of farmers' practices, their social vulnerability, and other factors that are not easily measured (Enenkel et al. 2020). However, such impact data are often measured infrequently or selectively and may not be available for every year or location under study. Thus, in determining a decision rule for taking action against drought, policy makers face both uncertainty within many possible instrumental measurements of weather and uncertainty between those measurements and the human

Denotes content that is immediately available upon publication as open access.

Corresponding author: Max Mauerman, mmauerman@iri.columbia.edu

DOI: 10.1175/WCAS-D-22-0019.1

impacts of concern—the latter of which may only be partially known.

Examples of policies that face this type of problem include parametric weather insurance (Hochrainer et al. 2009; Maganga et al. 2021) and forecast-based early action protocols (Coughlan de Perez et al. 2015). Both types of program tend to rely on retrospective evaluation to assess their efficacy, as damaging weather events are infrequent and counterfactual data is by nature limited (Bucheli et al. 2021; Lobell et al. 2020; Osgood and Shirley 2012). Commonly, such evaluations compare the estimated historical payouts of the program under study with the economic value of participants' losses from adverse weather, as captured by loss of agricultural yields, damage to property, and so on (Meuwissen et al. 2019; Ortiz-Bobea et al. 2018; Schlenker and Roberts 2009). From such data, standard economic estimates of program value such as expected utility can be constructed (Benami et al. 2021).

However, in many of the countries where parametric insurance and forecast-based action are being practiced, impact data such as local historical yields are frequently incomplete, unreliable, or unavailable, for the very reasons that motivate parametric insurance—the prohibitive cost of and lack of infrastructure for routine individual assessments of loss (Enenkel et al. 2020). Consequently, there has been much work in the field on how to conduct constructive evaluations of such programs under conditions of limited information (Brahm et al. 2019; Osgood et al. 2018; Osgood and Shirley 2012). One way forward is to focus on the ordinal comparison of years—since financial constraints naturally put a limit on the number of years in which a program can pay out, the problem can be posed as one of assessing the similarity between the worst years in the observational data and the worst years in terms of farmer impact. This offers a way to assess program reliability even where "benchmark" data are limited (Enenkel et al. 2019).

The challenge, then, is to conduct such ordinal comparison in a rigorous way, accounting for a variety of potential weather measurements as well as their relationship to farmer impact, which may only be partially known. We frame this problem as one of constructing a probability distribution for the relative severity of each year in the record. This probability distribution should give decision-makers a concise historical summary of where there is a convergence of evidence for or against a given catastrophic event (such as drought), as well as where the record is conflicted, which may indicate a probable basis risk event. Such a probability distribution should satisfy the following practical and theoretical considerations:

- First, it ought to have a minimal number of distributional assumptions—that is, it should not be constrained to fit a certain parametric family such as a Gaussian distribution, as the distribution of risk exposure data is often irregular, multimodal, and nonnormal (Svensson et al. 2017).
- Second, it ought to allow the user to incorporate a semistructured source of prior information on human impact—for instance, farmers' reported bad harvest years, experts' knowledge, or food security measurements—all of which could act as a check on the range of plausible distributional outcomes [see also Camenisch et al. (2022) and Nicholson (2001) for similarly motivated approaches to estimating historical climate impact].
- Third, it ought to be robust to small N and partially missing data—this is a common occurrence for climate and social data in developing countries, and many expectation maximization (EM)-based methods, such as mixture models, struggle to fit such data (Golan 2017).

In this paper, we present a computationally simple, extensible statistical method for estimating the probability distribution of relative weather severity that satisfies all three of these requirements and can be applied to a wide class of important problems in climate risk analysis. We refer to this method, which is based on the principle of maximum entropy (Golan 2017), as the "MaxEnt" probability model.

We can think of the MaxEnt method as modeling the most parsimonious data-generating process that is consistent with users' prior information on which historical events were impactful. Such a model blurs the distinction between "exploratory" and "confirmatory" data analysis; as Hullman and Gelman (2021) observe, much exploratory analysis is in fact a process of reconciling discrepancies between observations and the user's mental model of the world—that is, their Bayesian priors (see also Karduni et al. 2021; Kim et al. 2019; Wickham et al. 2010).

We begin with the basic framework of the MaxEnt model (following Faynzilberg 1996b,a), gradually introduce refinements, then present a case study of the model's application. In the case study, we use actual satellite and farmer survey data from an index insurance project in Malawi to accomplish two tasks: Summarizing the historical record of drought indicators and identifying potential "basis risk" years in which an indicator failed to capture a probable drought event. This case study builds on several of the authors' prior involvement in evaluating the Africa Risk Capacity (ARC) sovereign insurance facility in Malawi. In the conclusions section, we discuss future extensions to the MaxEnt model and possible policy applications beyond drought insurance.

2. Methodology

a. *Problem statement*

We have an $n \times k$ matrix $\mathbf{X}$ that consists of k different observational series of weather indicators measured over n years, all of which measure a particular hazard, such as drought. (Note that we will use years as the unit of aggregation in this example, but this method could be applied to other temporal or spatial units.) Each observation is expressed in terms of its position in the historical ranking for that indicator, ranging from 1 (the worst year on record) to 0 (the least bad). We can also think of this value as the reciprocal of the return period. For instance, in a dataset of 20 years of drought measurements, a value of 0.2 would correspond to the 16th driest year on record, that is, a return period of 1/5. (Continuous data must first be normalized, i.e., expressed in ordinal terms, in order to fit this format.)

We also have an $n \times 1$ matrix $\mathbf{Y}$, which represents the aggregate outcome for each year. In its simplest form, $\mathbf{Y}$ is

TABLE 1. Simulated example data, raw.

Year	Source 1	Source 2	Source 3	Prior
2000	1	1	1	1
2001	2	2	2	Missing
2002	0	0	1	5
2003	5	5	0	2
2004	4	3	4	4

TABLE 2. Simulated example data, ranked and averaged.

Year	Source 1	Source 2	Source 3	$E(X)$	Prior
2000	0.4	0.4	0.6	0.47	0.4
2001	0.6	0.6	0.8	0.67	Missing
2002	0.2	0.2	0.6	0.33	1
2003	1	1	0.2	0.73	0.6
2004	0.8	0.8	1	0.87	0.8

calculated as $E[X(n),]$—that is, the mean for that year as calculated over every observational indicator—but it could also incorporate higher-order moments, like the standard deviation (Golan 2017; Harte 2011).

We aim to construct a generative probability model (Läderach et al. 2017)—for each year, given some observed hazard indicators $\mathbf{X}$ and some aggregate constraints y, we wish to estimate the probability distribution for its likeliest position in the historical ranking (aka its return period)—that is, $P(\mathbf{X}|\mathbf{Y} = \mathbf{y})$. This is in contrast to discriminative methods like logistic regression, which estimate $P(\mathbf{Y}|\mathbf{X} = \mathbf{x})$. The goal of our model is not to classify observations according to their severity, as in a discriminative model, but to reconstruct the most plausible probability distribution that could have generated our observations, subject to some constraints.

Our goal, then, is to find the most plausible probability distribution for the relative severity of each year that could have generated the observed average outcome y, given the data x. To find this, we need to fix two more concepts: The outcome space of the model, and the objective function for determining which distribution is the most "plausible."

The distribution-generating model $S(l)$ is very simple: it breaks the outcome space $(0, l)$ into l discrete chunks. For instance, $S(5)$ can take on the values $(0, 0.25, 0.5, 0.75, 1)$. Each state in S is assigned a probability p. To be consistent with the aggregate outcomes, our model must satisfy the constraint

$$Sp = E(x) = y. \tag{1}$$

There are many distributions P over $S(l)$ that could generate a given y, so we must establish an objective function for selecting the most plausible one. A natural choice is the Shannon entropy, a metric that summarizes the amount of information present in a probability distribution (Shannon 1948). Shannon entropy defines information as "distinctiveness" or "surprise"—observing a likely event carries little information about the shape of a distribution but observing an unlikely event does. Over the discrete probability space $S(l)$, entropy is calculated as

$$-\sum_{p\in S(l)} p\log(p). \tag{2}$$

Absent any constraints, entropy is maximized over a space by the uniform distribution, in which no realized state provides any information about the probability of other states in the space. The principle of maximum entropy for model selection states that given some constraints, the most plausible (in the sense of being the most epistemically conservative) probability distribution is the one that maximizes entropy. Here, we apply the principle of maximum entropy to a generative problem, following other applications in climate and social sciences such as Harte (2011) and Läderach et al. (2017).

The fundamental form of our model is therefore

$$\max -\sum_{p\in S(l)} p\log(p) \text{ s.t. } \sum_{p\in S(l)} p = 1 \text{ and } S(l)P = y. \tag{3}$$

To illustrate the output of this model, we show its results when applied to some simulated example data (Table 1). In these data, a larger number means a drier year according to that source (units here are arbitrary). Table 2 shows what the data look like after being normalized, that is, expressed in ordinal terms, as well as the average outcome for each year. Figure 1 shows the resulting MaxEnt probability distribution for each year.

Using the year 2004 as an example, our model takes as input the average severity from our three data sources (0.86, shown as the green tick in the margins) and calculates that there is a 55% percent chance that 2004 was the driest year on record (the rightmost red bar), a 25% chance it was the second driest year on record (the second to rightmost red bar), and so on.

For reference, the graph shows the uniform distribution in gray. This represents what the principle of maximum entropy would yield if there were no observational data—that is, a belief that all outcomes for a given year would be equally likely. The difference between the estimated and the uniform distribution tells us how much information we have about the probable outcome in that year, relative to a state of total uncertainty.

Next, we show how this basic model can be extended to address the practical decision-making needs discussed in the introduction.

b. *Extensions to model*

1) MAXIMUM LIKELIHOOD CONSTRAINT

The basic model presented above yields estimated distributions that are consistent with the mean outcome for each year. However, we also want to represent heterogeneity between the individual predictors, especially if the distribution of predictors is multimodal. To capture this, we can introduce an additional constraint to Eq. (1):

$$\sum_{q\in S(l)} (N_q/N)\log[(N_q/N)/p] < (1/N)\log(1/\alpha), \tag{4}$$

where N_q/N is the number of predictors in state q divided by the total number of predictors and α (or alpha) is some

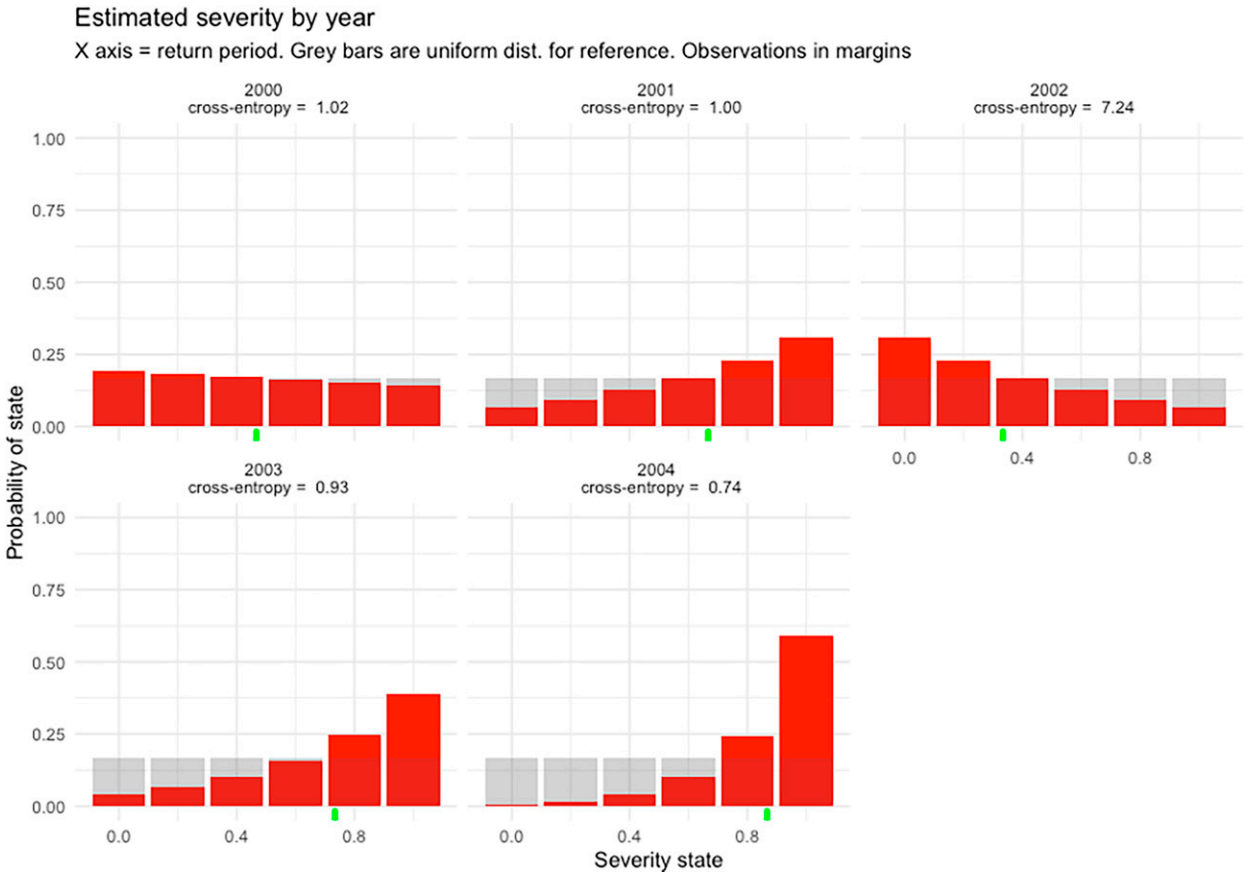

FIG. 1. Basic model results with example data. The x axis depicts the return period or relative severity, where 0 is the best/wettest year on record and 1 is the worst/driest year on record. The y axis depicts the likelihood of each return period/severity state.

arbitrary constant. This constraint states that the Kulback–Leibler divergence (i.e., the differential entropy) between the empirical frequency of each state and its estimated probability cannot be larger than some constant. As α or $N \rightarrow$ infinity the estimated distribution essentially becomes a histogram.

2) PRIOR

An additional refinement to the basic model is to add a source of prior knowledge that acts as a check on the range of possible distributions. Specifically, suppose we have a prior distribution Q of the same dimension as P. We do not want our estimated distribution to be "too different" than the prior. We can operationalize this through the inclusion of a cross-entropy term in the objective function:

$$\max - \sum_{p \in S(l)} p \log(p) + \sum_{q, p \in S(l)} p \log(q). \tag{5}$$

A uniformly distributed Q is equivalent to having no prior at all.

We can extend this farther by putting a user-defined scalar, beta, on the cross-entropy term. Beta > 1 means a greater weight on the prior than the default. Beta < 1 means a lower weight. This parameter can be adjusted as the use case demands.

The prior represents our information on the actual humanitarian impact in our sector of concern. In a drought estimation problem, it might come from survey data on farmers' reported worst harvest years, or yield data. The prior can be considered to be a "benchmark"—it represents our best ex ante knowledge about the possible states of the world. If we do not observe the humanitarian impact in a given year, we can treat the prior as uniform, that is, providing no additional information about the probable outcome in that year. Note that if the prior data source takes the form of a single point per year, it may be desirable to first transform it into a probability distribution via the iterative application of the maximum entropy method, as described above.

To illustrate the effect of a prior, consider what happens if we introduce a prior in our previous example, Fig. 1. The distribution with prior is shown in Fig. 2. In 2000, 2004, and 2003, the prior data do not differ significantly from the observational data, so the shape of the estimated distribution does not change. Likewise, in 2001, there is no prior, so the distribution is not affected. However, in 2002, there is a strong prior (at a severity of 1), diverging from the observational data, that pulls the distribution away from a right-tailed shape and closer to a uniform distribution.

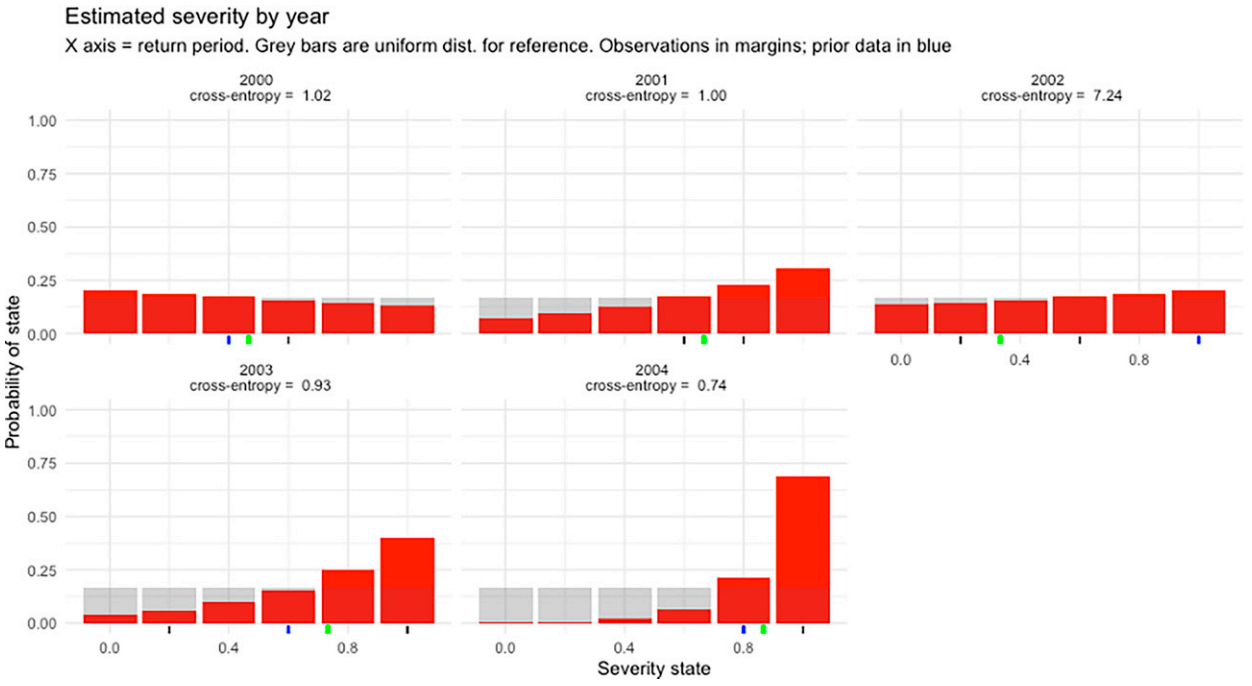

FIG. 2. Example results with prior added.

3) ERROR TERM

Last, we may want to introduce an error term into our model. There are two reasons that this might be useful. First, in a complex model with all the extensions described above, some problems may not always be tractable. Second, even if we know a feasible solution to the problem exists, we may have conceptual reasons for allowing for imperfect measurement of the aggregate outcome **Y**. In both cases, an error term is useful, and the maximum entropy framework gives us an easy way to integrate one.

We define the error term analogously to the main outcome model: it can take one of f discrete states in the interval $(-1, 1)$. Each of those states has a probability e and an optional prior of its own, υ. Then, the moment-consistency condition becomes

$$S(l)P + \text{error}(f)e = y, \tag{6}$$

and the objective function becomes

$$\max -\sum_{p\in S(l)} p\log(p) + \sum_{q,p\in S(l)} p\log(q) + \sum_{e\in S(f)} e\log(e) + \sum_{e,\upsilon\in S(f)} e\log(\upsilon). \tag{7}$$

Under this new model, if the solver struggles to fit the moment-consistency constraint, it can place some uncertainty into the error term. We can refine this by putting a scalar weight gamma on the error component of the objective function, analogous to the beta weight on the prior—as gamma increases, the solver "cost" of putting uncertainty into the error term increases. We can use this parameter to adjust the model's accuracy versus precision trade-off.

We might also want to include a prior on the error term, because it is reasonable to assume that not all error sizes are equally likely ex ante—we expect larger error values to be less likely than small values. A reasonable choice of the prior is the (discrete approximation of) the normal distribution $\sim N(0, 1)$. We can obtain this for a given error space by maximizing its entropy subject to the constraints $E(\upsilon) = 0$ and $E(\upsilon)^2 = 1$.

We can also visualize the distribution of the error term itself for each year (Fig. 3) to see where the model struggled most to fit the data. In this case, 2002 has the sharpest disagreement between the various components of the objective function and thus the largest error term.

c. *Diagnostic metrics*

One advantage of the maximum entropy framework is that it lends itself naturally to useful model diagnostics. The first is relative entropy, also known as efficiency, which tells us how much overall information content our recovered distribution has, relative to a reference distribution over the same space:

$$\text{efficiency} = 1 - \text{entr}(p)/\text{entr}(\text{uniform}), \tag{8}$$

where "entr" is the entropy function as defined in Eq. (2). The most natural choice for the reference distribution is the uniform distribution, which has the highest possible entropy over a given space. It is equivalent to saying all outcomes are equally likely. No matter the choice of reference, efficiency tells us about the relative information content in our estimated distribution.

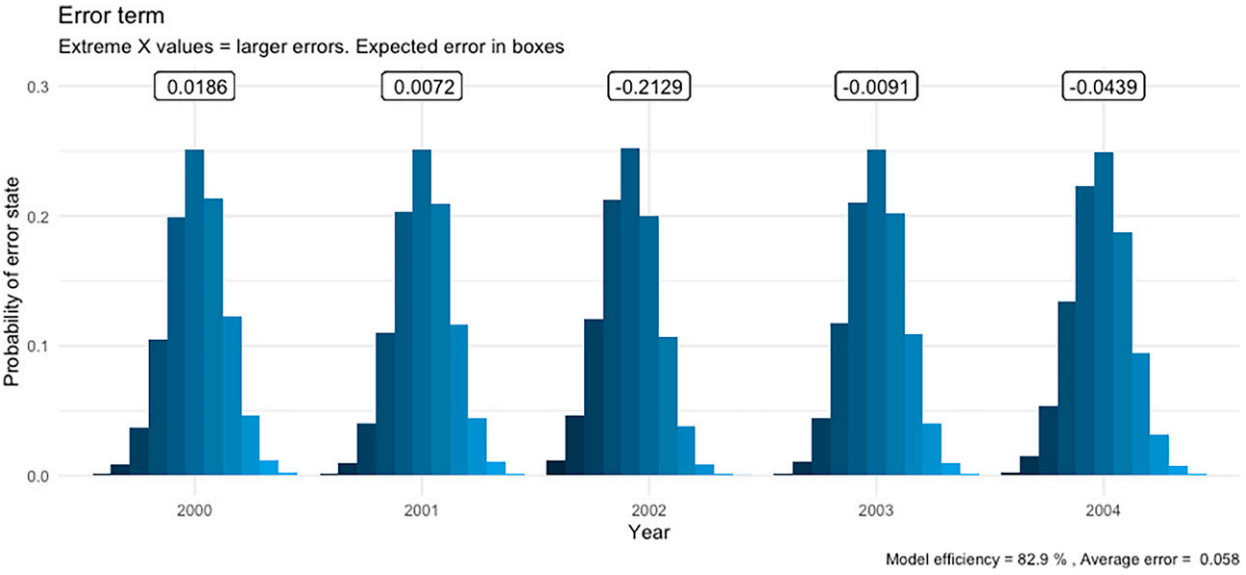

FIG. 3. Error term for example results.

The second useful diagnostic is cross entropy (CE) between the estimated distribution and the prior. For ease of interpretation, we can normalize cross entropy, in a way comparable to efficiency:

$$\mathrm{NCE} = \mathrm{CE}(p, q)/\mathrm{entr}(q). \tag{9}$$

If normalized cross entropy (NCE) is equal to or close to 1, it means the prior is uninformative about the state of the world. If NCE is < 1, it means that the prior and posterior distributions are in agreement (lower = more agreement). Conversely, if NCE is >1, it means the prior and posterior distributions are in conflict (higher = less agreement). We can use NCE to assess the convergence of evidence.

3. Case study

a. Overview

We present a case study of how this method can be applied in a humanitarian setting to facilitate practical decision-making. This case comes from the Satellites for Weather Index Insurance–Agricultural Early Warning System (SatWIN-ALERT) project, which aims to characterize basis risk (where compensation does not match farmers' observed losses) in index insurance, using drought insurance in Malawi as a case study. Basis risk in satellite-based drought insurance can arise for a variety of reasons, such as error or imprecision in satellite measures of rainfall, the disjuncture between meteorological drought (rainfall deficit) and agricultural drought (soil moisture deficit) or crop failure, and unobserved differences in farming practices, to name a few commonly cited reasons (Benami et al. 2021). These issues are relevant in Malawi, where programs like the ARC intergovernmental weather insurance facility have suffered from basis risk in the past (African Risk Capacity 2016).

To unpack these differences, the SatWIN-ALERT project examined a number of different satellite rainfall sources, along with remote sensing-based vegetation indices and models of soil moisture and water stress. To understand the human side of basis risk, the project also conducted focus groups with farmers on their worst harvest years due to drought in memory. While farmers are useful ground information sources—they have intimate knowledge of the local landscape/weather/agricultural system—farmer focus groups may suffer from their own biases, such as imperfect recall, motivated reporting, or exclusion of marginalized groups from participation.

Thus, the SatWIN-ALERT project required a statistical method that could summarize these various sources of historical risk information—satellites, models, farmers—holistically, with the goal of generating a catalog of years when biophysical measurements diverged from farmers' perceptions (and why). We demonstrate how the MaxEnt method can be applied to this task to generate new insights, allowing a nontechnical user to visualize at a glance where there is a convergence or divergence of evidence between the many sources of data, and to diagnose what may have happened during probable basis risk events.

b. Data sources

Two types of data were combined: Biophysical indicators of drought and farmer surveys.

The spatial unit of analysis for all types of data is the extension planning area (EPA), an administrative unit defining the territory of agricultural extension officers in Malawi (Fig. 5). Of the 187 EPAs, 18 were selected for this study—shown in Fig. 4—based on the criteria of being persistently food insecure and prone to drought (see appendix A for full details).

The biophysical indicators include three satellite-based rainfall estimation datasets: African Rainfall Climatology (ARC), Climate Hazards Group Infrared Precipitation with Stations

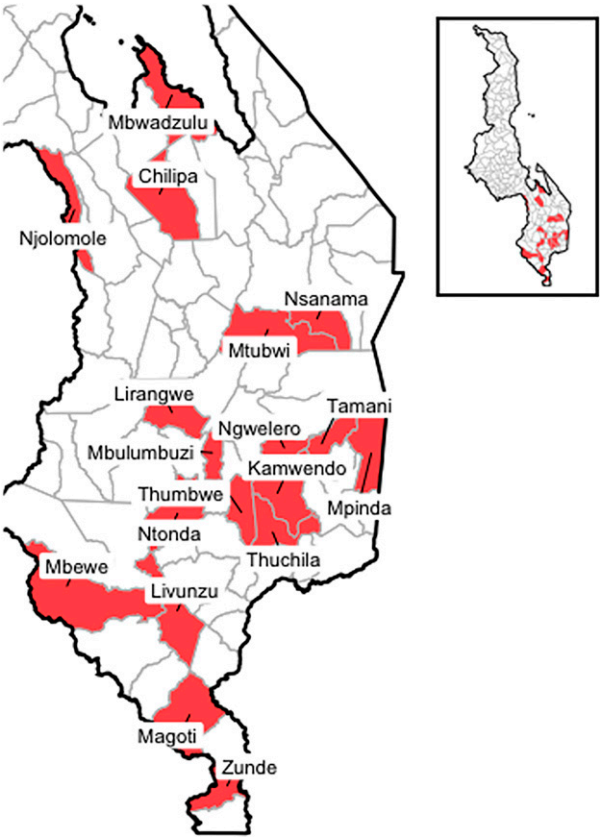

FIG. 4. Map of EPAs selected for this study (red).

(CHIRPS), and Tropical Applications of Meteorology Using Satellite and Ground Based Observations (TAMSAT); a satellite-derived measure of vegetation productivity: normalized vegetation difference index (NDVI); and two estimates of water availability: TAMSAT–Agricultural Early Warning System (ALERT) soil moisture and water stress requirement index (WRSI). These data sources were chosen to be representative of the types of data that are commonly used for drought insurance in the region. It is important to recognize that the choice of sources is subjective and will affect the resulting distribution—for example, NDVI measures vegetation cover, a quantity that may be only partially related to the other sources considered here, which are derived from rainfall. As such, appendix D presents robustness checks that omit NDVI from the analysis.

ARC, version 2 (ARC2), combines satellite infrared data from the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and gauge observations from the Global Telecommunication System to provide daily rainfall accumulations over Africa at 0.1° resolution (Novella and Thiaw 2013).

The CHIRPS dataset is a quasi-global rainfall dataset incorporating the CHPclim climatology, high-resolution satellite imagery and in situ station data to create a daily 0.05° gridded rainfall time series from 1981 to the present (Funk et al. 2015).

The TAMSAT rainfall dataset is based on high-resolution Meteosat thermal-infrared observations for all of Africa, available from 1983 to the present and updated in near–real time, providing daily rainfall estimates for all of Africa at a 4-km resolution (0.0375°) (Maidment et al. 2017).

Historic measures of NDVI, which measures vegetation greenness, were obtained from the Global Inventory Monitoring and Modeling System (GIMMS) project's NDVI. Specifically, we used the NDVI3g.v1 dataset, which extends from July 1981 to December 2015 and provides the bimonthly NDVI measures at an 8-km resolution, globally (Pinzon and Tucker 2014).

With regard to soil moisture and WRSI, TAMSAT-ALERT uses historical meteorological variables (e.g., TAMSAT rainfall and NCEP wind, temperature, and relative humidity) to drive a land surface model to estimate soil moisture and WRSI. Both of these variables are derived from an agroeconomic model that relates meteorological measurements to the growing conditions of a crop—in this case, maize. These models may capture conditions in a way that cumulative rainfall alone does not, hence their inclusion. The historic dataset spans all of Africa from 1983 to the present and provides daily estimates at a 0.25° resolution (Asfaw et al. 2018; Boult et al. 2020).

All biophysical indicators were spatially averaged over the EPA and temporally summed over the tasseling period (from December through January of the following year) to generate one measurement per year and EPA. The tasseling period was chosen because it is one of the times during which variation in weather can have the greatest impact on crop performance and is typical of the cover period for index insurance; however, unpacking the implications of this choice on potential basis risk is one of the goals of this analysis.

The farmer survey data come from community focus-group discussions, one per EPA. Participants were asked to rank the top 8 worst drought years of the last 35 years (1984–2017), identifying whether each was an early, mid-, or late-season drought event (not mutually exclusive). See appendix B for full details on the focus-group methodology. This survey methodology has been developed by authors Osgood and Diro (among others) as a way to reliably elicit accurate recall of historical crop failures from farmers; for more details, see Brahm et al. (2019) and Osgood et al. (2018).

c. *Model specification*

To incorporate these data into the MaxEnt framework, we first transformed each data source into an ordinal ranking (1 being the worst, and 0 being the best). For the purposes of this step of the calculation, years with missing data—that is, those not mentioned by farmers in their top 8 drought years—were treated as missing values; that is, a uniform prior for that year and location.

The biophysical indicators were used as the observational data for this model, and the farmer surveys were used as the prior data. Values chosen for alpha, beta, and gamma were 10^{-5}, 20, and 2, respectively. These values were chosen on the basis of providing a balance between goodness of fit and model solvability when tested on simulated data from a

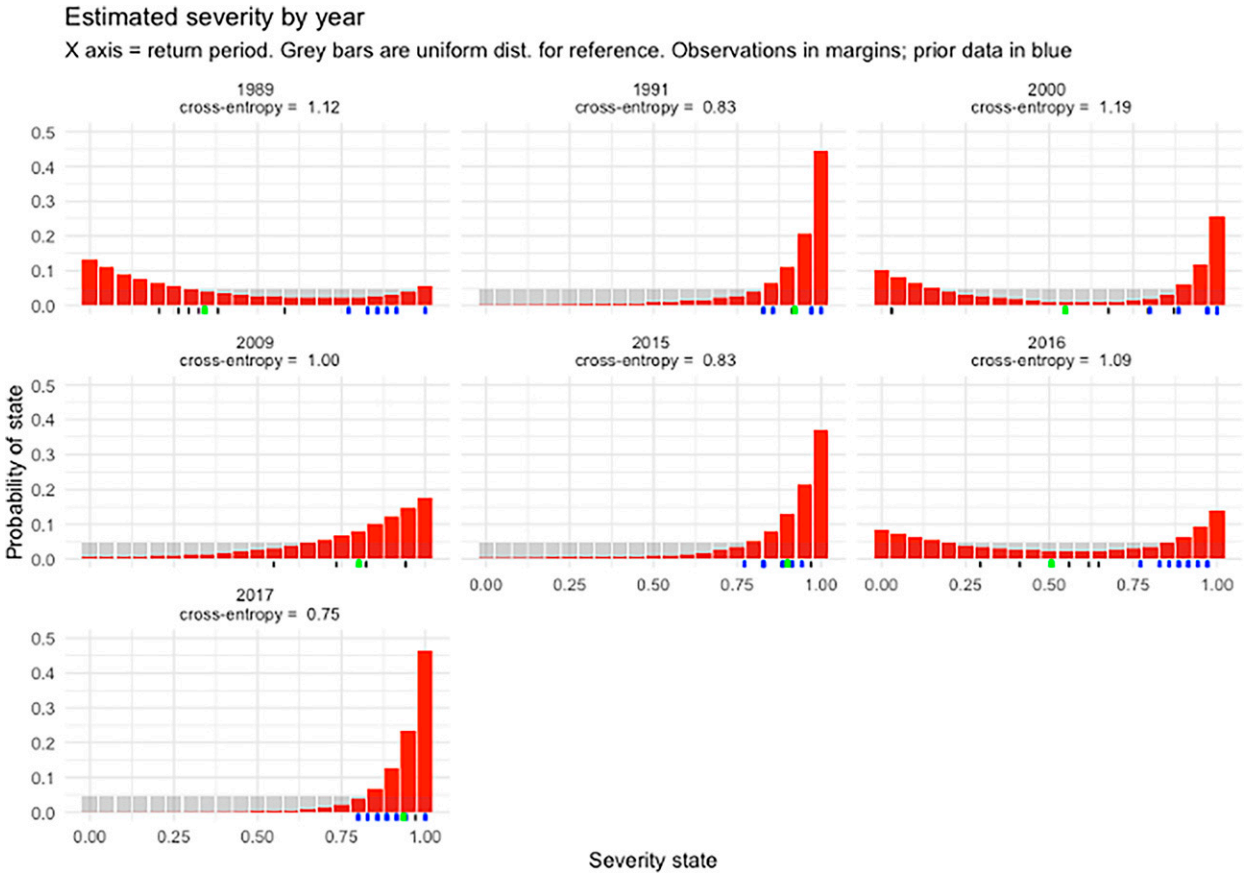

FIG. 5. MaxEnt results for Malawi SatWIN-ALERT data, for select years. (The uniform distribution is shown in gray as reference.)

known data-generating process (taken from Faynzilberg 1996b). Appendix D presents robustness checks to other choices of these values, as well as to omission of NDVI or WRSI from the distribution.

The biophysical indicators exhibited relatively little spatial variation between EPAs in comparative historical severity (see appendix C for details), and thus were averaged over all EPAs before calculation of the MaxEnt distribution. However, the farmer surveys had a large degree of variation between EPAs, which may be due to differences in farmers' cropping practices, their social vulnerability, and other important factors that may affect the relationship between biophysical conditions and the human impacts of drought. To preserve this heterogeneity in the final results, separate prior distributions were first estimated for each EPA separately and then averaged together.

Years that were not mentioned in the farmer focus groups were treated as uniformly distributed for the purposes of calculating the prior—in other words, the prior is completely uninformative for those years.

The MaxEnt results shown in the following sections came from the full specification [Eqs. (6) and (7)]—including the ML, prior, and error term components.

d. Results

1) OVERALL DISTRIBUTION

Figure 5 shows the resulting probability distribution for the comparative severity of selected harvest years (e.g., the 2016/17 season is shown in the graph for "2017"). A graph of the full data from 1984 to 2017 is available in appendix D.

We can see that some years appear as significantly below average in most or all datasets, including both the biophysical data (black ticks in the margins of the graphs) and farmer data (blue ticks)—for instance, 2017, 2015, and 1991. Those years have an NCE of <1, indicating a convergence of evidence.

Other years, such as 2009, are identified as moderate droughts in the biophysical data but were not mentioned by farmers. These years have an NCE of ~ 1, indicating that the prior contributed no evidence on such cases.

A third category of years, such as 2016, 2000, and 1989, exhibited a divergence between the biophysical and farmer data, with the former identifying them as good to average years and the latter identifying them as droughts. These years have an NCE of >1, indicating a disagreement between the biophysical data and the prior information on drought impact coming from farmers.

TABLE 3. Summary of findings, including identification of divergent years, indicated with boldface type. Years are referred to by the calendar year of harvest, e.g., "1984" is the 1983/84 agricultural season.

Year	Expected return period (full distribution); 1 = driest year	Expected return period (farmers only); 1 = driest year	Cross entropy
1984	0.22	0.54	1.03
1985	**0.14**	**0.59**	**1.11**
1986	0.55	0.57	1
1987	0.73	0.50	1
1988	0.69	0.51	0.99
1989	**0.36**	**0.64**	**1.12**
1990	0.65	0.59	0.96
1991	0.90	0.60	0.83
1992	0.34	0.53	1.02
1993	0.87	0.62	0.84
1994	0.76	0.55	0.95
1995	0.43	0.53	1.01
1996	0.24	0.53	1.03
1997	0.29	0.54	1.03
1998	**0.27**	**0.62**	**1.13**
1999	0.78	0.65	0.87
2000	**0.57**	**0.78**	**1.19**
2001	0.51	0.56	1
2002	0.54	0.55	1
2003	0.69	0.59	0.96
2004	0.46	0.55	1.01
2005	0.35	0.53	1.02
2006	0.25	0.53	1.03
2007	0.26	0.51	1.01
2008	0.34	0.58	1.06
2009	0.78	0.50	1
2010	0.43	0.56	1.02
2011	0.52	0.57	1.01
2012	0.52	0.51	1
2013	0.58	0.66	1.03
2014	**0.37**	**0.68**	**1.19**
2015	0.88	0.64	0.83
2016	**0.52**	**0.69**	**1.1**
2017	0.93	0.79	0.75

2) CATALOG OF DIVERGENT YEARS

Drawing from the results above, we can create a catalog of years for which farmers' consensus pointed toward a drought while the biophysical indicators did not. We define these so-called divergent years as years in which the normalized cross entropy is > 1.1 (i.e., where the biophysical observations and the prior information from farmers actively disagree). This cutoff is arbitrary, and we chose it simply to highlight some of the most important years in this case study—more generally, the MaxEnt method allows for a continuous measurement of degree of agreement via the cross-entropy value.

In these years, we observe a bimodal distribution in the MaxEnt results (Fig. 5), with the observations pointing in one direction and the priors in another. These years would be probable basis risk events in an index insurance application based on the chosen datasets. Table 3 summarizes those findings, with the divergent years highlighted in boldface type.

Note that we find the rank order of years and the identified divergent years to be robust to different model parameter choices as well as the omission of NDVI or WRSI—see appendix D for full robustness checks.

There is a meaningful policy distinction between years where satellite and farmer data strongly disagree, and years where the information from either source is inconclusive. The MaxEnt method allows us to distinguish between these two types of years—divergent years and inconclusive years—in terms of their information content, without the need to rely on a binary threshold to define "drought." In our MaxEnt results, the former type tend to look like bimodal distributions with an NCE of >1, and the latter look more like uniform or close to uniform distributions with an NCE of ~1. We only include the former type of years in the following discussion.

In Malawi, divergent years include 1985, 1989, 1998, 2000, 2014, and 2016. For each of these years, we can draw on expert discussion, secondary data sources, and temporal breakdowns of the sensor data to assess what might have gone wrong. Divergent years could either be basis risk events—that is, farmers experienced losses from drought, but the biophysical indicators registered the year as good or average, perhaps due to mistimed or inaccurate observations—or a year that farmers mistakenly identified as drought due to errors in recall, such as misidentifying years in the distant past or recalling losses due to events other than drought. To unpack what happened during these years, we can draw on more detailed data on seasonal progression, as well as news reports and local expertise. The following are some examples:

The year of 2016 was known as a year for which some risk models failed to capture the extent of El Niño–induced drought in southern Malawi—particularly those that made strong assumptions about seasonal timing, like ARC (African Risk Capacity 2016). While the biophysical data used for this analysis, as well as the water stress models commonly used by humanitarian agencies in Malawi, focused on the tasseling period (December–January), the 2016 drought did not occur until February—this can be observed in below-average February rainfall and February–March vegetation cover but not in the model of soil moisture (Fig. 6).

Similarly, 2014 was a year in which both tasseling period flooding and sowing period drought affected crops during the same season (United Nations 2015), meaning that the December–January rainfall signal did not clearly show the effect of drought, although the vegetation index did (Fig. 7).

In 2000/01, a major nondrought disruption to agricultural production occurred: the government of Malawi halted its input subsidy program, following a production surplus in the preceding year. This led many farmers to experience the subsequent drop in yields as a significant harm to their livelihoods (Denning et al. 2009).

The year of 1989 was a year that some remote sensing data sources—namely, NDVI and soil moisture—identified as a drought year while WRSI, ARC2, and CHIRPS did not. Similar to 2014, this may have been due to the earlier than usual onset of drought in this year, which showed up in December–January vegetation indices but not in rainfall over the same period (Fig. 8).

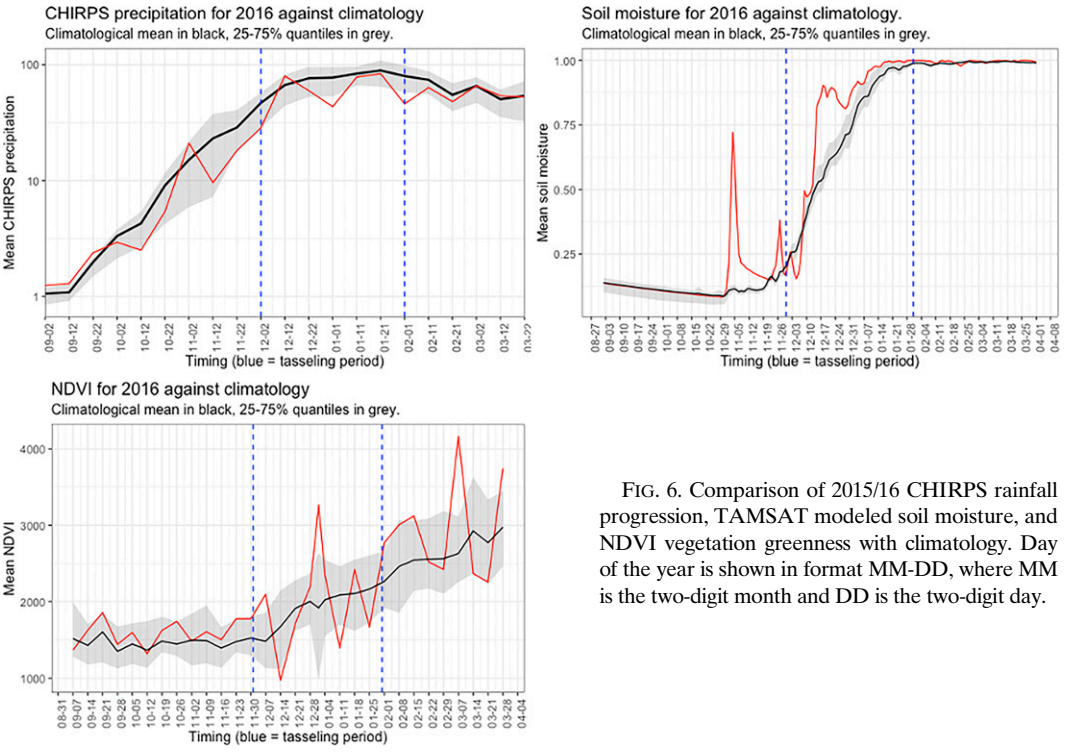

FIG. 6. Comparison of 2015/16 CHIRPS rainfall progression, TAMSAT modeled soil moisture, and NDVI vegetation greenness with climatology. Day of the year is shown in format MM-DD, where MM is the two-digit month and DD is the two-digit day.

Last, the divergence observed in 1985 may be a recall issue, as the previous year was known as a major drought year (Masih et al. 2014), and farmers may be misidentifying it as 1985 because of its distance in the past.

This method builds on previous evaluations of satellite drought product reliability such as Brahm et al. (2019) that rely on a binary definition of drought versus not-drought years and the comparison of one pair of data sources at a

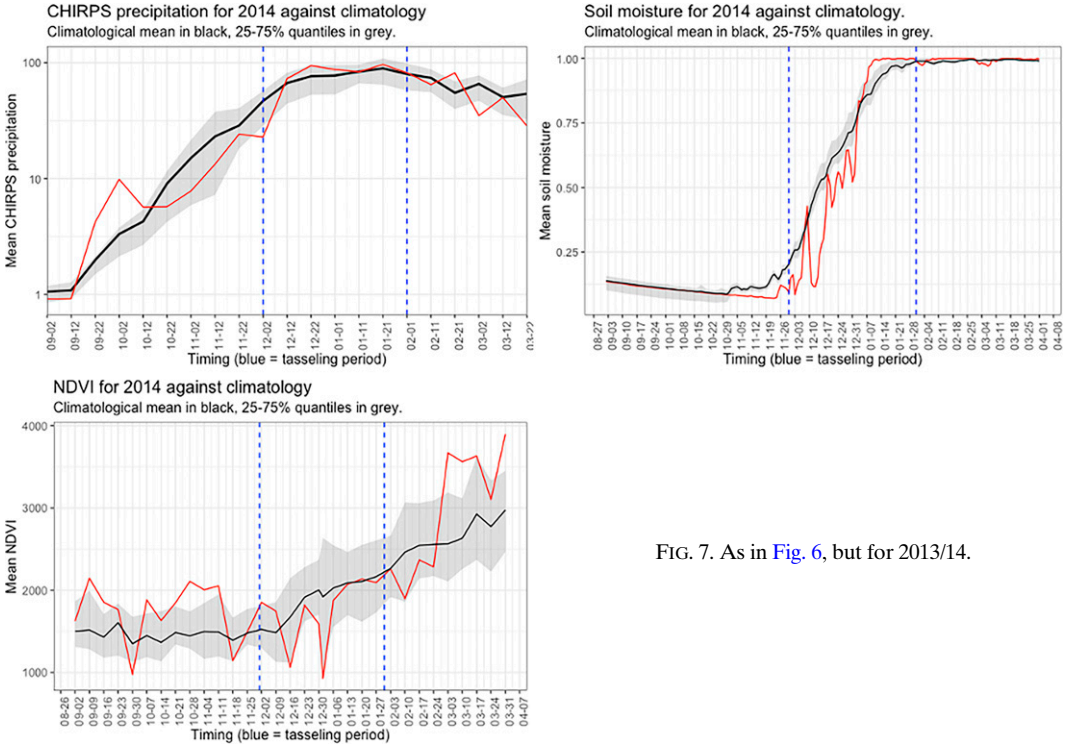

FIG. 7. As in Fig. 6, but for 2013/14.

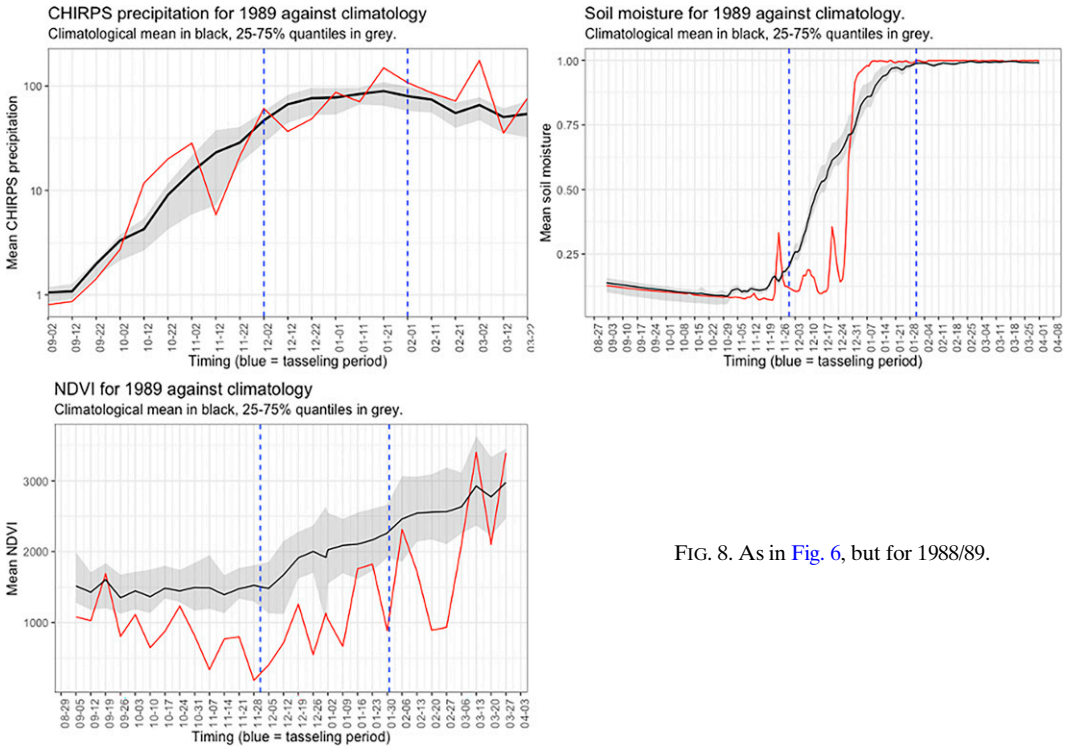

FIG. 8. As in Fig. 6, but for 1988/89.

time. Our method extends those discrete comparisons into a continuous distribution, freeing us from the need to define an arbitrary threshold for drought and allowing us to compare the joint information from many indicators at once. This allows us to easily identify divergent patterns in the data that may not be apparent from individual comparisons, using a combination of simple metrics (such as cross entropy) and visual intuition (such as where the distribution appears bimodal). Looking forward, this probability distribution could be used to evaluate policy questions like how reliable a given satellite would be as a trigger for early action, and at which threshold is early action most reliable.

4. Conclusions

This paper offers a simple statistical method for synthesizing many indicators of historical weather hazard and relating them to human impact, addressing a commonly expressed need from climate risk management practitioners. Among practitioners, this kind of analysis has conventionally taken the form of "counting on one's fingers," given the limited availability of benchmark data—for instance, if rain gauges, satellite precipitation estimates and yield data all suggest that a year is a 1-in-5-yr drought or worse, a humanitarian analyst might feel confident in deeming the year bad. The MaxEnt method moves from those discrete comparisons for a fixed return period to a continuous probability measure over many return periods, maintaining the intuitive interpretation of describing "consensus" among disparate data sources but allowing for much more sophisticated and flexible analysis. While the case study presented here focuses on drought insurance, the method could be applied to any use case for which a policy maker must determine a quantitative threshold for action on the basis of uncertain historical information about weather hazard and impact.

The MaxEnt method is better suited to such inductive reasoning than conventional statistical methods, which tend to require strong parametric observations and/or a large number of observations for consistent estimation and cannot easily accommodate partial or missing data (Golan 2017). Likewise, the MaxEnt method is richer than purely descriptive methods of data presentation like histograms or kernel density estimators, allowing for the incorporation of prior knowledge and the quantitative analysis of information content (Hullman and Gelman 2021).

The major limitation of this method is that it only accounts for ordinal differences between years, throwing out information on, for example, the absolute difference in mm of rainfall between years. However, in doing so, it allows both structured (like annual biophysical measurements) and semistructured (like farmers' partial ranking of the worst harvest years) data to be made commensurate. Moreover, most humanitarian programs are concerned primarily with the ordinal comparison of years and not their absolute differences, as triggers for humanitarian action are typically defined in reference to some frequency of action (like triggering during a 1-in-5-yr drought or worse) (Enenkel et al. 2020). That said, it is important to stress that a focus on ordinal comparison necessarily limits the strength of the claims we can make from the method, particularly when the historical record is short.

FIG. B1. Example of focus-group flip-chart output.

The SatWIN-ALERT case study presents one of many potential applications of the method: the descriptive analysis of years to identify likely basis risk events. Another promising application of the MaxEnt method is to use the probability distribution as a benchmark for evaluating the reliability of a decision rule for humanitarian action. Different trigger data sources and/or different thresholds for action could be compared in terms of how likely they are to coincide with the consensus from many data sources (i.e., the MaxEnt distribution). For example, practitioners could use this method to learn which period of the season would provide the most reliable trigger for drought monitoring, how severe a drought must be in order to be consistently identifiable via satellite observation, or how likely ex gratia payments to compensate for basis risk are expected to be necessary, by expressing these questions in a form that considers both intersensor consistency and those sensors' relationship to human impact.

Besides observational data, another potential application of the method is to forecast-based early action. Forecast users often face the problem of how to summarize the predictions of an ensemble of forecasts (or a number of different forecast products) in a way that is concise and easily interpretable (Vigaud et al. 2018; Wilks and Hamill 2007). The members of an ensemble, and/or the results of multiple forecast products, could be treated like any other data source, using the MaxEnt method to summarize consensus among indicators in each year, including the future. This addresses an expressed need from practitioners for more flexible, transparent probabilistic forecasts (Coughlan de Perez et al. 2015).

In sum, this paper presents a simple, extensible method for summarizing historical risk data with a minimum of assumptions

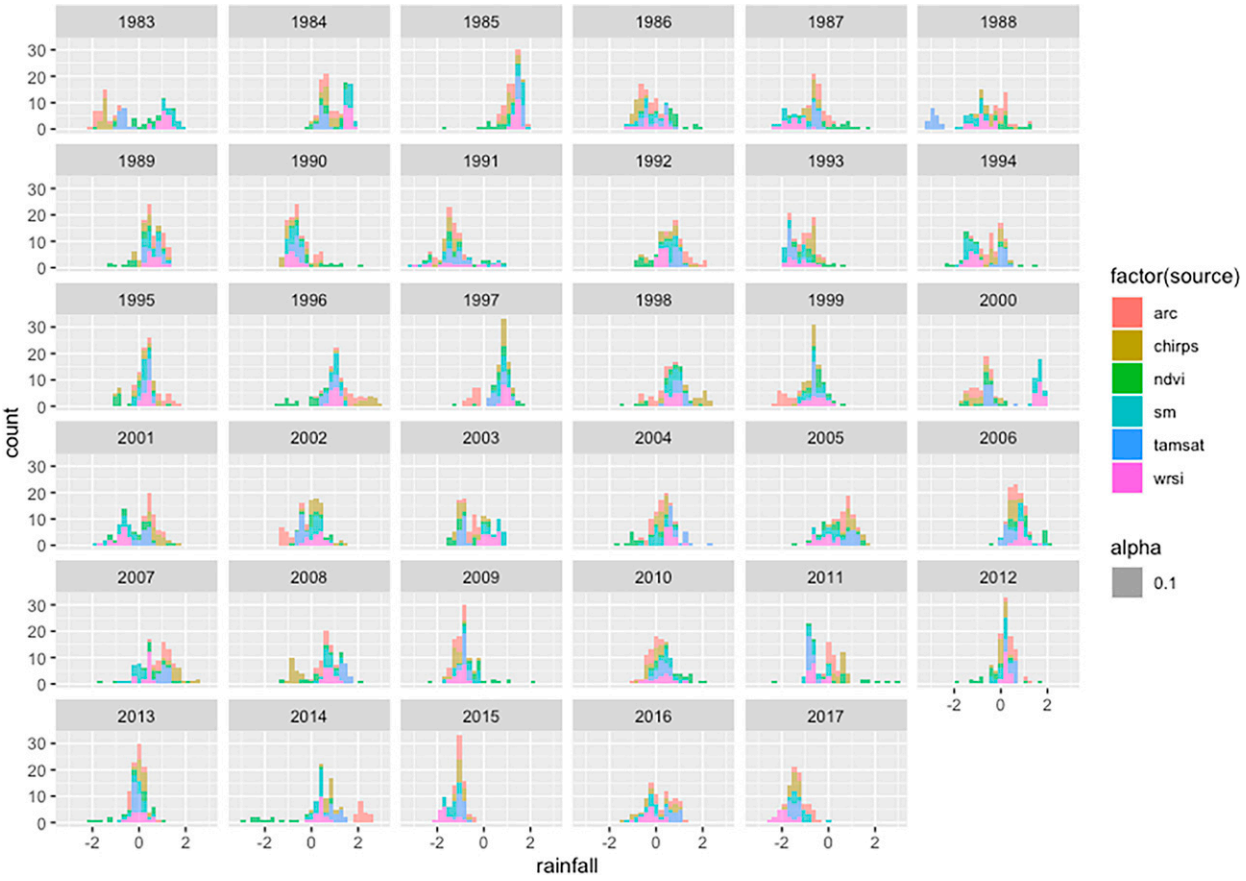

FIG. C1. Summary of spatial variation in biophysical variables by year, measured in standard deviations.

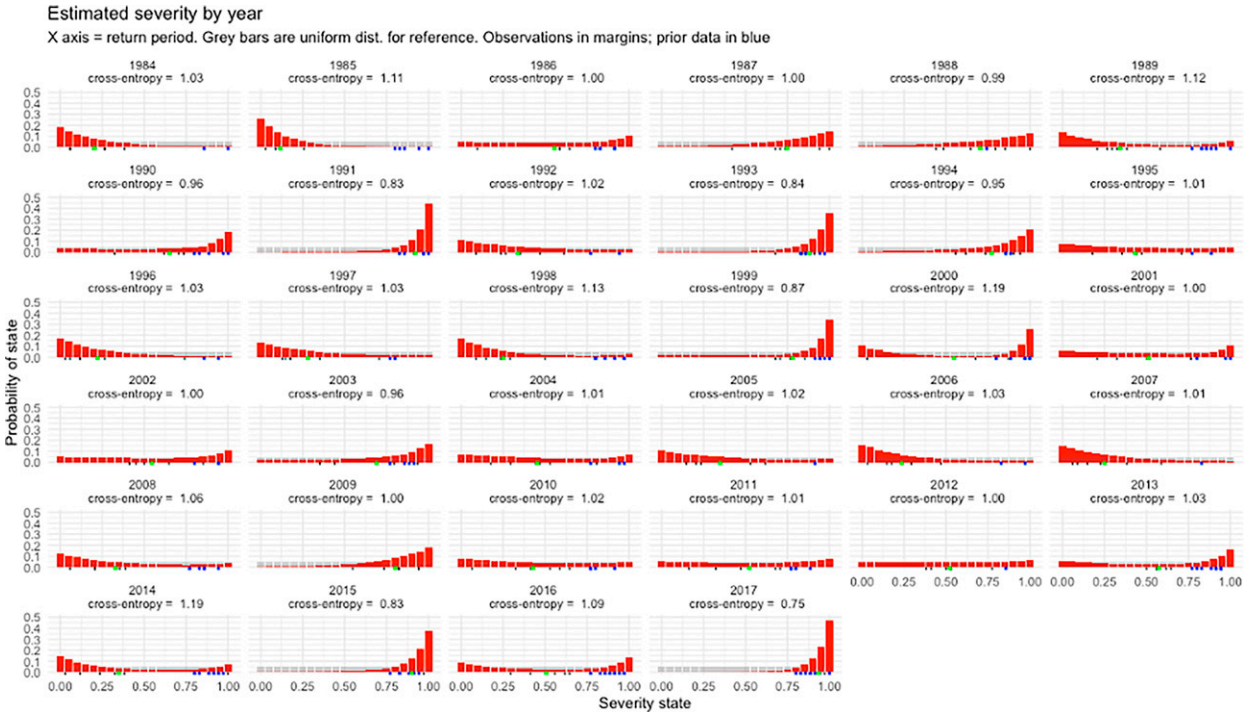

FIG. D1. MaxEnt estimated distribution for each harvest year in Malawi, 1984–2017.

about distribution and data completeness, and its results have a number of intuitive applications in climate risk analysis.

Acknowledgments. Support for this paper was provided by the NERC SatWIN-ALERT project (Grant NE/R014116/1). Field data collection was carried out by the Centre for Agricultural Research and Development (CARD) at the Lilongwe University of Agriculture and Natural Resources, led by Dr. Thabbie Chilongo.

Data availability statement. CHIRPS, ARC2, MODIS NDVI, and TAMSAT are available from their prospective providers' websites (listed in the references). The soil moisture and WRSI models are available upon request. Because of its proprietary nature, the farmer focus group cannot be made openly available. The R code to conduct the MaxEnt analysis is available online (https://github.com/mm5330/maximum-entropy).

APPENDIX A

Focus-Group Sampling Procedures

Step 1: Ranking districts for selection

We list down the districts based on their drought risk profile and number of food insecure households according to the Malawi Vulnerability Assessment Committee (MVAC).

Step 2: Drop districts with low level of food security

We drop all districts with that experienced food insecurity 1–3 times during the 10 years analyzed. These are Chitipa, Kasungu, Mchinji, and Nkhotakota.

Step 3: Ranking the remaining districts according to their severity of drought

We rank the remaining districts according to the drought return period criterion. Districts with 7–10 years of drought return period are classified as high frequency, and districts with drought return periods of 4–6 years are classified as medium frequency.

Step 4: Selection of districts

In this stage, first we rank the high frequency and medium frequency districts based on the cumulative number of food insecure households over the 10-year period. Then, the top six droughtprone districts from high frequency and medium frequency district are selected.

Step 5: Village selection

In each district, we will purposely choose two EPAs based on their remoteness and other criteria such as direction and number of food insecure households. From each EPA, we will randomly choose three sections. In each section, three villages will be randomly selected for household surveys. In total, there will be 18 villages for household surveys.

Step 6: Focus-group discussions (FGDs)

In each section, we will randomly select one village (excluding the three household survey villages) and conduct a focus-group discussion. Thus, a total of 6 FGDs will be conducted per district bringing the overall total of FGDs to 72. For each FGD, 10–15 villagers will be involved, ensuring that there is representation for youths, by gender, and, most important, to include some elderly people of 50 years and above who have stayed in the areas for longer periods to facilitate recall of drought events that happened in the area in the past 35 years.

TABLE D1. Relative severity of years under different parameter assumptions.

Year	Baseline	Alpha = 10^{-3}	Alpha = 10^{-7}	Beta = 10	Beta = 40	Gamma = 1	Gamma = 4	NDVI omitted	WRSI omitted
1984	0.22	0.22	0.22	0.24	0.21	0.22	0.23	0.23	0.25
1985	0.15	0.15	0.15	0.17	0.13	0.15	0.15	0.12	0.15
1986	0.56	0.56	0.56	0.56	0.56	0.56	0.57	0.65	0.55
1987	0.73	0.73	0.73	0.72	0.74	0.73	0.73	0.79	0.70
1988	0.69	0.69	0.69	0.68	0.70	0.69	0.69	0.73	0.68
1989	0.36	0.36	0.36	0.38	0.35	0.36	0.37	0.32	0.38
1990	0.65	0.65	0.65	0.66	0.65	0.65	0.67	0.72	0.64
1991	0.90	0.90	0.90	0.89	0.91	0.90	0.92	0.90	0.92
1992	0.35	0.35	0.35	0.36	0.34	0.35	0.35	0.30	0.32
1993	0.87	0.87	0.87	0.86	0.88	0.86	0.89	0.88	0.87
1994	0.76	0.76	0.76	0.76	0.77	0.76	0.77	0.74	0.74
1995	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.39	0.45
1996	0.24	0.24	0.24	0.26	0.23	0.24	0.24	0.15	0.26
1997	0.30	0.30	0.30	0.31	0.29	0.30	0.30	0.33	0.32
1998	0.27	0.27	0.27	0.29	0.26	0.27	0.28	0.26	0.28
1999	0.78	0.78	0.78	0.78	0.78	0.77	0.80	0.80	0.80
2000	0.57	0.57	0.57	0.59	0.56	0.56	0.60	0.51	0.67
2001	0.52	0.52	0.52	0.52	0.52	0.51	0.52	0.52	0.47
2002	0.55	0.55	0.55	0.55	0.55	0.55	0.56	0.57	0.56
2003	0.69	0.69	0.69	0.69	0.69	0.69	0.70	0.64	0.74
2004	0.46	0.46	0.46	0.47	0.46	0.46	0.47	0.39	0.49
2005	0.36	0.36	0.36	0.37	0.35	0.36	0.36	0.36	0.31
2006	0.25	0.25	0.25	0.27	0.25	0.25	0.26	0.29	0.26
2007	0.26	0.26	0.26	0.28	0.26	0.26	0.27	0.27	0.24
2008	0.35	0.35	0.35	0.36	0.34	0.34	0.35	0.33	0.37
2009	0.78	0.78	0.78	0.77	0.79	0.78	0.78	0.83	0.78
2010	0.44	0.44	0.44	0.44	0.43	0.43	0.44	0.51	0.44
2011	0.53	0.53	0.53	0.53	0.53	0.52	0.54	0.60	0.49
2012	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.47	0.52
2013	0.59	0.59	0.59	0.60	0.58	0.58	0.60	0.57	0.59
2014	0.37	0.37	0.37	0.39	0.36	0.36	0.39	0.25	0.38
2015	0.89	0.89	0.89	0.88	0.89	0.88	0.91	0.89	0.87
2016	0.52	0.52	0.52	0.54	0.52	0.51	0.54	0.52	0.49
2017	0.93	0.93	0.93	0.93	0.94	0.92	0.96	0.93	0.92

APPENDIX B

Focus-Group Methodology

For this exercise farmers will identify the worst drought years and they will discuss the reasons why they were the worst. During this exercise, farmers will name the worst 8 years (that they can remember) in the past 31 years for their main crop. Facilitators should conduct the exercise as follows and THEN record the results from the exercise into the Kobo form.

1) Supplies needed

- Chalkboard/dry-erase board/flip chart
- Markers/chalk to write on the board
- Several different types/colors of Post-its
- Paper and pen for the facilitators to record the results of the exercise

Note: if the facilitators cannot use the suggested supplies in the field, some of these items can be replaced by more basic objects. As an example, the years might be represented by 15 boxes placed on the ground and the Post-its can be replaced by beans or leaves of different colors.

2) Focus-group instructions

- Assign participants in groups of 4–5 people. At least one person in each group should be able to read/write (can be a farmer or an assistant). Provide each group with several Post-its of the same color (or other objects suitable to use in the field). Each group should have 3–6 Post-its.
- Ask each group to select a representative who would be responsible to place the Post-its in the appropriate place on the board once the group reaches consensus on the worst 8 years.
- Discuss with the farmers and the participants to identify which part of the season caused the loss in production: early, late, or both. It is important to know which months the farmers consider part of the "early" season and which months correspond to the "late" season.
- Explain to each group that this is an interactive exercise in which participants in each group must discuss which years were the worst, based on their experience and memory. Ask each group to discuss and identify the 8 worst years. The eight worst years will be represented by the Post-its each group receives (or other objects that groups receive). The groups should rank the worst years

TABLE D2. Cross entropy of years under different parameter assumptions.

Year	Baseline	Alpha = 10^{-3}	Alpha = 10^{-7}	Beta = 10	Beta = 40	Gamma = 1	Gamma = 4	NDVI omitted	WRSI omitted
1984	1.03	1.03	1.03	1.03	1.04	1.03	1.03	1.03	1.03
1985	1.11	1.11	1.11	1.1	1.12	1.11	1.1	1.11	1.11
1986	1	1	1	1	1	1	0.98	0.97	1
1987	1	1	1	1	1	1	1	1	1
1988	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
1989	1.12	1.12	1.12	1.11	1.13	1.14	1.09	1.14	1.12
1990	0.96	0.96	0.96	0.97	0.96	0.98	0.93	0.94	0.97
1991	0.83	0.83	0.83	0.85	0.8	0.84	0.79	0.83	0.82
1992	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
1993	0.84	0.84	0.84	0.86	0.83	0.86	0.81	0.84	0.85
1994	0.95	0.95	0.95	0.95	0.94	0.95	0.94	0.95	0.95
1995	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
1996	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.04	1.03
1997	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
1998	1.13	1.13	1.13	1.12	1.14	1.14	1.11	1.14	1.13
1999	0.87	0.87	0.87	0.89	0.86	0.92	0.81	0.86	0.86
2000	1.19	1.19	1.19	1.15	1.22	1.3	1.05	1.26	1.06
2001	1	1	1	1	1	1.01	0.98	1	1.01
2002	1	1	1	1	1	1	0.98	0.99	0.99
2003	0.96	0.96	0.96	0.96	0.96	0.97	0.94	0.98	0.94
2004	1.01	1.01	1.01	1.01	1.01	1.02	1	1.03	1.01
2005	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03
2006	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
2007	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
2008	1.06	1.06	1.06	1.06	1.07	1.07	1.05	1.06	1.06
2009	1	1	1	1	1	1	1	1	1
2010	1.02	1.02	1.02	1.02	1.03	1.03	1.02	1.01	1.02
2011	1.01	1.01	1.01	1.01	1.01	1.01	1	0.99	1.02
2012	1	1	1	1	1	1	1	1	1
2013	1.03	1.03	1.03	1.02	1.03	1.06	0.98	1.04	1.03
2014	1.19	1.19	1.19	1.16	1.21	1.22	1.14	1.26	1.19
2015	0.83	0.83	0.83	0.85	0.82	0.84	0.8	0.83	0.84
2016	1.09	1.09	1.09	1.08	1.1	1.13	1.03	1.09	1.12
2017	0.75	0.75	0.75	0.77	0.74	0.79	0.68	0.75	0.77

(with the worst year being 1). Remember to designate the elected group representative to lead the discussion.

- Give the groups 30 min to decide on the worst 8 years.
- Now, ask the group representative to walk to the chart and place the Post-its next to the worst years decided by his/her group (Fig. B1).
- After all groups have placed all their Post-its on the flipchart, compare the years reported by the different groups. If there are differences, give the participants a chance to discuss in one big group until they reach agreement on 8 common worst years going back to 1983.
- Groups will also have to specify whether the rainfall was particularly low at the start or at the end of the season. The facilitator can give an example like, "If you had the worst year for your main crop because of a drought, was the precipitation lower than usual in the early or the late part of the rainy season?"
- From all of the years that the farmers listed, select the years for which there was strong disagreement between the farmer groups, and discuss the effects of the worst years until consensus is reached.

APPENDIX C

Spatial Variation in Biophysical Measures

In this appendix, a summary of spatial variation in biophysical variables by year is presented (Fig. C1). The figure shows that biophysical indicators exhibited relatively little spatial variation between EPAs in comparative historical severity. This result was used to justify averaging of these variables over all EPAs before subsequent calculation of the MaxEnt distribution.

APPENDIX D

Full Results and Robustness Checks

Figure D1 shows the MaxEnt estimated distribution for all harvest years from 1984 to –2017, expanding the results shown in section 3d(1). Note that the choice of data sources is subjective and will affect the resulting distribution. This appendix reports the results of a few sensitivity analyses in which datasets and observations were removed to examine the effect on the results. One robustness check is to remove

NDI from the analysis, and another is to remove WRSI from the distribution. Other robustness checks looked at the effects of choosing values of alpha, beta, and gamma that differ from the those ultimately selected for final use. Under these different parameter assumptions, we examined the relative severity (Table D1) and cross entropy (Table D2) for each year from 1984 to 2017.

REFERENCES

African Risk Capacity, 2016: Malawi 2015/16. Africa RiskView Special Rep., 74 pp.

Asfaw, D., and Coauthors, 2018: TAMSAT-ALERT v1: A new framework for agricultural decision support. *Geosci. Model Dev.*, **11**, 2353–2371, https://doi.org/10.5194/gmd-11-2353-2018.

Benami, E., Z. Jin, M. R. Carter, A. Ghosh, R. J. Hijmans, A. Hobbs, B. Kenduiywo, and D. B. Lobell, 2021: Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nat. Rev. Earth Environ.*, **2**, 140–159, https://doi.org/10.1038/s43017-020-00122-y.

Boult, V. L., and Coauthors, 2020: Evaluation and validation of TAMSAT-ALERT soil moisture and WRSI for use in drought anticipatory action. *Meteor. Appl.*, **27**, e1959, https://doi.org/10.1002/met.1959.

Brahm, M., D. Vila, S. M. Saenz, and D. Osgood, 2019: Can disaster events reporting be used to drive remote sensing applications? A Latin America weather index insurance case study. *Meteor. Appl.*, **26**, 632–641, https://doi.org/10.1002/met.1790.

Bucheli, J., T. Dalhaus, and R. Finger, 2021: The optimal drought index for designing weather index insurance. *Eur. Rev. Agric. Econ.*, **48**, 573–597, https://doi.org/10.1093/erae/jbaa014.

Camenisch, C., F. Jaume-Santero, S. White, Q. Pei, R. Hand, C. Rohr, and S. Brönnimann, 2022: A Bayesian approach to historical climatology for the Burgundian Low Countries in the 15th century. *Climate Past.*, **18**, 2449–2462, https://doi.org/10.5194/cp-18-2449-2022.

Coughlan de Perez, E., B. van den Hurk, M. K. van Aalst, B. Jongman, T. Klose, and P. Suarez, 2015: Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Nat. Hazards Earth Syst. Sci.*, **15**, 895–904, https://doi.org/10.5194/nhess-15-895-2015.

Denning, G., and Coauthors, 2009: Input subsidies to improve smallholder maize productivity in Malawi: Toward an African green revolution. *PLOS Biol.*, **7**, e1000023, https://doi.org/10.1371/journal.pbio.1000023.

Enenkel, M., and Coauthors, 2019: Exploiting the convergence of evidence in satellite data for advanced weather index insurance design. *Wea. Climate Soc.*, **11**, 65–93, https://doi.org/10.1175/WCAS-D-17-0111.1.

——, and Coauthors, 2020: Why predict climate hazards if we need to understand impacts? Putting humans back into the drought equation. *Climatic Change*, **162**, 1161–1176, https://doi.org/10.1007/s10584-020-02878-0.

Faynzilberg, P. S., 1996a: Meal estimation: Acceptable-likelihood extensions of maxent. *Maximum Entropy and Bayesian Methods*, K. M. Hanson and R. N. Silver, Eds., Fundamental Theories of Physics, Vol. 79, Springer, 387–392, https://doi.org/10.1007/978-94-011-5430-7_49.

——, 1996b: Statistical mechanics of choice: MaxEnt estimation of population heterogeneity. *Ann. Oper. Res.*, **68**, 161–180, https://doi.org/10.1007/BF02205453.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, https://doi.org/10.1038/sdata.2015.66.

Golan, A., 2017: *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*. Oxford University Press, 488 pp., https://doi.org/10.1093/oso/9780199349524.001.0001.

Harte, J., 2011: *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*. Oxford University Press, 280 pp., https://doi.org/10.1093/acprof:oso/9780199593415.001.0001.

Hochrainer, S., R. Mechler, and G. Pflug, 2009: Climate change and financial adaptation in Africa. Investigating the impact of climate change on the robustness of index-based microinsurance in Malawi. *Mitigation Adapt. Strategies Global Change*, **14**, 231–250, https://doi.org/10.1007/s11027-008-9162-5.

Hullman, J., and A. Gelman, 2021: Designing for interactive exploratory data analysis requires theories of graphical inference. *Harv. Data Sci. Rev.*, **3**, https://doi.org/10.1162/99608f92.3ab8a587.

Karduni, A., D. Markant, R. Wesslen, and W. Dou, 2021: A Bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations. *IEEE Trans. Visualization Comput. Graphics*, **27**, 978–988, https://doi.org/10.1109/TVCG.2020.3029412.

Kim, Y.-S., L. A. Walls, P. Krafft, and J. Hullman, 2019: A bayesian cognition approach to improve data visualization. *Proc. CHI Conf. on Human Factors in Computing Systems*, New York, NY, Association for Computing Machinery, 1–14, https://doi.org/10.1145/3290605.3300912.

Läderach, P., J. Ramirez-Villegas, C. Navarro-Racines, C. Zelaya, A. Martinez-Valle, and A. Jarvis, 2017: Climate change adaptation of coffee production in space and time. *Climatic Change*, **141**, 47–62, https://doi.org/10.1007/s10584-016-1788-9.

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray, 2020: Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis. *Amer. J. Agric. Econ.*, **102**, 202–219, https://doi.org/10.1093/ajae/aaz051.

Maganga, A. M., L. S. Chiwaula, and P. Kambewa, 2021: Parametric and non-parametric estimates of willingness to pay for weather index insurance in Malawi. *Int. J. Disaster Risk Reduct.*, **62**, 102406, https://doi.org/10.1016/j.ijdrr.2021.102406.

Maidment, R. I., and Coauthors, 2017: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Sci. Data*, **4**, 170063, https://doi.org/10.1038/sdata.2017.63.

Masih, I., S. Maskey, F. E. F. Mussá, and P. Trambauer, 2014: A review of droughts on the African continent: A geospatial and long-term perspective. *Hydrol. Earth Syst. Sci.*, **18**, 3635–3649, https://doi.org/10.5194/hess-18-3635-2014.

Meuwissen, M. P. M., and Coauthors, 2019: A framework to assess the resilience of farming systems. *Agric. Syst.*, **176**, 102656, https://doi.org/10.1016/j.agsy.2019.102656.

Nicholson, S. E., 2001: Climatic and environmental change in Africa during the last two centuries. *Climate Res.*, **17**, 123–144, https://doi.org/10.3354/cr017123.

Novella, N. S., and W. M. Thiaw, 2013: African Rainfall Climatology version 2 for famine early warning systems. *J. Appl. Meteor. Climatol.*, **52**, 588–606, https://doi.org/10.1175/JAMC-D-11-0238.1.

Ortiz-Bobea, A., E. Knippenberg, and R. G. Chambers, 2018: Growing climatic sensitivity of U.S. agriculture linked to technological change and regional specialization. *Sci. Adv.*, **4**, eaat4343, https://doi.org/10.1126/sciadv.aat4343.

Osgood, D., and K. E. Shirley, 2012: The value of information in index insurance for farmers in Africa. *The Value of Information: Methodological Frontiers and New Applications in Environment and Health*, R. Laxminarayan and M. K. Macauley, Eds., Springer, 1–18, https://doi.org/10.1007/978-94-007-4839-2_1.

——, and Coauthors, 2018: Farmer perception, recollection, and remote sensing in weather index insurance: An Ethiopia case study. *Remote Sens.*, **10**, 1887, https://doi.org/10.3390/rs10121887.

Pinzon, J. E., and C. J. Tucker, 2014: A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sens.*, **6**, 6929–6960, https://doi.org/10.3390/rs6086929.

Schlenker, W., and M. J. Roberts, 2009: Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc. Natl. Acad. Sci. USA*, **106**, 15 594–15 598, https://doi.org/10.1073/pnas.0906865106.

Shannon, C. E., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Svensson, C., J. Hannaford, and I. Prosdocimi, 2017: Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications. *Water Resour. Res.*, **53**, 999–1018, https://doi.org/10.1002/2016WR019276.

United Nations, 2015: Major floods in Malawi, Mozambique force thousands to flee their homes—UN. *UN News*, accessed 4 October 2021, https://news.un.org/en/story/2015/01/488442.

Vigaud, N., M. K. Tippett, and A. W. Robertson, 2018: Probabilistic skill of subseasonal precipitation forecasts for the East Africa–West Asia sector during September–May. *Wea. Forecasting*, **33**, 1513–1532, https://doi.org/10.1175/WAF-D-18-0074.1.

Wickham, H., D. Cook, H. Hofmann, and A. Buja, 2010: Graphical inference for infovis. *IEEE Trans. Visualization Comput. Graphics*, **16**, 973–979, https://doi.org/10.1109/TVCG.2010.161.

Wilks, D. S., and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390, https://doi.org/10.1175/MWR3402.1.