Statistical Weather-Impact Models: An Application of Neural Networks and Mixed Effects for Corn Production over the United States

JORDANE A. MATHIEU AND FILIPE AIRES

Laboratoire d'Etudes du Rayonnement et de la Matière en Astrophysique et Atmospheres, and Estellus, Paris, France

(Manuscript received 26 January 2016, in final form 7 June 2016)

ABSTRACT

Statistical meteorological impact models are intended to represent the impact of weather on socioeconomic activities, using a statistical approach. The calibration of such models is difficult because relationships are complex and historical records are limited. Often, such models succeed in reproducing past data but perform poorly on unseen new data (a problem known as overfitting). This difficulty emphasizes the need for regularization techniques and reliable assessment of the model quality. This study illustrates, in a general way, how to extract pertinent information from weather data and exploit it in impact models that are designed to help decision-making. For a given socioeconomic activity, this type of impact model can be used to 1) study its sensitivity to weather anomalies (e.g., corn sensitivity to water stress), 2) perform seasonal forecasting (yield forecasting) for it, and 3) quantify the longer-term (several decades) impact of weather on it. The size of the training database can be increased by pooling data from various locations, but this requires statistical models that are able to use the localization information—for example, mixed-effect (ME) models. Linear, neural-network, and ME models are compared, using a real-world application: corn-yield forecasting over the United States. Many challenges faced in this paper may be encountered in many weather-impact analyses: these results show that much care is required when using space–time data because they are often highly spatially correlated. In addition, the forecast quality is strongly influenced by the training spatial scale. For the application that is described herein, learning at the state scale is a good trade-off: it is specific to local conditions while keeping enough data for the calibration.

1. Introduction

The impact of weather or climate on social and economic activities is important in many domains such as environment, agriculture, energy (Adams et al. 1998; Kaylen and Koroma 1991), logistics in sales, insurances against catastrophes, or tourism (Jewson and Brix 2005). The analysis of the weather/climate impact on human activities requires building models that describe the links and effects between weather and socioeconomic data. These models are referred to here as ''impact models'' (see, e.g., Aires 2012).

Impact models can be used (e.g., for decision-making) in two different frameworks. First, weather-related natural variability and hazards can directly impact some human activities over the short term. Impact models can then be used to avoid or mitigate weather-related risks in the present climate. For instance, weather variability has a direct influence on the management of activities such as agriculture or tourism (Marteau et al. 2004; Jewson and Brix 2005). Second, impact models can be used to better anticipate the long-term evolution resulting from climate change (Leckebusch et al. 2002; IPCC 2007; Rauber et al. 2005): changes in agricultural practices, higher risk related to hurricanes, or modification of energy consumption and production. For these types of applications, long-term impact models can be associated with climate forecasts (from global climate models) to estimate the consequences of climate change on human activities (Schimmelpfennig 1996). This type of study allows optimization of long-term investments like agricultural practices (Lewandrowski and Schimmelpfennig 1999), investments in wind facilities or hydraulic dams, or health policies (Greenough et al. 2001). It allows us to define adequate mitigation and adaptation strategies. This paper addresses the first type of application related to short-term weather-impact models.

Corresponding author address: J. A. Mathieu, Lerma, 77 av. Denfert-Rochereau, 75014 Paris, France.
E-mail: jordane.mathieu@obspm.fr

DOI: 10.1175/JAMC-D-16-0055.1

To demonstrate the construction of an impact model, we examine an application in agriculture, a field in which weather is important and relationships are complex. We focus here on the forecast of corn yields across the United States. Many of the challenges that are faced in this paper can be generalized to many weather-impact analyses.

Weather is the major uncontrollable factor influencing the development of crops. In the scientific literature, two major approaches have been used to study the relationship between weather/climate and crop yield: 1) crop-simulation models and 2) empirical regression models. The crop models are useful to simulate the effects of environmental characteristics on crop growth with high detail (Hoogenboom 2000; Jones et al. 2003). They are often site-specific, requiring extensive and local information such as soil properties. In contrast to this *model-driven* scheme, it is possible to adopt a *data-driven* approach (Kotlowski 2007). These empirical models are not only useful for forecasting crop yield; the exploration of the historical data can also improve our understanding of the processes.

Empirical regression models have been intensively used to study the impacts of climate change on crop yields. Kandiannan et al. (2002) adopted multiple-regression models to study the effects of weather on turmeric yield, using a 20-yr dataset. With a modified multiple-regression model, first developed by Thompson (1988), Tannura et al. (2008) studied the relationship between corn/soybean yields and monthly weather conditions during 1960–2006 for Illinois, Indiana, and Iowa. Using a regression model, Lobell et al. (2007) analyzed the relationship between crop yield and three climatic variables (i.e., minimum temperature, maximum temperature, and precipitation) for 12 major crops in California during 1980–2003.

Practical difficulties can occur when using regression models. A first difficulty comes from the fact that causal relationships can be corrupted by other factors. For instance, precipitation effects on crop yield can be largely damped by local irrigation, mainly in the western part of the United States. Therefore, it is not easy to distinguish between the influence of weather and the influence of other important factors such as irrigation, fertilization, mechanization, or soil type. Because the goal is to investigate only the weather impact on corn yield, it is often necessary to limit the impact of nonweather factors during the preprocessing stage (Kotlowski 2007).

Detection of these nonweather factors is complicated by spatial heterogeneity of the weather impact on crop yield. For instance, Schlenker and Roberts (2009) explored how the temperature–yield relationship varies across different U.S. regions and found that southern crops have lower sensitivity to extreme heat. Cai et al. (2014) attempted to compare the impacts of weather variations on crop yields across the U.S. regions and showed that the relationship between corn yield and weather has a large spatial variability. In this paper, we propose two ways to study the spatial heterogeneity of the weather-impact models:

1) An analysis of the advantages provided by linear mixed-effect (ME) models is done. In the United States, the 50 states are divided into about 300 districts, which are themselves divided into about 3000 counties. Each crop datum is referenced to a specific county, but data can be gathered by district or state. In this article, the term "group" will refer to a spatial classification (the different counties, districts, or states). The ME models are based on the idea that each piece of information belongs to a particular group, and the model accounts for the particularities of each one of these groups (Fahrmeir and Tutz 1994).
2) The spatial scale used by the impact model (county, district, state, or nation) is then optimized for various statistical models (linear, neural network, and ME) to analyze their advantages and limitations.

Another challenge comes from the size limitation of the datasets used to calibrate the statistical impact models (Aires 2012). In agriculture, one of the biggest challenges is the lack of crop data since a sample in the statistical dataset is associated with one full year. For example, a 20-yr record represents only 20 samples! The complexity of the weather–crop relationships requires as much information as possible in the inputs of the impact model, but not enough data are available to fit a complex model. This results in the "bias/variance dilemma" or overfitting (Geman et al. 1992). A model overfits when it works well on training data but has poor predictive performance on new data. For instance, overfitting occurs when a model is excessively complex, such as having too many parameters and not enough observations to calibrate them.

In this paper, an emphasis will be put on reliable crop-yield quality measures and on regularization strategies to limit these dataset-limitation issues. The importance of these considerations will be demonstrated with examples. The method used in this paper is very general and can by employed for other weather impacts on socioeconomic activities.

The paper is organized as follows: Section 2 presents the data sources and the impact models used in this study. After examining the experimental configuration, particular attention is given to the reliable assessment of forecasting ability. The impact models are compared in

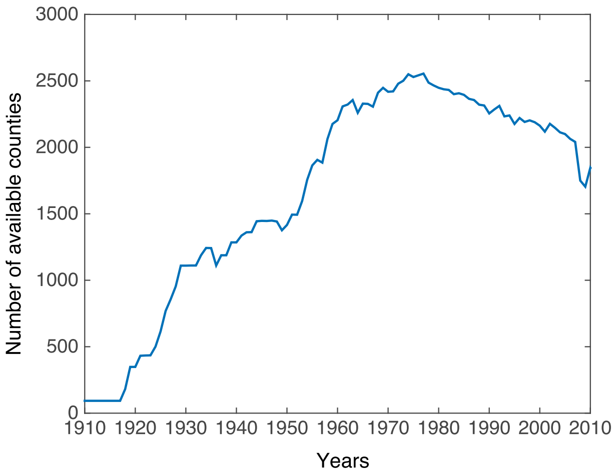

FIG. 1. Number of available counties with corn-yield information.

section 3, and a detailed study is done on the state of Illinois. Last, an impact model is tested for a corn-production forecast and for seasonal forecasting.

2. Material and methods

a. Databases

1) AGRICULTURAL DATA

Corn is one of the major crops in the United States, and we focus here on this cereal. The corn-yield data (in bushels of corn per acre; 1 bushel of corn per acre is approximately equal to 63 kg corn ha^{-1}, assuming that 1 bushel of corn = 56 lb. = 25.4 kg) were collected by the U.S. Department of Agriculture National Agricultural Statistics Service (USDA-NASS). A long historical record is available from 1910 to 2013, but the yield time series are incomplete in some counties, districts, and states. Figure 1 shows the evolution of the number of counties with available crop data from 1910 to 2013. From 1910 to 1975, an increasing number of counties were registered, and a peak was reached in 1975 with more than 2500 available counties. Since then, the amount of data is decreasing. The number of corn-yield records at the county scale depends highly on year, and this can be a problem because statistical models need to be developed with homogeneous data. The meteorological data presented in the next section are available for 1979–2013. We will then focus on these 35 years. This is not a limitation because, as a result of drastic change in agricultural practices, the crop data from earlier years are probably too different to be pooled with the most recent data.

In the United States, the 50 federal states are divided into about 300 agricultural districts, which are themselves divided into about 3000 counties. Figure 2 represents the time series of corn yield for counties of four spatially distant states (Alabama, Texas, Minnesota, and Iowa). Each line refers to the yield time series of a county; for each state, counties of a particular crop district are plotted with the same color. The counties within a particular district have yields that vary in the same way. Length of time series can be short, as for Texas, or longer, as for Minnesota. Some states (e.g., Texas) have crop districts with different orders of magnitude of yield, whereas other states (e.g., Iowa) have districts with similar yield values. The range of district yield variations for a state can be large (see, e.g., Iowa from 1975 to 1995). From 1979 to 2013, the U.S. yield of corn grain increased from 86 to 158 bushels per acre as a result of improved genetics and agricultural practice.

The spatial distribution of the average yield of corn grain in 2000 is represented in Fig. 3. Counties that are left blank indicate that data are not available for this year. The eastern United States produces more corn than the western part because of climate and topography. The Corn Belt region (mainly Iowa, eastern Nebraska, Minnesota, Illinois, and Indiana) is known to be the biggest corn-producing region of the United States. In 2000, these five states produced more than 62% of the U.S. corn production. Yield values have a smooth spatial distribution. This regularity crosses state boundaries and extends over vast areas.

2) METEOROLOGICAL DATA

Temperature and precipitation data were collected at the monthly scale, for 1979–2013, from the ERA-Interim reanalysis of the European Centre for Medium-Range Weather Forecasts (Uppala et al. 2005). The U.S. territorial organization and its subdivision into counties are used to project data from their original 75 km × 75 km regular grid into county-level data. Maps of the 1979–2013 averaged monthly temperature and total precipitation are given in Fig. 4. Large spatial variations can be observed: in general, temperature decreases from south to north while precipitation decreases from south to north and from east to west. Counties with the highest corn yield are located in areas with strong precipitation but moderate temperature. Corn-yield forecasting is reduced to 1979–2013 because the weather data that we use are available during that time. Furthermore, it makes more sense to calibrate an impact model with these 35 latest years than to mix data from very different agricultural practices.

Irrigation is very important in some of the western regions (see Fig. 5). According to the U.S. Geological

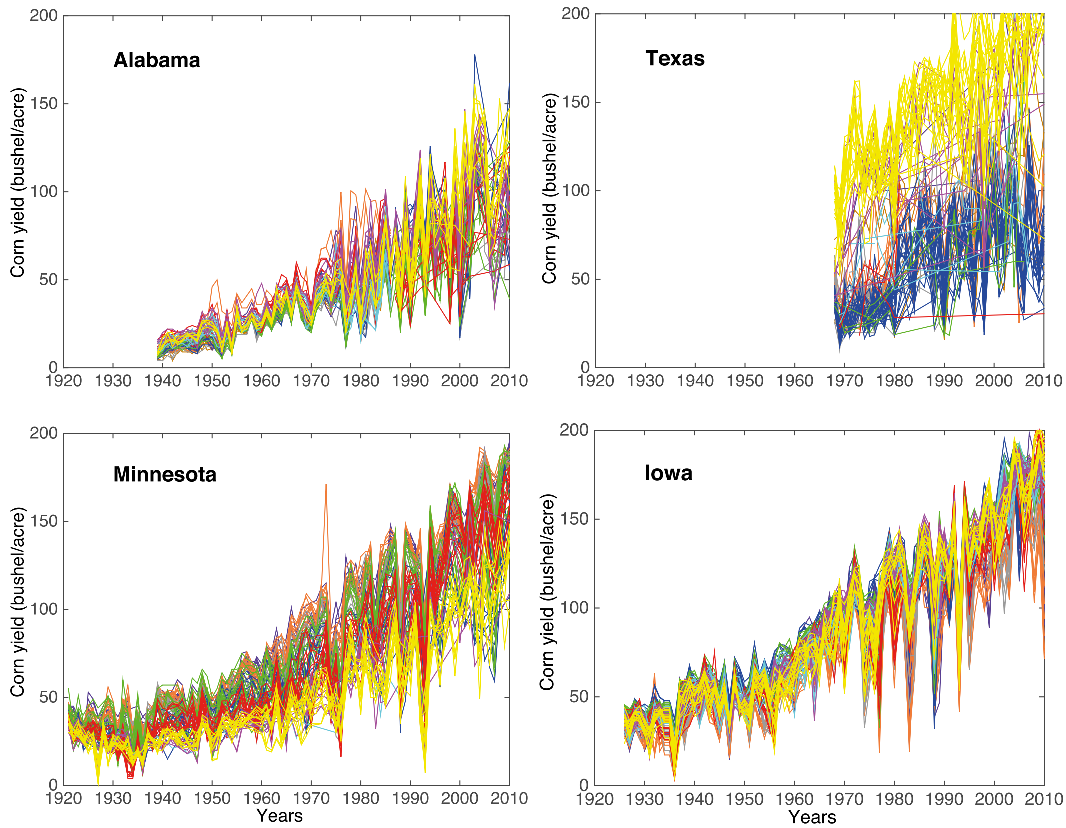

FIG. 2. Time series of corn yield for the counties of four spatially distant states: Alabama, Texas, Minnesota, and Iowa. The counties of a given crop district are plotted with the same color. These four states have been chosen on the basis of the location, the importance of yield values, and the length of the available time series.

Survey (Kenny et al. 2009, p. 23), in 2005 ‘‘[t]he majority of withdrawals (85 percent) and irrigated acres (74 percent) were in the 17 conterminous Western States’’ (west of the black solid line on the map in Fig. 5). The 17 western states are located in areas where average annual precipitation typically is less than 20 in. (1 in. = 2.54 cm) and is insufficient to support crops without supplemental water. Surface water was the primary source of water in the arid West, except in Kansas, Oklahoma, Nebraska, Texas, and South Dakota, where more groundwater was

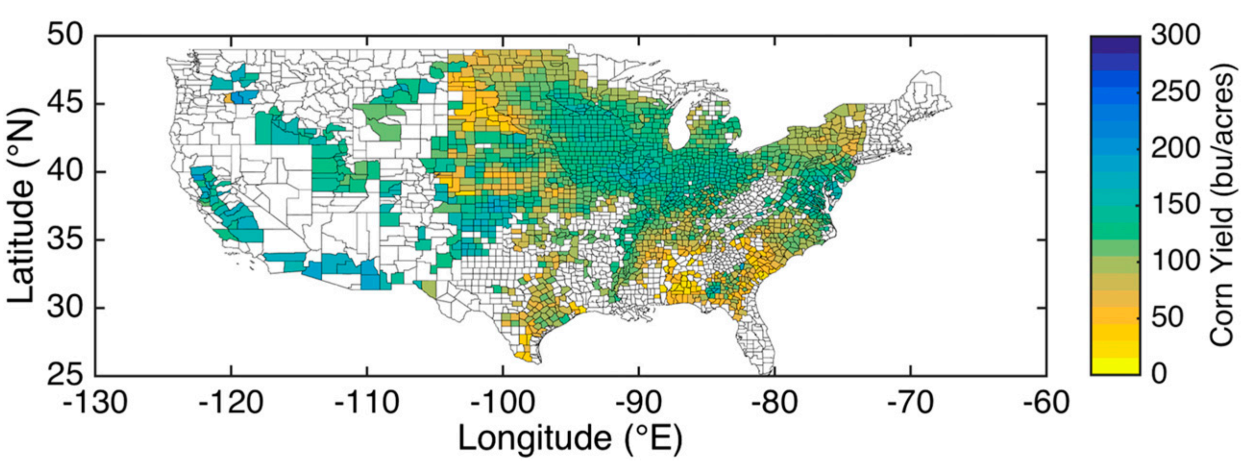

FIG. 3. Corn yield (bushels per acre) in U.S. counties for 2000. A blank county indicates that data are not available.

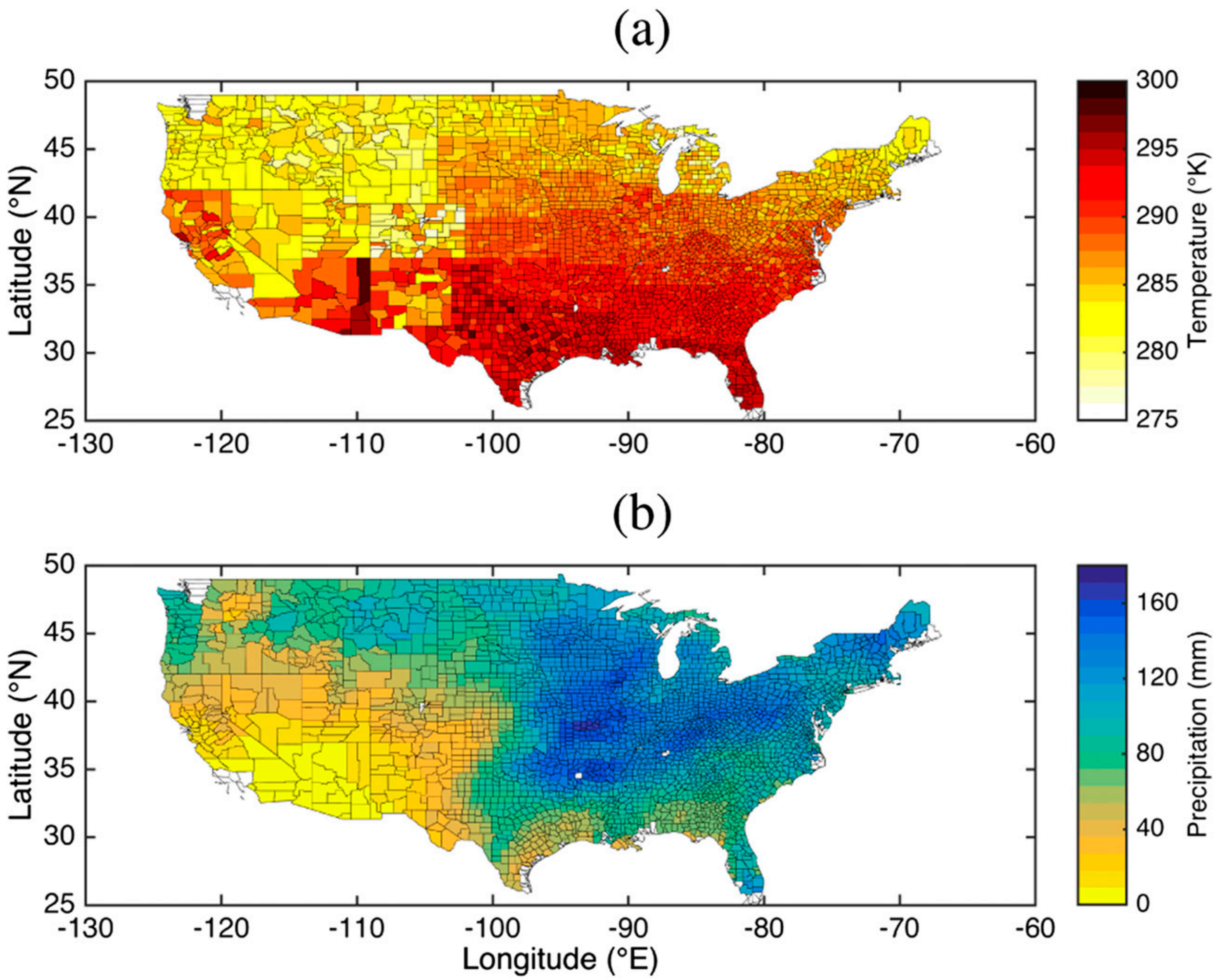

FIG. 4. Monthly-mean values of (a) temperature and (b) precipitation for May, averaged from 1979 to 2013.

used. The 17 western states cumulatively accounted for 93% of total surface-water irrigation withdrawals and 69% of total groundwater irrigation withdrawals. These regions should be discarded so as to obtain a clearer link between precipitation and corn yield. Therefore, all states located west of $-103°$E and some other isolated states (e.g., Nebraska and Texas) have not been considered in this study.

3) NOTATIONS

Let us introduce here some important notations used throughout the paper. The yield of corn grain for year t is $y(t)$, and the long-term trend value is given by $\overline{y}(t)$. This trend represents the slow evolution of the corn yield (mostly from changes in agricultural practices). Both $y(t)$ and $\overline{y}(t)$ are in bushels per acre. The corn-yield anomaly $a(t)$ is then defined as a percentage variation around the trend:

$$a(t) = \frac{y(t) - \overline{y}(t)}{\overline{y}(t)} \in [-1, 1]. \quad (1)$$

If $a(t) > 0$, then the yield in that year is higher than in a regular year; if $a(t) < 0$, then the yield is lower than in a regular year. For instance, an anomaly of $a(t) = 0.25$ means that the corn yield for year t is 25% higher than the annual trend.

The average monthly temperatures from January to December are referred to as $T_{\text{jan}}, \ldots, T_{\text{dec}}$, and the cumulative monthly precipitations are $P_{\text{jan}}, \ldots, P_{\text{dec}}$, using the first three letters of each month as the identifying label. Weather anomalies will also be considered, but since the trend from 1979 to 2013 for the climatological data is relatively low in comparison with interannual variations, these anomalies will not be normalized into percentages, and absolute values will make it easier to interpret the sensitivities.

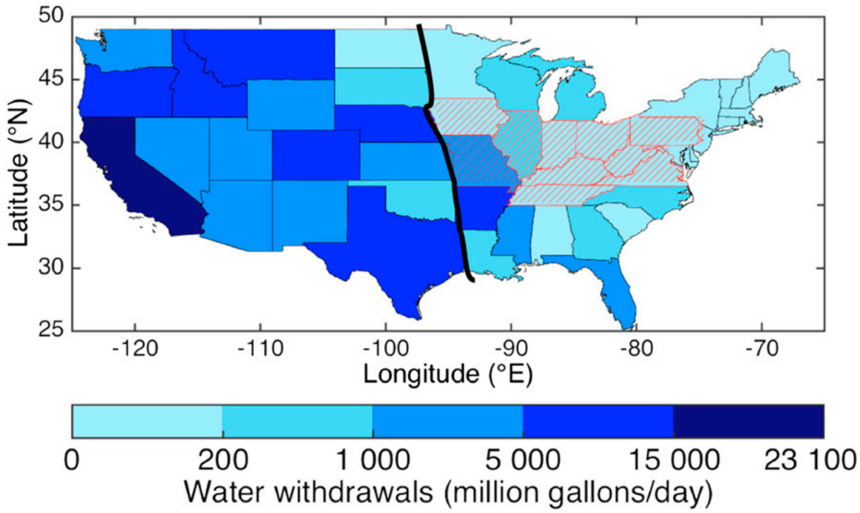

FIG. 5. Total irrigation water withdrawals, in millions of gallons per day, by state in 2010 (Maupin et al. 2014). The areas with red crosshatching represent the states that are most sensitive to weather and rely least on irrigation.

b. Impact models

The objective of our impact models is to estimate the impact of weather anomalies on the corn-yield anomalies. Three statistical models are introduced here.

1) Linear regression (LIN)

The relation between the observations $a(t)$ of corn-yield anomalies and the weather-input anomalies X_{ti} ($i = 1, \ldots, p$, where p is the number of input variables) is formulated as

$$a(t) = \beta_0 + \beta_1\phi_1(X_{t1}) + \cdots + \beta_p\phi_p(X_{tp}) + \varepsilon_t, \quad t = 1, \ldots, T. \tag{2}$$

In theory, the terms $\phi_1, \ldots, \phi_p$ may be nonlinear functions. The "linear" name of this model relates to the use of the regression coefficients $(\beta_i)_{i=1,\ldots,p}$ in a linear way in the above relationship. In this study, all functions $(\phi_i)_{i=1,\ldots,p}$ are chosen as the identity function. The quantities $(\varepsilon_t)_{t=1,\ldots,T}$ are random variables that represent the errors of the model, that is, the part of the target variance in $a(t)$ that is not explained by the inputs of the model.

As with most other regression models, the training of this model (estimating the coefficients β) can be done with a "pooled dataset" that includes data from all of the spatial locations (Fig. 3); conversely, the training can be done independently for each spatial group.

2) Mixed-effects model

The ME models are based on the idea that each datum belongs to a particular group, and the ME model takes into account the particularities of each group. In statistics, the term "effect" is related to the response of the model to an input perturbation. Effects are said to be "fixed" when they are constant for a whole population. They are assumed to be the same each time that data are collected. Estimating these fixed effects is the traditional domain of regression modeling.

Random effects, by comparison, are sample-dependent random variables, meaning that the effects are varying from one type of population to another. A random effect is represented in the model as additional terms whose values are drawn from a probability density function (PDF) that needs to be estimated from data. This PDF is intended to describe the systematic differences between groups. Random effects are useful when data fall into natural groups, such as geographic categories, soil, or land-use groups (Mathieu et al. 2015).

The ME models account for both fixed and random effects. Adding random effects to a model extends the reliability of the inferences by taking into account the variability between groups. As with all statistical regressions, the purpose of ME models is to represent a relationship between a response variable and predictors. By adding group information, ME models provide a compromise between 1) ignoring the groups, pooling all of the available data of all groups into a single dataset used to calibrate a general model (a method called pooling), and 2) fitting independently each group with a separate model (i.e., no pooling).

Let us first consider a random-effect model. Let us suppose that the data used to calibrate an impact model fall into one of m distinct groups, $i = 1, \ldots, m$. To account for the groups in the impact model, the response j in group i is written as

$$a_{ij} = f(\boldsymbol{\phi}, X_{ij}) + g(\boldsymbol{\phi}, X_{ij})\varepsilon_{ij} \quad \text{for} \quad i \in [1, m] \quad \text{and} \quad j \in [1, n_i], \tag{3}$$

where a_{ij} is the response (corn-yield anomaly in our case), the X_{ij} are the vector of predictors (i.e., weather-variable anomalies), and $\boldsymbol{\phi}$ is a vector of model parameters. The terms ε_{ij} are the measurement or process errors. They are supposed to be independent and identically distributed with expected value $E(\varepsilon_{ij}) = 0$ and variance $\mathrm{var}(\varepsilon_{ij}) = 1$. The ε_{ij} are usually assumed to be normally distributed. Here, n_i is the number of observations in group i. The function f specifies the slope of the model (i.e., structural model), and the function g describes the covariance of the error term (residual-error model, usually chosen to be constant).

In Eq. (3), the parameters $\boldsymbol{\phi}$ describe the whole population, with the assumption that they are the same for all groups. If, however, they vary with the group then the model becomes an ME model:

$$a_{ij} = f(\phi_i, X_{ij}) + g(\phi_i, X_{ij})\varepsilon_{ij} \quad \text{for} \quad i \in [1, m] \quad \text{and} \quad j \in [1, n_i], \tag{4}$$

In the ME model of Eq. (4), ϕ_i may be a combination of fixed and random effects:

$$\phi_i = \mathbf{F}\boldsymbol{\beta} + \mathbf{R}b_i, \tag{5}$$

where the random effects b_i are usually described as multivariate normally distributed, with zero mean and covariance matrix $\boldsymbol{\Psi}$. The covariance matrix $\boldsymbol{\Psi}$ of the random effects describes the between-group variability. Estimating the fixed effects $\boldsymbol{\beta}$ and the covariance of the random effects $\boldsymbol{\Psi}$ provides a description of the population that does not assume the parameters ϕ_i are the same across groups. Estimating the random effects b_i also gives a description of specific groups within the data (between- and within-group variability).

The matrices $\mathbf{F}$ and $\mathbf{R}$—called design matrices—are used to identify parameters with linear combinations of fixed and random effects, respectively.

If f is a linear function of parameters ϕ_i, then we obtain a linear ME model (Lindstrom and Bates 1988). In this paper, we use a *nonlinear* ME model (more precisely, with a logistic function) for the identification of the yield tendency (section 2c), and *several linear* ME models for the corn-anomaly forecasting (section 3a): the linear ME models for corn-yield-anomanomaly forecasting are fitted with fixed effects and uncorrelated random effects for the weather inputs (predictor variables), grouped by spatial scales that combine administrative regions (states, or crop districts, or counties), as described at the end of the introduction.

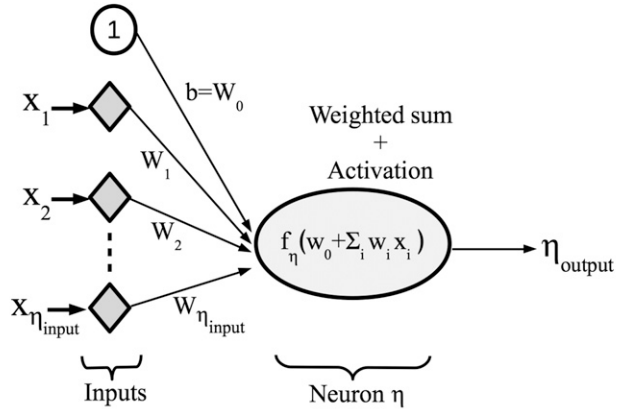

FIG. 6. Schematic representation of an artificial neuron.

3) NEURAL NETWORK (NN)

In this section, the possibility of using a highly nonlinear function for the impact model is considered. Neural networks are good candidates. We only consider here feed-forward neural networks trained using a supervised-learning approach (Bishop 1995). This type of NN has only forward connections from the input to the output layers.

An artificial neuron η (Fig. 6) is a model characterized by 1) n_{inputs} inputs, 2) a vector of weights $\mathbf{W}$, 3) an activation function f_η: $\mathbb{R} \to \mathbb{R}$, and 4) an output η_{output}. The inputs of an artificial neuron are multiplied by weights (plus a bias), and this weighted sum is then passed through its activation function.

A two-layer feed-forward network with one sigmoid function in the hidden layer and one linear output neuron is used here. So a corn-yield anomaly a is modeled by the following equation:

$$\begin{aligned} a &= \sum_{j=1}^{n_{\text{neuron}}} w_j \times \operatorname{logsig}\left(\sum_{i=1}^{n_{\text{inputs}}} x_i w_{ij} + b_j\right) + b_{\text{output}} \\ &= w \times \operatorname{logsig}\left(\sum_{i=1}^{n_{\text{inputs}}} x_i w_i + b_{\text{hidden}}\right) + b_{\text{output}}, \end{aligned} \tag{6}$$

where the x_i are the predictor weather variables, the b_j are the different bias of the NN, n_{inputs} is the number of inputs, and n_{neuron} is the number of neurons in the hidden layer.

Table 1 synthesizes the pros and cons of the three impact models presented in this section (LIN, ME, and NN). By design, a model is imperfect and any statistical modeling requires trade-offs. By comparing these

TABLE 1. Benefits and limitations of linear, mixed-effects, and neural-network models.

Model	Benefits	Limitations
LIN pooling	More data available	Mix of different population behaviors
LIN no-pooling	Locally specialized	Too few data; overfitting risk
ME	Many data available for training; variability between and within groups; information sharing	Complex coding; intensive computational effort required
NN	Simple to implement; nonlinear model	Training less open to interpretation; bias/variance dilemma; need to be regularized

models using a similar database, this study quantifies how much the learning is affected by their disadvantages and improved by their benefits.

c. Experimental configuration

1) Trend identification

As mentioned in the introduction, the first step in a weather-impact analysis is to remove the impact of nonweather factors by trend identification. The long-term trend in corn yield is not the result of climate change but rather is the result of changes in agricultural practice (fertilization, tools, irrigation practices, genetics of crop seeds, etc.). The U.S. corn yield has grown exponentially from 1920 to 2013, and this growth is not the result of climate. We use a nonlinear ME model with a logistic function

$$\overline{y}(t) = c + \frac{K}{1 - pe^{-rt}}$$

that is defined by four parameters. The initial values of these four parameters are chosen by using the relation

$$\mathrm{logit}\{[\overline{y}(t) - c]/K\} = \ln\left(\frac{1}{pe^{-rt}}\right) = rt - \ln(p).$$

The choice of the initial value for the parameters is a critical issue. Figure 7 illustrates the trend identification for two different counties in Alabama and Texas. The resulting yield anomalies are also represented. Despite the different lengths of time series, the ME model appears to be well suited for this task. The sharing of information between counties with many data and neighboring counties with fewer available data gives a good-quality regression for the neighboring county (e.g., the Texas county in Fig. 7): this helps the use of trend values even for periods with few available data (from 1920 to 1970). The similar functional relationship between yield and years adapts well to any maximum value, minimum value, and amplitude of variations. The deducted anomalies are often between −0.5 and 0.5, which means that the yield-relative variation ranges from −50% to +50% of the trend.

2) Input selection

Among all of the possible weather information [section 2a(2)], only six inputs are selected for the impact models to satisfy as well as possible the parsimony principle (Crawley 2011) and therefore limit the overfitting issues. A forward-induction method is used: the first selected weather input is the one that shares the greatest covariance with corn anomalies. The second input is chosen among all of the predictors (except the first selected input) as the one that, combined with the first selected input, defines a two-input linear model that shares the greatest covariance with corn anomalies. This process is repeated until six inputs are selected. This method ensures the selection of the most complementary inputs for the yield modeling. Adding the squared predictors to the list of the potential inputs only slightly improves the results, and therefore no quadratic input information is used here. All of the models in this paper have the same six inputs (weather anomalies): T_{may}, T_{jun}, T_{jul}, T_{aug}, P_{jul}, and P_{aug}; this situation makes result comparisons easier and more legitimate.

3) Group considerations

Linear and NN models are tested for each 1) county, 2) district, and 3) state and 4) for the whole United States. An ME model is built at the national, state, and district level only and not at the county level since county data could not be subdivided.

d. Reliable assessment of the forecast ability

To make meaningful comparisons between the models, the evaluations need to be made at the same spatial scale. Therefore, even if trained at higher spatial scale, the validation will always be done at the county level. This means that a higher-spatial-scale model (e.g., at state level) will be applied at the county level, and then comparisons will be made among models at that spatial scale.

1) Performance metrics

Let us consider a yield anomaly dataset with n samples $a_1, \ldots, a_n$, each associated with a predicted value $\hat{a}_1, \ldots, \hat{a}_n$ (from the regression models), and $\overline{a}$ being the

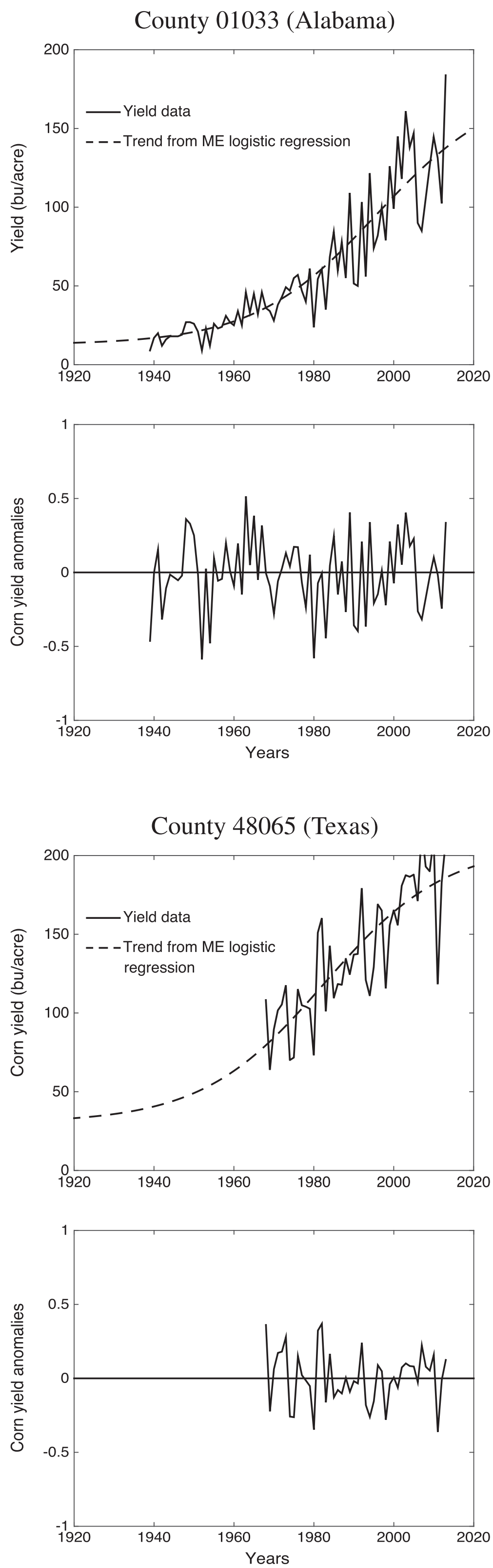

FIG. 7. Trend identification and yield anomalies for one county of Alabama and one county of Texas.

mean of the observed data. The coefficient of determination R^2 is the fraction

$$R^2 = \frac{SS_{\mathrm{Reg}}}{SS_{\mathrm{Tot}}} = \frac{\sum_i (\hat{a}_i - \overline{a})^2}{\sum_i (a_i - \overline{a})^2}. \qquad (7)$$

It represents the percentage of variance explained by the regression model (e.g., corn-yield-anomaly forecasting). The SS_{Tot} is the total sum of squares (proportional to the variance of the data), and SS_{Reg} is the regression sum of squares (also called the explained sum of squares). The coefficient R^2 is a well-known criterion to compare the quality of regressions. The closer to 1 that R^2 is, the closer are the observations to the forecast.

The estimation of R^2 is particularly sensitive to the number of samples, especially in the validation set. A second performance metric is the correlation coefficient Corr: for the linear model, Corr is exactly the square root of R^2. Corr quantifies the dynamical quality of a model and not its absolute values. A third performance metric is the mean-square error (MSE), which measures the departure from the absolute values. Using the same notations, the MSE of the predictor $\hat{a}$ is defined by

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\hat{a}_i - a_i)^2 .$$

Many optimization/regression methods seek to reduce this MSE. These quality metrics can be used on the dataset that is used to calibrate the impact models (i.e., training dataset), but we will mainly use them on an independent dataset (i.e., the testing dataset) to better assess the forecasting ability.

2) SPATIALLY CORRELATED INPUTS

A cross-validation analysis is often conducted in statistical learning to check the overfitting issue and to assess the predictive ability of the model. For this purpose, a temporal random splitting of the global dataset can be performed, with, for instance, 80% of the data used for the model to learn and the remaining 20% used to validate the model. The fitting assessment will be conducted here on 100 draws of these learning/validation datasets. These 100 runs provide an average estimate of the R^2 and an associated uncertainties estimate (characterized by the standard deviation of R^2).

Much care is required when using space–time data because data of spatially close pixels are often highly correlated. As shown by Le Rest (2013) or by Araújo et al. (2005) (in the ecology domain), the cross-validation method in the presence of spatial autocorrelation needs to be adapted, otherwise it will lead to nonreliable results and false conclusions. For cross validation, the learning and validation sets should be independent (Arlot and Celisse 2010)—a critical prerequisite that is almost always ignored. For weather or climate, unfortunately, most data (monthly-mean temperatures and precipitation) are spatially correlated, even at the scale of states. Therefore, learning and validation sets are rarely independent if chosen randomly in the overall weather dataset. To solve this issue, our analysis uses independent years for the learning dataset and the validation dataset. Using truly independent training and validation sets is sometimes ignored in the literature, and it makes the comparison of results among studies difficult.

e. Regularization strategies

1) Bias/variance dilemma and regularization

An excessively simple model makes important bias errors because its lack of complexity does not allow a good representation of the data. A model with high complexity (e.g., an important number of parameters) can yield important variance errors because of overfitting. The overfitting issue means that the model fits the training data very well but the application of the model to new data gives poor results. The model cannot be generalized (Hawkins 2004; Bishop 1995). Choosing the right complexity for a statistical model is associated with the bias/variance dilemma (Geman et al. 1992). Regularization techniques can be used to attenuate this issue. They can act on the data representation, on the fitting algorithm, and on the structure of the model itself (Hastie et al. 2001).

For instance, one could choose to use a smaller number of model inputs (parsimony principle) by analyzing the correlations between inputs (see section 2c). Another strategy is to keep all available inputs but use a regularization technique involving a penalization term: ridge regression (L^2 regularization; e.g., the weight decay in neural networks), lasso (L^1 regularization), or elastic net, which is a linear combination of the lasso and ridge-regression methods (Hastie et al. 2001). Early stopping is another regularization approach to avoid overfitting when training a model with an iterative method. The higher the number of learning iterations is, the smaller are the errors of the training sample but the more sensitive is the model to overfitting. Early stopping consists in stopping the learning as soon as the validation error increases.

2) Neural-network regularization

In this article we call the different methods used to prevent—as well as possible—overfitting "regularization": it refers to all techniques used to limit the complexity of the model and to focus on the bias/variance dilemma (e.g., low number of inputs, low number of parameters, penalization term, or early stopping). Statistical learning with NN needs to determine the number of hidden layers and the number of neurons in these layers. The number of parameters to be estimated is determined by this important choice. To avoid selecting too many parameters (overfitting), NN need to be regularized. Two regularization techniques will be used. First, a penalization term—called weight decay—can be added in the quality criterion that is minimized during the learning stage.

A weight decay d is an additional term in the weight-update rule that causes the weights to decay to zero and so to prevent overfitting (Geman et al. 1992). The penalty term in weight decay, by definition, penalizes large weights. The weight-decay penalty term causes the weights to converge to smaller absolute values than they otherwise would. The larger the weights are, the clearer the model can predict training data. Large weights can hurt generalization, however, because they cause excessive variance of the output. Reducing the absolute value of weights leads to a less flexible model with less specializing to the data used for training. As with all penalization terms, a small weight decay means there is limited attention paid to the issue of large weights, whereas a large weight decay means inference really takes this constraint into account.

So, the weight decay limits overfitting by preventing parameters from taking values that are too large (Bishop 1995). Second, the number of parameters will also be reduced by limiting the number of neurons in the hidden layer. The choice of the optimal regularization parameters (weight decay and number of neurons) is based on the performance metrics used to validate the model. Assuming a fixed training and validation sampling, several NN are built with a weight decay d ranging from 0 to ¼ and a number of neurons N ranging from 1 to 20. For each (N, d) couple, performance metrics such as R^2 and MSE are computed for the learning and testing datasets. This process is repeated 100 times to use a Monte Carlo strategy.

According to the regularization results, none of the feed-forward NN with more than one neuron performs the learning better. An NN with one neuron in its hidden layer represents a traditional logistic regression (i.e., the model results in a linear weighting of the inputs followed by a sigmoid function). This is not a surprise: the weather-to-yield relationships are simple and are close to linear, and the nonlinearity allows one to obtain saturation effects for some input configurations. The saturation effect appears when some inputs have no impact on the output once they reach a certain threshold.

Nationwide, the optimal weight decay is 0, whereas at the scale of the states it is 1/128, and at the scale of a district

TABLE 2. Generalization performance metrics [R^2 (%), Corr (%), and MSE $\times 10^3$ (bushels per acre)] for the five impact models trained at the national scale (i.e., complete pooling): LIN, NN, an ME model with classification by 1) federal state (ME clas. by state), 2) district (ME clas. by district), and 3) county (ME clas. by county). Results are provided for two spatial domains: the total statistics, which include—after irrigation screening—all of the counties where corn-yield data are available, and the central regions (Missouri, Illinois, Indiana, Ohio, Pennsylvania, Virginia, West Virginia, Kentucky, and Tennessee). The impact models are assessed at different spatial scales: at the county, district, state, and national scale. Statistics in boldface font are represented in Fig. 8.

		Total			Central regions		
Model	Validated on	R^2	Corr	MSE	R^2	Corr	MSE
LIN	United States	15	37	42	—	—	—
	State	22	42	44	31	53	33
	District	25	44	45	34	54	31
	County	27	**41**	43	32	**48**	27
NN	United States	16	38	43	—	—	—
	State	17	38	43	17	38	41
	District	19	39	42	18	38	41
	County	30	**48**	43	30	**48**	42
ME clas. by state	United States	16	40	44	—	—	—
	State	22	41	44	28	50	33
	District	24	43	46	31	52	32
	County	32	**50**	45	38	**57**	33
ME clas. by district	United States	16	39	43	—	—	—
	State	19	38	44	25	47	33
	District	22	41	45	28	48	32
	County	30	**48**	45	35	**53**	33
ME clas. by county	United States	21	45	40	—	—	—
	State	23	42	41	33	54	30
	District	25	43	43	34	55	29
	County	34	**51**	42	41	**59**	30

or county it is 1/32. A larger weight decay for learning at the scale of a district or county is consistent because fewer data are available for a county or district training: it is then necessary to further penalize the model complexity to avoid overfitting. Indeed, when learning state by state, fewer data are provided to the model—relative to a nationwide training—to estimate the same number of parameters. So, less information is provided to the NN and the risk of overfitting is more important. Therefore, it is to be expected that the weight decay determined by a cross-validation process is higher than for nationwide training.

3. Illustration for corn yield over the United States

a. Model comparison

Table 2 represents the R^2, Corr, and MSE generalization statistics (section 2d) for the five impact models trained at the U.S. scale (i.e., complete pooling): a LIN model, an NN, and an ME model with classification by 1) state (ME clas. by state), 2) district (ME clas. by district), and 3) county (ME clas. by county). Results are provided for two spatial domains: the "total" statistics, which include—after irrigation screening—all of the counties where corn-yield data are available, and the "central regions," which include the most corn-weather-sensitive regions (Missouri, Illinois, Indiana, Ohio, Pennsylvania, Virginia, West Virginia, Kentucky, and Tennessee). Even if trained at the U.S. scale, the impact models can be assessed at different spatial scales: statistics can evaluate the model at the county, district, state, and national scale. To compare the different models, it is necessary to use the same spatial scale. We will focus here on evaluation at the county level (i.e., the last table row for each model). To facilitate the discussion, cells with boldface font emphasize the quantities that will be mostly discussed in the following. The ME models are better than the LIN model (Corr is ~50% as compared with 41%). This result can be explained by the fact that the linear ME models are equivalent to LIN except that supplementary information is provided to them (i.e., the localization class). When pooling at large spatial scales, such as the state level, the use of district or county information helps to adapt the ME model to more-local conditions. When a reduced-scale training is applied to a model, the ME can benefit from more samples from the large scale, whereas the LIN model could use only a limited dataset, making it sensitive to overfitting issues [section 2b(2)]. The ME models are also always better than or equivalent to the NN. The nonlinear advantage of the NN relative to the linear ME models does not appear to be relevant for the corn-yield application: the weather-to-corn-yield relationship is simple enough to be modeled by a linear model. This statement needs to be moderated, however; the NN has better statistics than the LIN (Corr of 48% as compared with 41%), showing that the nonlinearity of the NN allows fine-tuning to local conditions. This state dependency of the NN is almost as good as the ME models, even if no class information is provided (only weather information). This means that when no class information is available, the NN is a very interesting candidate, but when class information is available, an ME model can benefit from this additional information. The ME models provide about 50% correlation for the total domain, but this number reaches 59% for the central regions where the weather–yield relationship is stronger. This means that R^2 = 41% of the corn-yield variability can be explained by the monthly-mean weather information. The remaining variance is explained by other factors (different agricultural practices) or more-detailed information on weather.

TABLE 3. As in Table 2, but for models trained at the state level.

Model	Validated on	Total			Central regions		
		R^2	Corr	MSE	R^2	Corr	MSE
LIN	State	20	40	44	26	47	32
	District	23	42	45	28	48	31
	County	32	**50**	44	36	**54**	32
NN	State	20	39	45	25	46	36
	District	20	39	46	26	47	35
	County	32	**50**	45	36	**53**	35
ME clas. by district	State	21	40	44	27	49	32
	District	23	42	46	30	50	31
	County	31	**49**	45	37	**56**	32
ME clas. by county	State	20	40	44	28	49	33
	District	23	41	46	30	50	32
	County	31	**49**	45	38	**56**	33

In Table 3, the training is applied independently to each state: the amount of data for the training of each impact model is therefore reduced relative to Table 2. The LIN and NN are improved, showing that a focus on state helps in describing local-condition differences among states. At the same time, the number of data available for each dataset at the state level seems to be enough to avoid overfitting issues. The ME models do not improve when trained at the state level. This was to be expected because the local conditions were already considered in the ME models of Table 2 through the state, district, or county class. A general comment is also that all models (LIN, NN, and ME) appear to be of similar quality; this situation means that focusing the impact models on each state is the best compromise between a focus on local conditions and having enough data for the learning of the models.

On the basis of the results of Tables 2 and 3, which model should be used? To answer this question, another point needs to be considered. As mentioned in sections 2d and 2e, much effort was made in this study to avoid overfitting of the model. The more localized the model is, however, from the whole United States to state, district, and county, the higher the difference between the test dataset and the learning dataset becomes (section 2d). For instance, for state training, for the LIN model, $\mathrm{Corr}_{\mathrm{training}} = 52\%$ and $\mathrm{Corr}_{\mathrm{test}} = 40\%$, but, for county training, $\mathrm{Corr}_{\mathrm{training}} = 68\%$ and $\mathrm{Corr}_{\mathrm{test}} = 46\%$. The amount of available data is reduced when focusing locally, and therefore the bias/variance dilemma becomes more stringent and, even when using regularization techniques, overfitting is a concern. Therefore, it is recommended that a model trained on as much data as possible be used, even at the U.S. level. The ME models with their additional information on spatial location have a true advantage, especially for applications for which the dataset sizes are limited and for which spatial information can be used to increase the dataset (Aires 2012).

As mentioned earlier, the best statistics are obtained on the central regions (Missouri, Illinois, Indiana, Ohio, Pennsylvania, Virginia, West Virginia, Kentucky, and Tennessee), with higher R^2 and Corr values: ME clas. by county reaches 59% when trained nationwide, and approximately 52% is reached for both the LIN and NN models when trained by state. Furthermore, a substantial decrease of MSE by 10 points for almost all models shows how corn yields in these states are more directly related to weather. In these states, we can see that weather has a more important/direct impact on corn yield because our model can extract this information. The central regions represent the two climate zones 4A and 5A of the U.S. climate-zones map from The International Code Council (2006): these zones have warm summers with a mixed-humid climate for the 4A zone and a cool-humid climate for the 5A zone. The states of the central regions produced nearly 40% of the corn in the United States in 2014. As shown in Fig. 5, the states of the central regions (Missouri excepted) rely least on irrigation, with less than 230 gallons per day (1 gallon = 3.785 L) in 2010 according to Maupin et al. (2014). Corn yields will be naturally more sensitive to weather variations in less-irrigated regions.

Figure 8 represents the spatial patterns of the Corr. The spatial consistency is good: neighboring counties have been estimated to have a similar quality since their correlations between observed and forecast values are close. This observation is not surprising because neighboring counties have very similar weather data (because of the smoothness of monthly-mean weather fields). For LIN and ME models, the central part of the United States is always better estimated than are regions at lower or higher latitudes. Northeastern regions provide many data, but our models cannot generalize well on these regions. Results for NN are less good on average but are much more spatially homogeneous than those of the ME or LIN, where some areas are clearly better learned than others. This homogeneity shows the efficient regularization of the NN model: there is no county that is very well learned that is close to a county that is poorly learned, which would have indicated an overfitting issue. These results are due to the NN parameters being chosen to provide a better generalization rate on average, for any given county, resulting in greater homogeneity. The ME and LIN give better results for the central regions even if LIN cannot be generalized outside this area. This difference between LIN and ME—both of which are linear regressions—indicates the improvement generated by group information, particularly for county grouping. Even if ME provides better

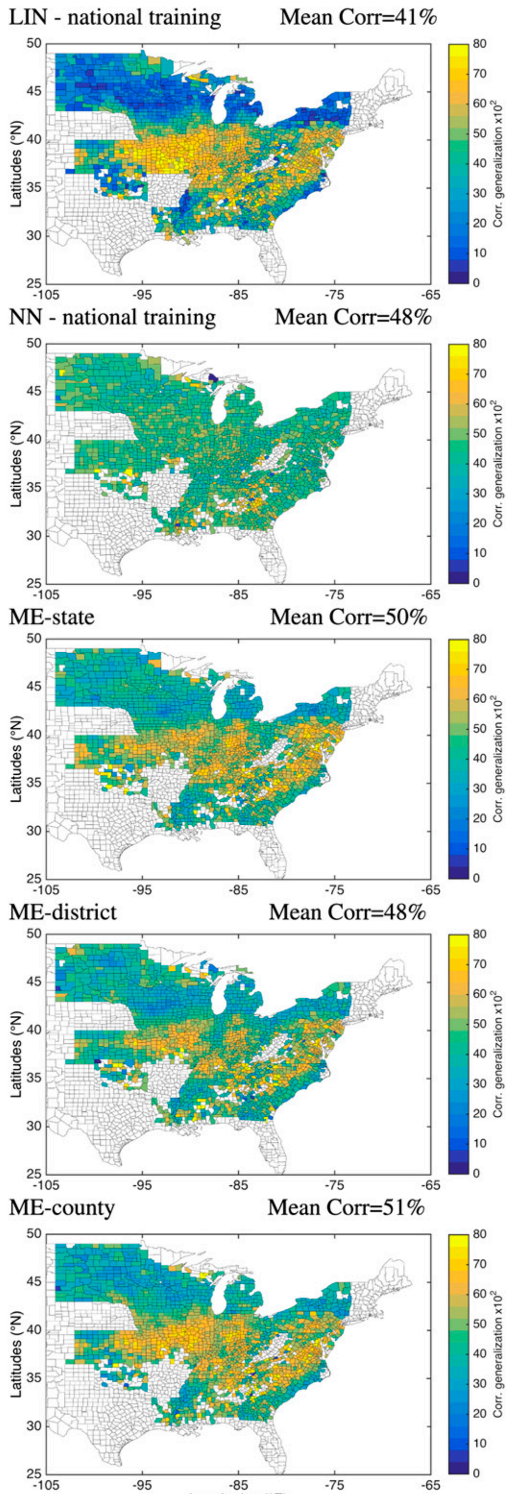

FIG. 8. County Corr values between observed and forecast yield anomalies for generalization when training is nationwide (pooling) and validation is on counties for five models of regression: LIN, NN, ME clas. by state, ME clas. by district, and ME clas. by county (Table 2). County-average values are indicated on each panel.

results in northeastern regions than does LIN, its performance is not as strong as in the central regions. A higher and clearer weather-to-yield sensitivity for the central regions may explain this difference.

When pooling by district or by county, the results are not better. The generalization quality degrades when the training resolution increases. Thus, for LIN or NN, the state spatial scale is a good trade-off, specializing to local conditions while keeping enough data to calibrate the LIN or NN model. ME clas. county is still the impact model that provides the best results.

b. *Focus on Illinois*

We focus here on the state of Illinois (in the so-called Corn Belt) using the ME impact model with classification by county: as mentioned above, for simplicity, this model will be referred to as ''ME clas. county'' in this section. Forecasts of corn-yield anomalies against observational anomalies are represented in Fig. 9. For this state, $R^2_{\text{test}} = 40\%$, $\text{Corr}_{\text{test}} = 58\%$, and $\text{MSE}_{\text{test}} = 26 \times 10^{-3}$. The large negative anomalies (under -0.4) are not well forecast by the impact model, and there is obviously a saturation effect for these low extreme values. Extreme low and high values are well identified by the impact model, which is nice, but low values are dampened.

This problem results from two causes: 1) these extremes are not well represented in the learning dataset, which is always a difficulty when training a statistical model (Coles 2001), and 2) the chosen impact model (ME) might not well represent such extreme cases. This can be an issue since the ME model is used here with a linear regression. A forthcoming study will focus on forecasting such extreme cases.

The corn-yield forecasting $y(t)$ is estimated using the corn-yield trend $\overline{y}(t)$ (section 2c) and the corn-yield-anomalies forecasting $a(t)$:

$$y(t) = \overline{y}(t)[1 + a(t)].$$

To assess the forecast quality of the impact model, Fig. 10 represents the time series of corn-yield observations (black) and the forecast (gray) for two counties from district 60 of Illinois. The forecasts are only generalization statistics: each forecasting year in the time series is the result of a forecast that was trained in a learning dataset that did not use this particular year. Even if the extreme events—in particular, the low yield values (e.g., 1983 or 1988)—are not well retrieved by the impact model (as discussed in Fig. 9), the forecast appears to be satisfactory, with a correlation of 80%. It is a surprise to obtain such a high correlation considering that the yield anomaly for Illinois is ~60% (Fig. 9), but this increased correlation results from the use of the

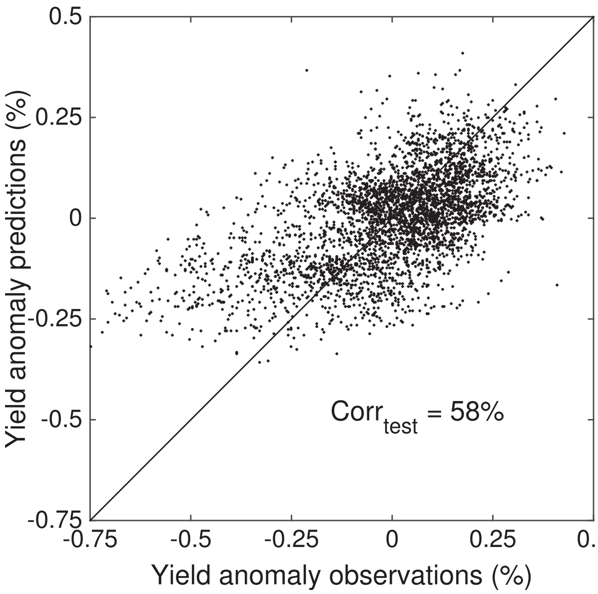

FIG. 9. Forecast vs observed corn-yield anomalies for Illinois using the ME clas. county impact model.

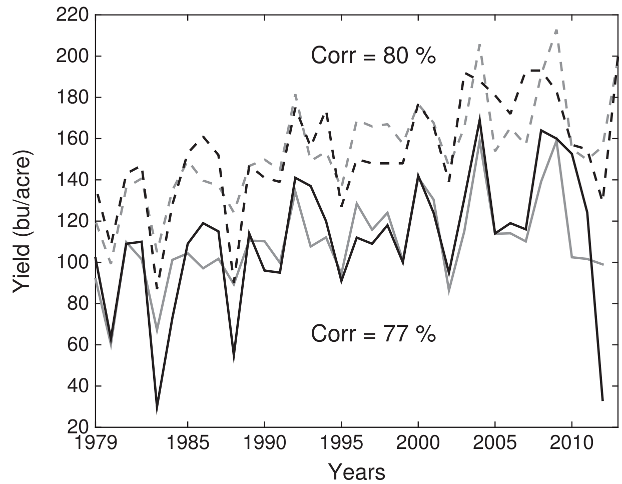

FIG. 10. Observed (black) and forecast (gray) corn-yield time series for two counties (solid and dashed lines) in district 60 of Illinois. The forecast is made using the ME clas. county impact model. Correlations between the observed and forecast time series are indicated.

trend that is here known without uncertainty. Besides, the yield time series and the trend lines are, by construction, highly correlated. This explains also why the corn-yield forecasts adjust so well to different counties in the same state: the trend was obtained in section 2 for each county independently.

c. *Exploitation of the impact model*

1) FORECAST OF CORN PRODUCTION

The ME clas. county impact model was developed in previous sections to quantify and analyze the links between weather and corn yield. It can also be used to evaluate the effects of weather on corn production $p(t)$ [$=y(t) \times s(t)$, where $y(t)$ is the corn-yield forecast (section 3b) and $s(t)$ is the harvested area (in acres; 1 acre = 0.4 ha)]. Figure 11 represents the resulting corn production for four contrasted counties in district 60 of Illinois. The agreement is very impressive, the interannual anomalies are well retrieved except for a few years (e.g., 2008), and the absolute differences between counties are well represented. As compared with the yield forecast in Fig. 10, in addition to the yield-anomaly and tendency information, this figure also uses the production area times series $s(t)$; this, of course, improves the forecasting quality. From Fig. 9 to Fig. 10 the correlation between observed and forecast values increased from ~60% for yield-anomalies forecasting to 80% for yield forecasting, underlining the amount of information provided by the trend. Similarly, from Fig. 10 to Fig. 11 the correlation between observed and forecast values increased from 80% for yield forecasting to 90% for production forecasting, underlying the amount of information provided by the value of harvested areas.

Figure 12 represents the corn-production forecasting ability for the three pooling ME models considered in this study: classification by county, district, or state. Results are represented for Montgomery County in Illinois, for district 60 of Illinois, and for the state of Illinois. It can be seen that the sensitivity of the quality of the ME impact model to the grouping spatial scale (county, district, or state) is limited since they all give similar results (confirming what was seen in section 3a). Furthermore, the statistics at the county, district, or state level are very similar (about 90% of correlation), and therefore the models can adequately be used at the three spatial scales, depending on the need.

2) SEASONAL FORECASTING

In this section, we quantify the information provided by weather inputs from May to August on the ME clas. county impact model: among the six inputs selected in section 2c (T_{may}, T_{jun}, T_{jul}, T_{aug}, P_{jul}, and P_{aug}), ME clas. county is first trained with the available inputs 1) until May (i.e., T_{may}). Another training is done with the available inputs 2) until June (T_{may} and T_{jun}), 3) until July (T_{may}, T_{jun}, T_{jul}, and P_{jul}), and finally 4) until August (all six selected inputs). From May to August, the predictive ability of this model rises from Corr = 1% to Corr = 74% for district 60 of Illinois (light-gray curve in top panel of Fig. 13). The weather in May does not influence corn yield. On the contrary, adding June and, more important, July weather variables to the model

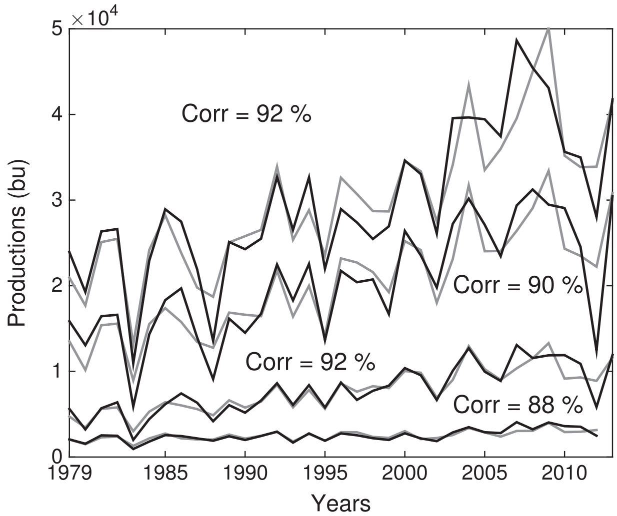

FIG. 11. Observed (black) and forecast (gray) corn production (bushels) for four counties in district 60 of Illinois. The forecast uses the ME clas. county impact model. Correlations between the observed and forecast time series are indicated.

inputs (T_{jun}, T_{jul}, and P_{jul}) highly improves the predictive ability of the model.

Figure 13 (middle panel) illustrates how the district corn-yield-anomaly forecasting is improved when adding weather-inputs information (until June, July, and August): the forecast of the district yield anomalies with only T_{may} input (the gray solid line) is roughly equal to zero for each year and is very far from the black line (the true yield anomalies). This means that, when using information only up to May, no information on the yearly yield is available. When July and then August weather information is added, the forecasting ability increases, emphasizing the importance of the latest months for the crop. Seasonal forecasting unfortunately does not seem to be really possible in this application, as the yield forecasting is reasonable only in July. Note, however, that no information on the crop status is used for this seasonal forecasting (such as the biomass of the plant at the forecasting time); this input information would be important for this kind of seasonal application.

Figure 13 (bottom panel) also illustrates the same four different input configurations but for the forecast of corn production. The three gray dashed lines (June, July, or August forecasting) are much closer to one another than they are for yield anomalies. Improvement with weather data of the successive months is often less important for corn production. It can be explained by the use of the same tendency and area data for the three configurations of inputs.

d. Discussion items

This paper focuses first on the development of weather-impact models for crop yield and on the statistically reliable assessment of the weather sensitivity. We compare five statistical models. According to the results, even if the various models seem close, the choice appears to depend on the chosen spatial resolution. If the U.S. crop data are available at the finest resolution (Table 2), the model ''ME with classification by county'' is a good solution, it provides the best results for the most-weather-sensitive regions and the most-corn-producing areas (the central regions). If data are available at the state level only (Table 3), the ME models become equivalent to the LIN model: fewer data are available for training and more parameters need to be estimated for the ME models than for the LIN one. At this state resolution, even if the ME model with classification by county gives the best results for the the central regions, the LIN model is a good and easy solution on average. The best quality of the NN model lies in its spatial homogeneity, and, even if it does not give the best generalization rates on average, it provides better results than do the LIN or ME models for some states (see Fig. 8).

An independent statistical forecasting of corn yield at the U.S. scale is provided by the NASS model from the USDA. The NASS currently uses two major survey techniques for crop-yield forecasting and estimation (Statistical Methods Branch 2006). In the first one, statistically sampled farmers are asked to report their final harvested yield or their best evaluation of potential yield as based on current conditions during the forecast season. The second one is named the Objective Yield Surveys; it utilizes plant counts and fruit measurements from random plots in selected fields. The overall NASS model does not use any weather data and is based on more direct yield information. Figure 14 compares our ME-model (with classification by county) forecasts with the results given by the September USDA-NASS report (Irwin et al. 2014). The ME model does not use weather data after the month of August, and therefore the two models can be compared. The USDA-NASS results are better, in particular for extreme yield values (improvements of our ME model in this respect will be considered in the future). The ME model, however, does provide better results than the NASS model 14 of 34 times, that is, for 41% of the 35 years (Fig. 14, bottom panel). The overall best results of the USDA-NASS model were to be expected because it uses inputs that are more directly related to corn production, whereas weather data affect corn production more indirectly. Our results are expected when using only weather information: indeed, according to Ray et al. (2015), climate variation explains a third of global crop yield variability.

The cost of the two necessary surveys for the USDA model (especially, the Objective Yield Survey) is much higher than the cost of collecting monthly temperatures

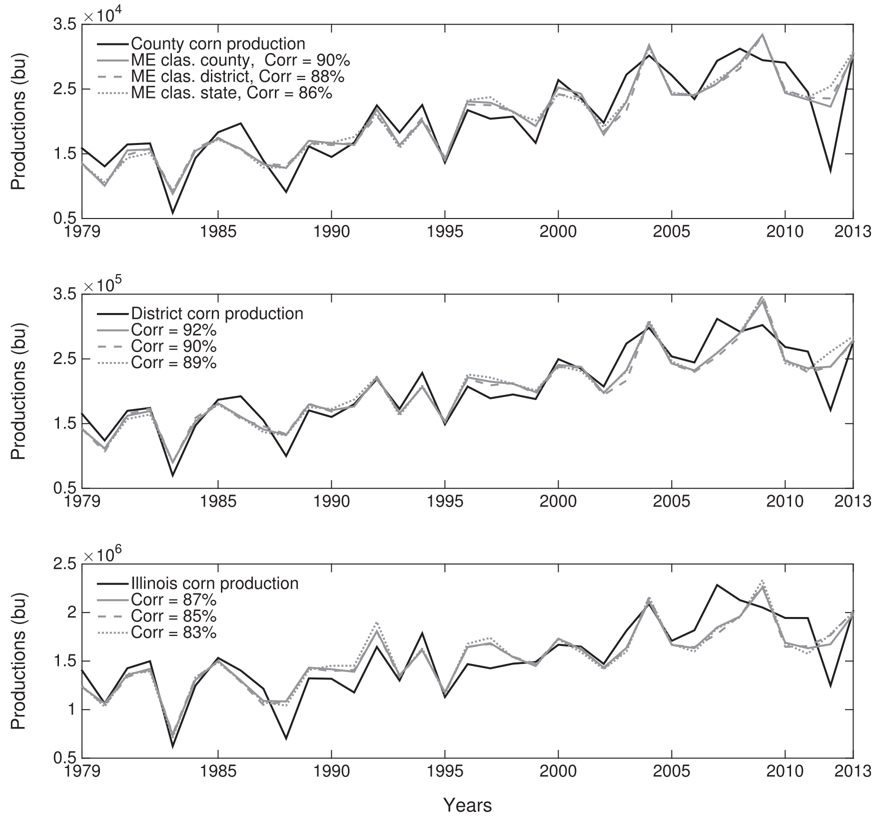

FIG. 12. Corn-production forecasting for the three geographic classifications of ME pooling model (classification by county, district, and state) for (top) Montgomery County (Federal Information Processing Standard code 17135) of Illinois, (middle) district 60 of Illinois, and (bottom) the whole state of Illinois.

and precipitation; agricultural-statistics methods are time consuming and very costly (Liang 2003). Besides, by using much simpler inputs, our models can perform past and future forecasting without any survey. Opportunities are also numerous to improve our weather-impact models: a future study will quantify the benefits of using 1) agroclimate indices instead of simple monthly-mean weather data and 2) satellite observations such as soil moisture (which is more directly linked to crop yield than is precipitation).

The USDA provides regular data on usual planting and harvesting dates for the different states of the United States (USDA 2010). In some states, corn is planted before May. Current model performance does not allow us to consider corn-yield seasonal forecasting starting in May. Results for seasonal forecasting are consistent with the simplicity of the model, however, and in particular with the low number of inputs. Forecast improvements are expected, in particular with regard to expanding the range of potential predictors.

Our statistical models have been developed to forecast anomalies with respect to the long-term trend in yield. This trend is not necessary for seasonal forecasting, but it is necessary for climate predictions. It can be seen that the U.S. corn yield has grown exponentially from 1920 to 2013; this long-term trend in corn yield is not the result of climate change but rather is the result of changes in agriculture practice (fertilization, tools, irrigation practices, genetics of crop seeds, etc.). So, an external trend yield for the next decades would be necessary for any kind of climate forecast, whatever the kind of model (statistical or vegetation). For agriculture, we could define several scenarios with a yield increase over the next decades, or we could use a scenario in which we assume a constant agriculture practice. Our goal is obviously not to quantify what the corn yield will be in 50 years but rather to see whether climate change will be beneficial/detrimental to current agriculture.

4. Conclusions and perspective

Our study introduces a method to build weather-impact models for agriculture. Weather-sensitivity assessments and seasonal-forecast applications are

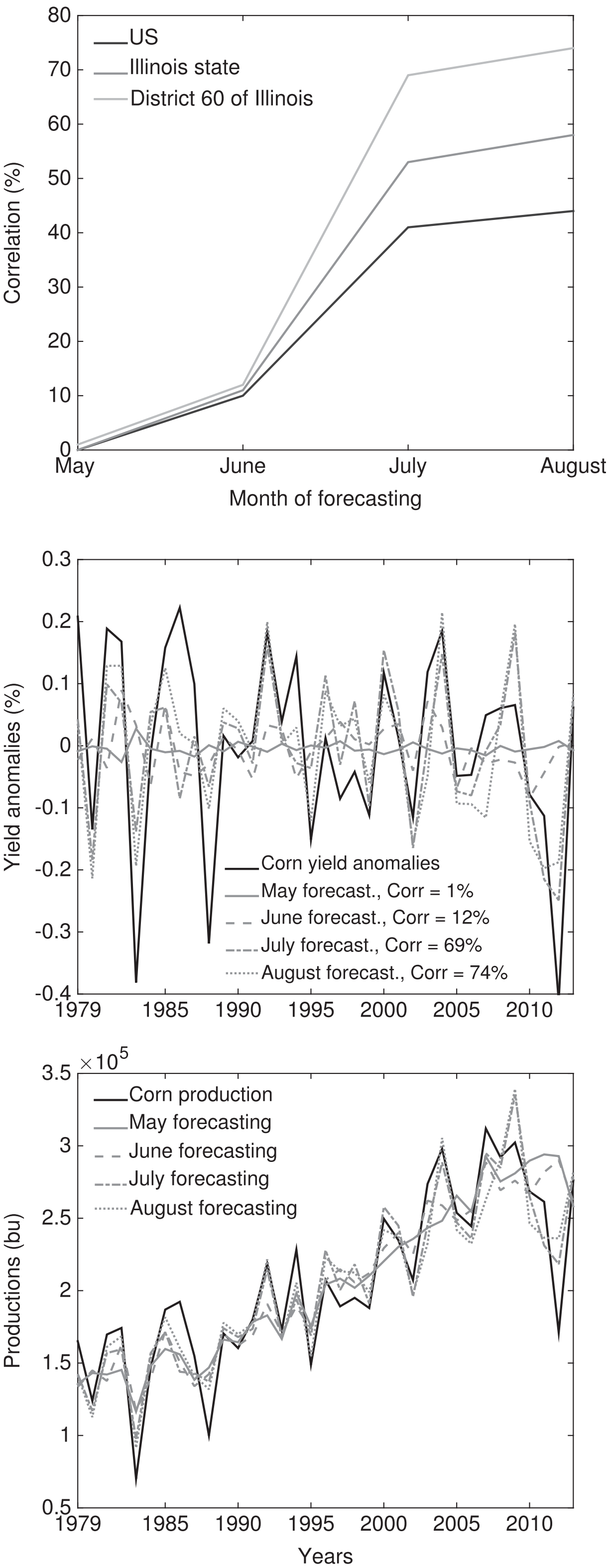

FIG. 13. Illustration of the seasonal forecasting. (top) Correlation between observed and forecast corn-yield values by ME clas. county at national, state, and district levels, according to the forecasting month. Also shown are time series of (middle) the yield-anomalies forecast or (bottom) the production forecast according to the forecasting month for district 60 of Illinois.

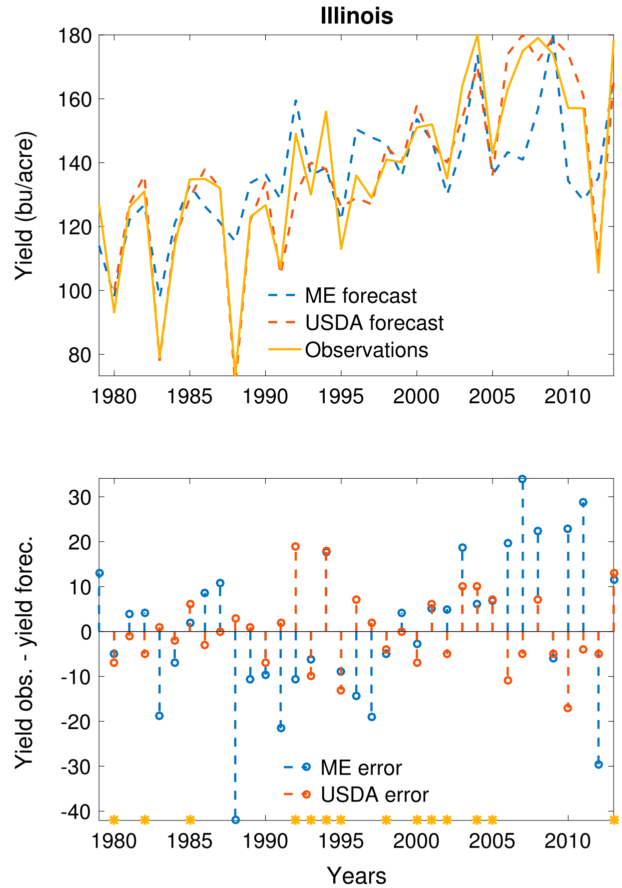

FIG. 14. Comparison of USDA-NASS forecast (September report) and ME clas. county forecast in Illinois. (top) Time series of corn-yield forecasting and observations in Illinois. (bottom) Errors of ME and NASS forecasts. Orange points underline the years that are predicted better by ME than by the NASS model.

discussed. In the context of agriculture, building such impact models suffers from some difficulties: in most cases, not enough data are available to fully represent the complex biophysical relationships. Furthermore, important information that is necessary to perform a good forecast can be missing (such as soil properties or agriculture practices). In this study, we focused on the issue of overfitting and on reliable assessment of the impact model in this low-data-number context. Such concerns are not always sufficiently detailed in the literature, resulting sometimes in an overestimation of what can be obtained with these kinds of models (for biophysical or statistical impact models). Three types of statistical models were considered: linear, neural-network, and mixed-effect models. The spatial scale of the model is an important aspect: the model can be set up at county, district, state, or national levels. Our results show that, for our particular application, state spatial scale is a good trade-off: it allows specializing to local conditions while keeping enough data to calibrate the linear model or the neural-network model. Even if in theory the nonlinearity of the neural-network model allows it to specialize to local conditions, the group information used in ME models is more direct information than what could be inferred by the NN on the basis of weather-input data only. Our ME model with classification by county can predict county corn-yield anomalies with a correlation of 50% between forecast and observed values. In more weather-sensitive regions (Missouri, Illinois, Indiana, Ohio, Pennsylvania, Virginia, West Virginia, Kentucky, and Tennessee), this correlation rises to 60%: this result means that ~40% of the variance can be explained by monthly weather information. As pointed out in Ray et al. (2015), these values are consistent with the simplicity of the considered models (in particular, the limited information carried out by monthly-mean weather predictors). A more detailed study over Illinois shows that the ME clas. county model gives a satisfactory forecast with an observation–forecast yield correlation that is between 70% and 80% at county level. Seasonal forecasting shows that, taking into account the inputs available in June only, the ME clas. county model gives a yield anomaly correlation of only 12%; this value increases to 70% when using the July information.

Our work can be extended in a number of interesting ways. First, we could extend our application to other cultures (e.g., wheat) and other spatial domains such as Europe. Moreover, the use of a nonlinear relationship within the ME model can be investigated for more complex weather-to-yield applications. These new nonlinear models could potentially benefit from both the NN nonlinearity and ME group specification. It was shown that extreme values need to be improved in our models: extreme meteorological events have an important impact on yield and these anomalous years (despite their limited occurrence) have a strong impact on crop production. Following Caubel et al. (2015), we could introduce more-sophisticated weather information as inputs to our impact model. In particular, so-called agroclimatic indices can be better linked to crop yield and therefore can improve the impact models. Indices that focus on extremes could also improve the behavior of the model for extreme years. Furthermore, to quantify the long-term crop-yield evolution (the next 50 years), it will be interesting to apply our impact models to the common climate simulations (from several carbon emission scenarios) and assume a constant agricultural practice so as to analyze the impact of climate change on crop yield over the next decades. The weather-impact models that are developed in this paper could also be used in other socioeconomic fields such as logistics in sales or insurance against catastrophes (Jewson and Brix 2005).

Acknowledgments. This work was supported by the PSL Research University. The authors are also extremely grateful to the National Agricultural Statistics Service of the U.S. Department of Agriculture for the quality of their survey and studies used in this work. We thank the European Centre for Medium-Range Weather Forecasts for providing meteorological data. We appreciate the interesting exchange with two agronomists, A. Perrier and A. Tuzet. We also deeply thank two reviewers and the handling editor for their useful comments.

REFERENCES

Adams, R. M., B. H. Hurd, S. Lenhart, and N. Leary, 1998: Effects of global climate change on agriculture: An interpretative review. *Climate Res.*, **11**, 19–30, doi:10.3354/cr011019.

Aires, F., 2012: Using random effects to build impact models when the available historical record is short. *J. Appl. Meteor. Climatol.*, **51**, 1994–2004, doi:10.1175/JAMC-D-11-0125.1.

Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard, 2005: Validation of species-climate impact models under climate change. *Global Change Biol.*, **11**, 1504–1513, doi:10.1111/j.1365-2486.2005.01000.x.

Arlot, S., and A. Celisse, 2010: A survey of cross-validation procedures for model selection. *Stat. Surv.*, **4**, 40–79, doi:10.1214/09-SS054.

Bishop, C., 1995: *Neural Networks for Pattern Recognition.* Oxford University Press, 504 pp.

Cai, R., D. Yu, and M. Oppenheimer, 2014: Estimating the spatially varying responses of corn yields to weather variations using geographically weighted panel regression. *J. Agric. Resour. Econ.*, **39**, 230–252. [Available online at http://www.waeaonline.org/UserFiles/file/JAREAug20146Caipp230-252.pdf.]

Caubel, J., I. G. de Cortázar-Atauri, M. Launay, N. de Noblet-Ducoudré, F. Huard, P. Bertuzzi, and A.-I. Graux, 2015: Broadening the scope for ecoclimatic indicators to assess crop climate suitability according to ecophysiological, technical and quality criteria. *Agric. For. Meteor.*, **207**, 94–106, doi:10.1016/j.agrformet.2015.02.005.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics, Springer, 224 pp.

Crawley, M. J., 2011: *Statistics: An Introduction Using R.* John Wiley and Sons, 342 pp.

Fahrmeir, L., and G. Tutz, 1994: *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer Series in Statistics, Springer-Verlag, 425 pp.

Geman, S., E. Bienenstock, and R. Doursat, 1992: Neural networks and the bias/variance dilemma. *Neural Comput.*, **4**, 1–58, doi:10.1162/neco.1992.4.1.1.

Greenough, G., M. McGeehin, S. M. Bernard, J. Trtanj, J. Riad, and D. Engelberg, 2001: The potential impacts of climate variability and change on health impacts of extreme weather events in the United States. *Environ. Health Perspect.*, **109** (Suppl. 2), 191–198. [Available online at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240666/pdf/ehp109s-000191.pdf.]

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics, Springer, 539 pp.

Hawkins, D. M., 2004: The problem of overfitting. *J. Chem. Inf. Comput. Sci.*, **44**, 1–12, doi:10.1021/ci0342472.

Hoogenboom, G., 2000: Contribution of agrometeorology to the simulation of crop production and its applications. *Agric. For. Meteor.*, **103**, 137–157, doi:10.1016/S0168-1923(00)00108-8.

International Code Council, 2006: 2006 International Energy Conservation Code. International Code Council Doc., 66 pp. [Available online at https://law.resource.org/pub/us/code/ibr/icc.iecc.2006.pdf.]

IPCC, 2007: *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Cambridge University Press, 976 pp.

Irwin, S., D. Good, and D. Sanders, 2014: Are USDA corn yield forecasts getting better or worse over time? University of Illinois at Urbana–Champaign *farmdoc daily* Vol. 4, No. 169. [Available online at http://farmdocdaily.illinois.edu/2014/09/usda-soybean-yield-forecasts-better-or-worse.html.]

Jewson, S., and A. Brix, 2005: *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations.* Cambridge University Press, 392 pp.

Jones, J. W., and Coauthors, 2003: The DSSAT cropping system model. *Eur. J. Agron.*, **18**, 235–265, doi:10.1016/S1161-0301(02)00107-7.

Kandiannan, K., K. K. Chandaragiri, N. Sankaran, T. B. Balasubramanian, and C. Kailasam, 2002: Crop–weather model for turmeric yield forecasting for Coimbatore district, Tamil Nadu, India. *Agric. For. Meteor.*, **112**, 133–137, doi:10.1016/S0168-1923(02)00125-9.

Kaylen, M. S., and S. S. Koroma, 1991: Trend, weather variables, and the distribution of U.S. corn yields. *Appl. Econ. Perspect. Policy*, **13**, 249–258, doi:10.2307/1349641.

Kenny, J. F., N. L. Barber, S. S. Hutson, K. S. Linsey, J. K. Lovelace, and M. A. Maupin, 2009: Estimated use of water in the United States in 2005. U.S. Geological Survey Circular 1344, 52 pp. [Available online at https://pubs.usgs.gov/circ/1344/pdf/c1344.pdf.]

Kotlowski, W., 2007: Qualitative models of climate variations impact on crop yields. IIASA Rep. IR-07-034, 30 pp. [Available online at http://pure.iiasa.ac.at/8423/1/IR-07-034.pdf.]

Leckebusch, G., U. Ulbrich, and P. Speth, 2002: Identification of extreme events under climate change conditions over Europe and the northwest-Atlantic region: Spatial patterns and time series characteristics. *Geophysical Research Abstracts*, Vol. 27, Abstract 1566. [Available online at http://adsabs.harvard.edu/abs/2002EGSGA..27.1566L.]

Le Rest, K., 2013: Méthodes statistiques pour la modélisation des facteurs influençant la distribution et l'abondance de populations: Application aux rapaces diurnes nichant en France (Statistical methods for modeling of the factors that influence the distribution and abundance of populations: Application to diurnal raptors breeding in France). These doctorat, Université de Poitiers, 153 pp. [Available online at http://www.cebc.cnrs.fr/Fthese/PUBLI/Le_Rest.pdf.]

Lewandrowski, J., and D. Schimmelpfennig, 1999: Economic implications of climate change for U.S. agriculture: Assessing recent evidence. *Land Econ.*, **75**, 39–57, doi:10.2307/3146992.

Liang, S., 2003: *Quantitative Remote Sensing of Land Surfaces.* John Wiley and Sons, 560 pp.

Lindstrom, M. J., and D. M. Bates, 1988: Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *J. Amer. Stat. Assoc.*, **83**, 1014–1022, doi:10.1080/01621459.1988.10478693.

Lobell, D. B., K. N. Cahill, and C. B. Field, 2007: Historical effects of temperature and precipitation on California crop yields. *Climatic Change*, **81**, 187–203, doi:10.1007/s10584-006-9141-3.

Marteau, D., J. Carle, S. Fourneaux, R. Holz, and M. Moreno, 2004: *La Gestion du Risque Climatique* (*Management of Climate Risk*). Economica, 211 pp.

Mathieu, J. A., C. Hatté, J. Balesdent, and E. Parent, 2015: Deep soil carbon dynamics are driven more by soil type than by climate: A worldwide meta-analysis of radiocarbon profiles. *Global Change Biol.*, **21**, 4278–4292, doi:10.1111/gcb.13012.

Maupin, M. A., J. F. Kenny, S. S. Hutson, J. K. Lovelace, N. L. Barber, and K. S. Linsey, 2014: Estimated use of water in the United States in 2010. U.S. Geological Survey Circular 1405, 56 pp. [Available online at http://pubs.usgs.gov/circ/1405/pdf/circ1405.pdf.]

Rauber, R. M., J. E. Walsh, and D. J. Charlevoix, 2005: *Severe and Hazardous Weather: An Introduction to High Impact Meteorology.* Kendall Hunt, 558 pp.

Ray, D. K., J. S. Gerber, G. K. MacDonald, and P. C. West, 2015: Climate variation explains a third of global crop yield variability. *Nat. Commun.*, **6**, 5989, doi:10.1038/ncomms6989.

Schimmelpfennig, D., 1996: Uncertainty in economic models of climate-change impacts. *Climatic Change*, **33**, 213–234, doi:10.1007/BF00140247.

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, doi:10.1073/pnas.0906865106.

Statistical Methods Branch, 2006: The yield forecasting program of NASS. USDA NASS Statistical Methods Branch Staff Rep. SMB 06-01, 109 pp. [Available online at https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/yldfrcst2006.pdf.]

Tannura, M. A., S. H. Irwin, and D. L. Good, 2008: Weather, technology, and corn and soybean yields in the U.S. Corn Belt. University of Illinois at Urbana–Champaign Dept. of Agricultural and Consumer Economics Marketing and Outlook Res. Rep. 2008–01, 123 pp. [Available online at http://www.farmdoc.illinois.edu/marketing/morr/morr_08-01/morr_08-01.pdf.]

Thompson, L. M., 1988: Effects of changes in climate and weather variability on the yields of corn and soybeans. *J. Prod. Agric.*, **1**, 20–27, doi:10.2134/jpa1988.0020.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

USDA, 2010: Field crops: Usual planting and harvesting dates—October 2010. USDA NASS Agricultural Handbook 628, 50 pp. [Available at http://usda.mannlib.cornell.edu/usda/nass/planting//2010s/2010/planting-10-29-2010.pdf.]