2018

Predicting county level corn yields using deep, long, short-term memory models in the Corn Belt

Zehui Jiang
Iowa State University

Predicting county level corn yields using deep, long, short-term memory models in the Corn Belt

by

Zehui Jiang

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Dermot Hayes, Major Professor
Chad Hart
Alejandro Plastina
Baskar Ganapathysubramanian
Soumik Sarkar

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I hereby sincerely acknowledge my major professor, Dr. Dermot Hayes, for his wise guidance and unlimited supply of time and patience throughout the course of this research. His insights and words of encouragement always inspired me to complete my research.

I would like to thank my committee members Dr. Soumik Sarkar and Dr. Baskar Ganapathysubramanian, who come from the Department of Mechanical Engineering, for their generous help in machine learning methods application. Thanks for offering me the access to hourly weather data and the high performance computing cluster in their lab. I would like to thank my committee members Dr. Chad Hart and Dr. Alejandro Plastina, for their constructive comments to help me complete this dissertation.

I would also like to thank Dr. Chao Liu, who was a postdoc in the lab of mechanics. I have been benefited a lot from the discussion with him about the dissertation.

Finally, I would like to give my special thanks to my family for their understanding, encouragement and love.

ABSTRACT

Having an accurate corn yield prediction is useful because it provides information about production and equilibrium post-harvest futures price prior to harvest. A publicly available corn yield prediction can help address emergent information asymmetry problems and, in doing so, improve price efficiency on futures markets. This paper is the first to predict corn yield using Long Short-Term Memory (LSTM), a special Recurrent Neural Network method. Our prediction is only 0.83 bushel/acre lower than actual corn yields in the Corn Belt, and is more accurate than the pre-harvest prediction from the USDA. And more importantly, our model provides a publicly available source that will contribute to eliminating the information asymmetry problem that arises from private sector crop yield prediction.

CHAPTER 1. INTRODUCTION

In the 2001 Nobel Prize winning paper "The Market for Lemons", George Akerlof shows that in a second-hand car market where asymmetric information exists, sellers know the quality of their cars but buyers do not, thus buyers offer a price based on expected quality. As a result, sellers of high quality cars worth more than the average price will exit the market, driving the proportion of low value cars up and offer prices down. Eventually, only "lemons" are left in the market and the market collapses. The key to this collapse is information asymmetry (i.e., sellers have more information than buyers).

The solution to information asymmetry is to provide public information to all participants in the market at the same time. Since 1964, the United States Department of Agriculture (USDA) has predicted national corn yield and production using estimated yield surveys from farms and enumerators making field visits in important corn production areas. These traditional statistical method results are subjective since they rely on farmers' estimations. USDA has also tried new sources of data such as satellite imagery from MODIS (moderate resolution imaging spectroradiometer). However, as of 2018, USDA continues to rely on the survey-based data, as they are still in the very beginning stage of developing new prediction methods.

Several private companies, such as Lanworth, Tellus Labs, and Climate Corp, are probably in a position to improve on the USDA survey. In contrast with the monthly state-level prediction from USDA, these companies set up plant growth models based on weather information and expert knowledge, monitor satellite imagery and weather patterns, and incorporate as many independent lines of evidence as possible into their estimates to produce daily yield estimates. Corn futures traders in the Chicago Mercantile Exchange who have preferential access to this information may be in a position to make profitable trades to the detriment of traders who do not have access. Corn

futures prices have a strong connection with expected corn yield. If a company or individual can predict yield more precisely, they will have better information about futures prices and can speculate in the futures market.

Figure 1.1 shows corn futures price, production, and yield from 1980 to 2016, demonstrating a negative relationship between futures price and corn productions. This indicates that accurate corn yield prediction data would be an important reference to the corn futures market.

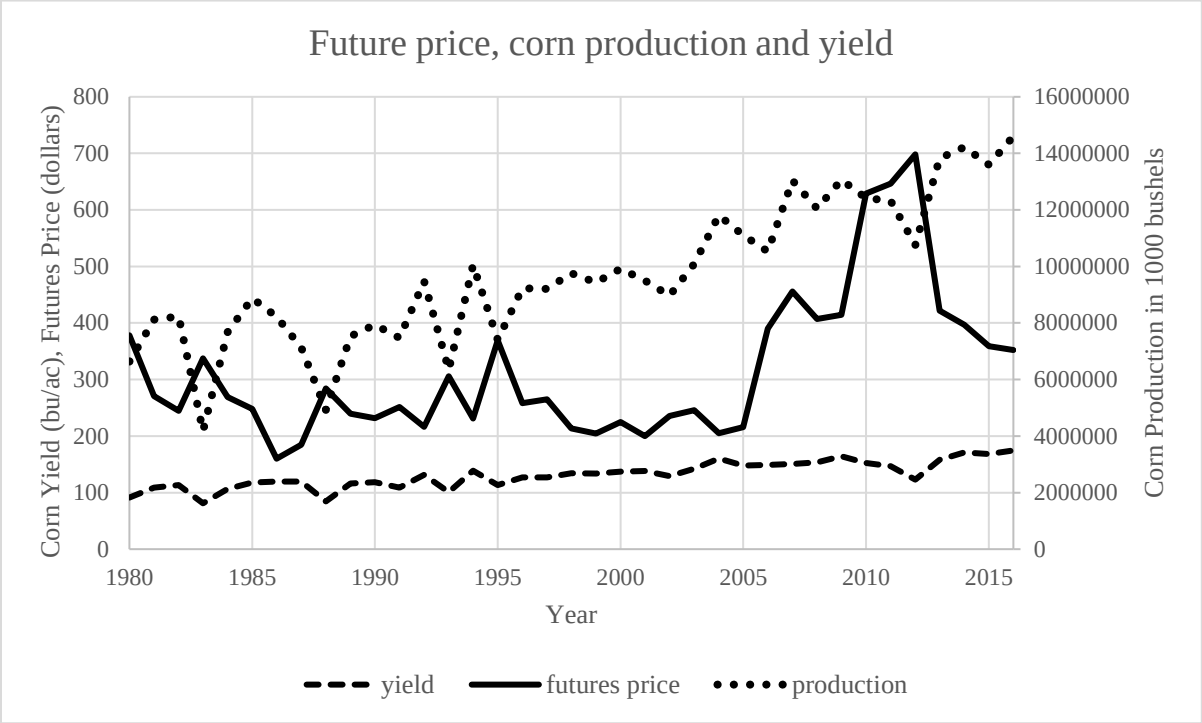


Figure 1.1: Corn futures price, production, and yield, 1980–2016

Our motivation is to provide the public high-quality corn yield prediction that can substitute for private information from companies, thus eliminating information asymmetry in the corn futures market in the long run. We attempt to do so by improving the accuracy and quality of USDA predictions. Two limitations of USDA predictions are that they only offer state-level, but not county level predictions, and that USDA only publishes four monthly prediction reports annually instead of a daily early prediction during the whole corn growing season. In this paper,

we utilize modern data science techniques to provide monthly Corn Belt corn yield predictions at the county level. This method can be used to provide a daily update on expected yield.

1.1 Background Knowledge

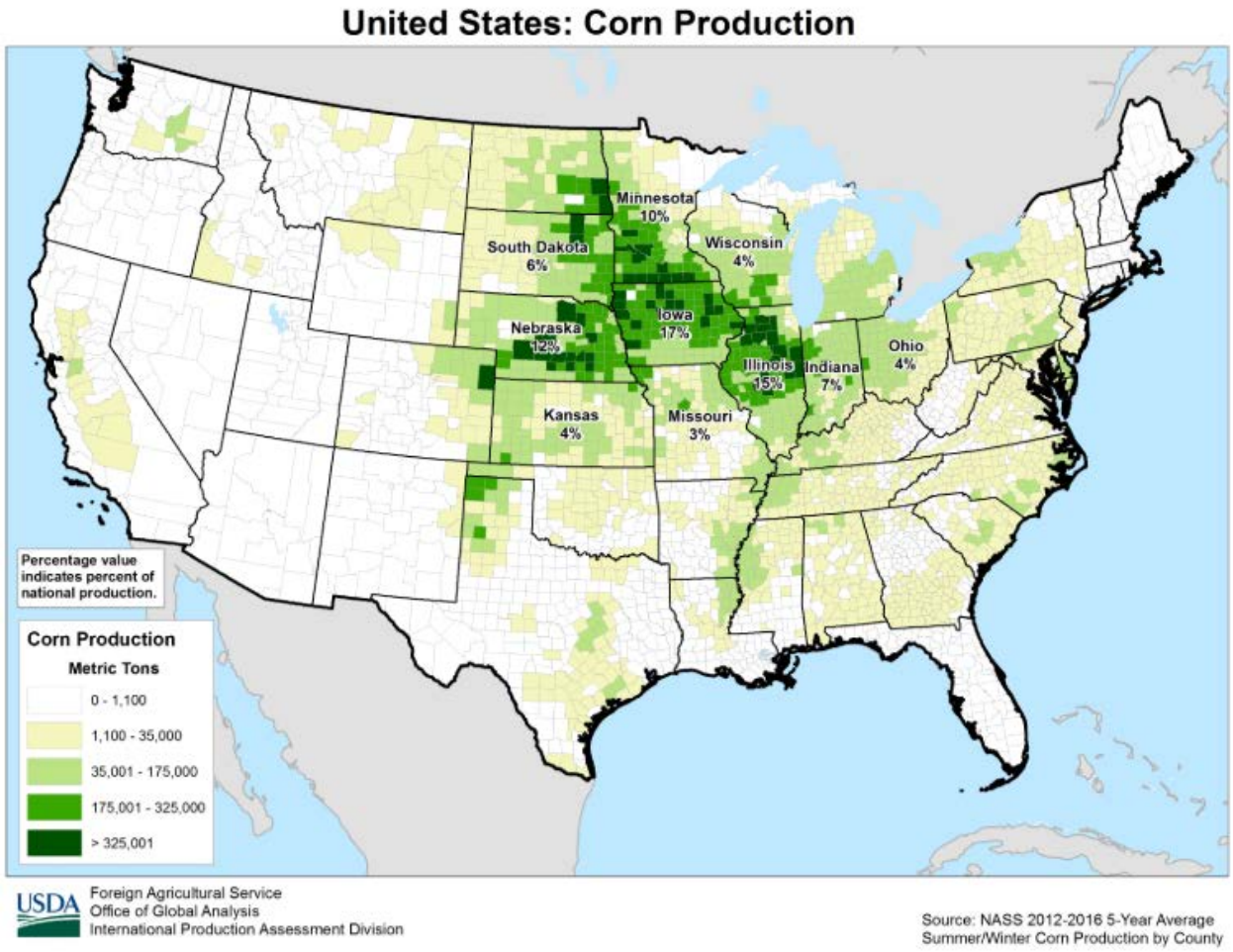


Figure 1.2: Percentage of U.S. corn yield grown in the Corn Belt

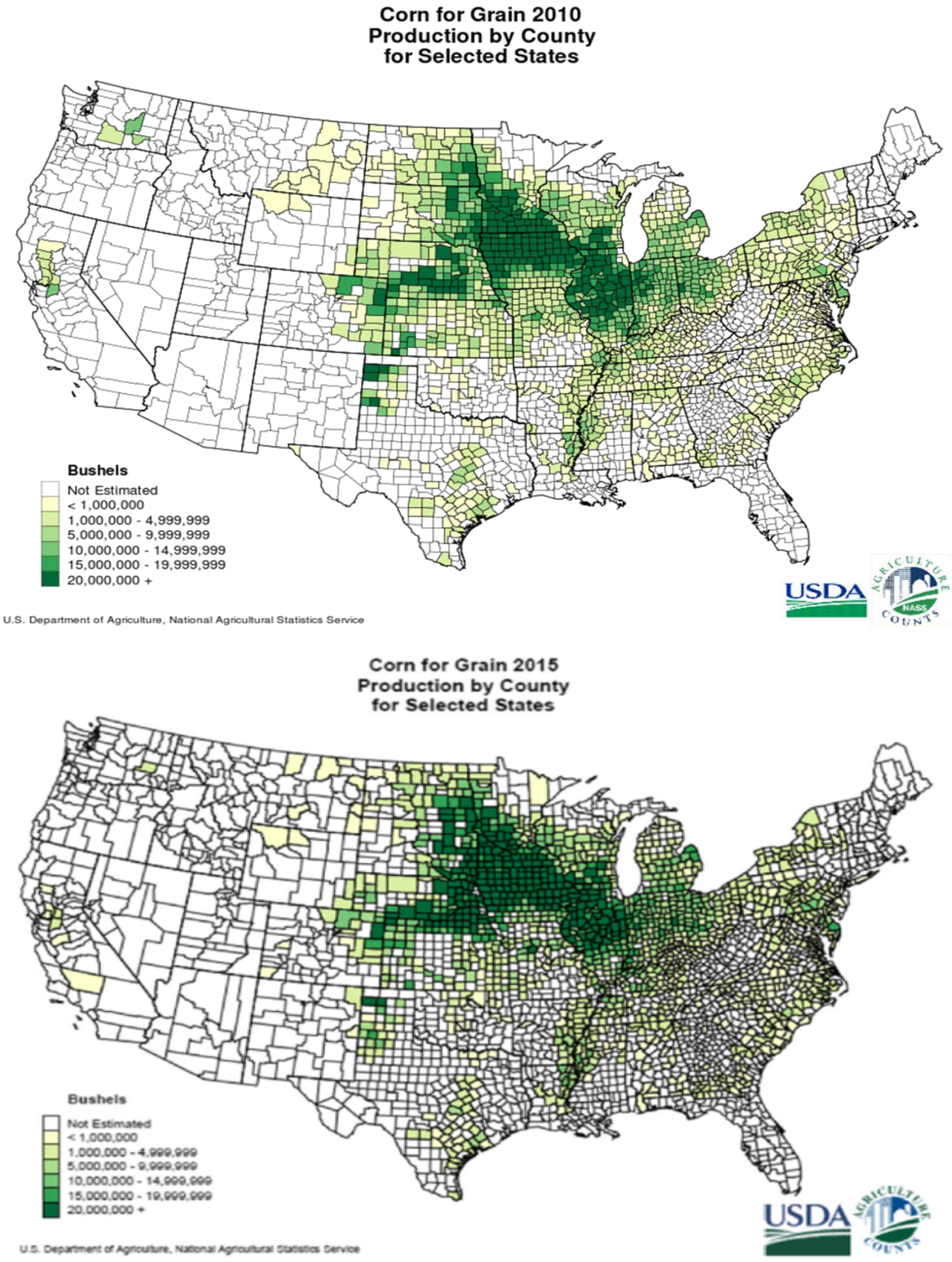


Figure 1.3: Changes in corn production by county, 2010 and 2015

Figure 1.2 shows that over 80% of U.S. corn is grown in the Corn Belt, and these regions kept changing over time due to climate change. By comparing corn planting areas in 2010 and 2015, as shown in Figure 1.3, we can see trend of corn planting moving north, possibly due to global warming.

The Corn Belt includes Iowa, Illinois, Indiana, southern Michigan, western Ohio, eastern Nebraska, eastern Kansas, southern Minnesota and parts of Missouri.

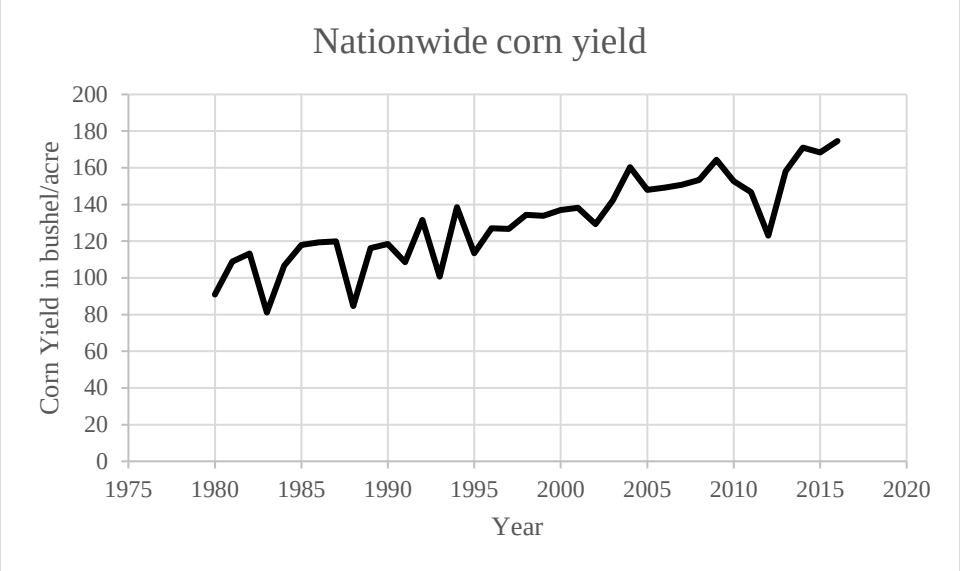


Figure 1.4: Nationwide corn yield, 1980–2016

Corn is typically planted in April and harvested in October. The USDA reports the nationwide county level corn yield in late February of the following year. The USDA provides a monthly estimate of expected yield beginning in August.

Corn yield growth increased rapidly after 1950 with genetic improvement in seed and farm management. By 1950, 99% of corn was grown from hybrids. Figure 1.4 shows an increasing trend in corn yield from 1980 to 2016.

Conventional crop forecasting is divided into two categories. A white-box approach is based on theoretical structure with calibration while a black-box approach is only based on data. The difference between white-box and black-box is the amount of prior information available. A white-box model supposes all necessary information is known and a black-box model has no prior information. In reality, some prior information can be figured out but there must be some unknown left. Thus, modern yield prediction proposes approaches called grey-box, which take advantage of both methods. Nonlinear regression and artificial neural network are examples of a grey-box approach that are popular in modern yield prediction.

Agricultural economists often focus on examining factors affecting crop yield and how they work. Statisticians and data scientists make efforts to improve the accuracy of crop forecasting. Chawla et al. (2016) present a novel, knowledge-based statistical forecasting approach to predict county-wide corn yield in the state of Iowa. This gray-box model was based on Bayesian networks (BN) to build a directed acyclic graph between predictors and yields. This model captured prior casual knowledge from previous agricultural papers and took advantage of machine learning techniques to improving prediction accuracy. This work first extends Chawla's work across more states and time and finds its inadequacies in prediction accuracy and proposes a more powerful supervised method to improve results.

At the beginning of the new century, traditional machine learning methods began to be considered for yield prediction. Past research using machine learning usually created models to predict yield with discrete weather variables. As weather conditions are continuous through the growing season, no research has explored the long-term dependencies between weather condition from any specific time point and final yield. With deep learning, we can see whether and how each point along the weather time series influences the final yield result. We use Long Short-Term

Memory (LSTM), a special form of Recurrent Neural Network (RNN). Its efficiency in capturing long-term dependencies and predicting time series with complex inner relations makes it a good choice for our work. Though LSTM is one of the most popular methods in deep learning, it has never been used in any other field except natural language processing. This work is the first to apply LSTM in crop yield prediction, which indicates its potential in solving other prediction problems, and contributes to improving the accuracy of publicly available corn yield prediction.

Our prediction is only 0.83 bushel/acre lower than actual corn yields in the Corn Belt, this difference is lower than the prediction from the USDA. Eighty percent of our LSTM model county-level corn yield predictions fall in the +/-20 bushel accuracy region. Results show that our LSTM model can provide good early prediction and accurate Corn Belt county-level corn yield prediction without farm management and corn seed data. Our model's shortcoming is that predictions do not perform well in extreme conditions because these are rare in the data.

The rest of this dissertation is organized as follows: Section 2 provides a literature review, related to the expert knowledge needed to build the model. Section 3 explains how model variables are selected, the process of data collection, and data cleaning. Methodologies used in the prediction and explains the details of various model settings are provided in section 4. Section 5 shows results and valuation of the prediction performance in Iowa. Section 6 extends to other Corn Belt states and early prediction. Section 7 concludes.

CHAPTER 2. LITERATURE REVIEW

Expert knowledge is important in input variable selection and preprocessing when building reasonable prediction models. Below (2008) explored factors affecting corn growth and found weather —rain, temperature, wind, and humidity—as key factors.

Foote and Bean (1951) were the first to investigate evidence of trends and patterns in crop yields associated with weather. Their study learned whether the available records on crop yields per acre contain variations from year to year that might be useful in anticipating future changes in per acre yields through conducting several statistical tests. Kaylen and Koroma (1991) suggest limiting weather variables to temperature and precipitation to model U.S. corn yields. They present a linear model using a stochastic trend and monthly rainfall and temperature variables from May to August. Deschênes and Greenstone (2007) estimate the effect of variation in temperature and precipitation on agricultural profit to measure the economic impact of climate change on US agricultural land. They find that yield decreases in temperature and increases in rainfall. Almaraz el at. (2008) study the relationship between climate variability (temperature and precipitation) and corn yield trends over a period of 33 years for south-western Quebec. They find that July temperature and May precipitation explain more than half of yield variability associated with climate.

All the aforementioned literature assumes linear relationships between corn yield and weather variables; however, weather impacts on corn yield are complex and subtle. Schlenker and Roberts (2009) conclude that temperature has a nonlinear effect on corn yield. They use nationwide county-level data and present a steep non-linear decline in yields when temperature is above 29^{o}C. Yu et al. (2011) examine the drought effect on crop yield in Iowa, Illinois, and Indiana and find significant results. They also estimate non-linear weather impacts on corn yield using a Bayesian

approach. Li et al. (2014) introduce a Bayesian dynamic linear model and estimated the impact of weather factors – temperature, amount of rainfall and drought – on corn yield. They find that weather events such as drought, flood and extreme heat cause considerable damage to corn yields and the critical temperature varies across the state of Iowa. Kefaya (2014) introduces a supervised classification method for crop yield prediction by improving a regularization technique that was used to obtain a computationally efficient classifier based on naïve Bayes. The proposed method is found to be much better than naïve Bayes model and can be extended to explore more complex predictor relations.

Basso et al. (2001) add spatial measurements – remote sensing in crop models. Remote sensing provide spatial inputs for the model and results show that a combination of crop model and remote sensing can identify management zones and causes for yield variability. Charles et al. (2014) use a spatial Bayesian regression model to predict maize yields in the Corn Belt. Though spatial smoothness among the regression coefficients will mitigate the effects of noisy data across regions and improve yield forecasting, their results indicate that corn yield prediction still remains a difficult problem. Gerlt et al. (2014) studied the relationship between farm-level yields and county-level yields by exploiting the fact that county-level yields are the aggregate of farm level yields to derive bounds that can be reduced to direct relationships between county- and farm-level yields under certain conditions.

Technology may also be an important factor. Tannura et al. (2008) investigated the relationship between weather, technology, and corn and soybean yields in the U.S. Corn Belt. Their findings provided strong evidence that precipitation, temperature, and a linear time trend to represent technological improvement explained all but a small portion of the variation in corn and soybean yields in the U.S. Corn Belt. As a result, Trend yield forecasts based on perceptions of a

rapid increase in technology may eventually lead to poor forecasts, as unfavorable weather in the future may lead to unexpectedly low corn yields.

Recently, data science has developed quickly and new techniques, such as machine learning, have been used to predict crop yields. Kaul, Hill, and Walthall (2005) use an artificial neural network model with rainfall data to predict corn and soybean yield. The Maryland Water Quality Improvement Act of 1998 requires mandatory nutrient management planning on all agricultural land in Maryland. In order to effectively predict yields for typical climatic conditions, they choose a machine learning method and find ANN models consistently produce more accurate yield predictions than regression models. Newlands and Townley-Smith (2010) were the first to apply a Bayesian Network (BN) to crop yield prediction and try to predict energy crop yield and present results of predicted probability distribution. They find predicted probability distribution could be mad, but no exact prediction result. Chawla et al. (2016) use BN to predict county-level corn yield in Iowa from 2005 to 2009 and find actual predicted yield, thus making it possible to assess the accuracy of this method.

Kim et al. (2016) use four machine learning approaches for corn yield estimation in Iowa—SVM (Support Vector Machine), RF (Random Forest), ERT (Extremely Randomized Trees), and DL (Deep Learning)—and use satellite images and climate data as explanatory variables. The differences between their predictions and USDA statistics were about 6%–8%, thus they conclude that machine learning can be a viable option for crop yield modeling. In particular, they find the results of deep learning methods were more stable. However, the deep learning method they used is quite different from our model, and they fail to offer county level or pre-harvest corn yield prediction.

RNN is a neural network originally used by biologist to mimic human brains. The basic RNN architecture was developed in the 1980s; however, it has a fundamental problem —it cannot learn to look far back into the past. Michael and Peter (2003) utilized the dynamic behavior of the RNN to categorize input sequences into different specified classes, as the prediction task strongly supports the development of a suitable internal structure, representing the main features of the input sequence, to solve the classification problem. Therefore, the speed and success of the training as well as the generalization ability of the trained RNN are significantly improved. The trained RNN provides good classification performance and enables the user to assess efficiently the degree of reliability of the classification result. Che et al. (2016) exploited the missing value patterns for effective imputation and improving prediction performance by developing a novel deep learning model. Their GRU-D model was based on Gated Recurrent Unit (GRU), a state-of-the-art recurrent neural network, and took two representations of missing patterns to incorporate them into a deep model architecture so that it not only captured the long-term temporal dependencies in time series, but also utilized the missing patterns to achieve better prediction results.

Hochreiter and Schmidhuber (1997) first rigorously analyze this problem in their paper "Long Short-Term Memory." LSTM network is a deep learning RNN that can solve the fundamental problems of traditional RNN models and has become very popular in the field of natural language processing. As of 2016, major technology companies use LSTM networks as fundamental components in new products. For example, Google uses LSTM for speech recognition on smartphones and Google Translate. The University of Montreal first developed a library for Python called Theano to manipulate deep learning models. Google developed their own library, TensorFlow, in 2015 to meet their needs for systems capable of building and training neural networks to detect and decipher patterns and correlations. Keras, also developed in 2015, is

a high-level neural networks API, written in Python and capable of running on top of either TensorFlow or Theano. For academics, Keras is one of the most popular packages for deep learning since it minimizes the number of user actions required for common use cases. More details about “Keras” can be found at keras.io.

CHAPTER 3. DATA COLLECTION

We collect data for ten states in the Corn Belt: Iowa, Illinois, Indiana, Minnesota, Nebraska, Kansas, Michigan, Ohio, Missouri, and South Dakota. All data is from 1980 to 2016. The data range is restricted due to availability of weather data. The first 33 years are selected as the training data, while the most recent four years are used as the testing data to explore the model's predictive capability. As Iowa is the dominant corn planting state, we use it as our model testing state. The data collection process consists of two parts: collection of raw data and feature selection and preprocessing.

3.1 Outcome Variables

Historical corn yield data is collected through Quick Stats from the National Statistics Service (NASS) for 37 years. Corn yield data is a yearly collected data at county level. The data include 99 counties in Iowa. Hence, there are 37*99=3663 records of historical yield data available, with 3267 of them acting as training samples. Due to genetic gain in corn growth, corn yield has increased through time. We adjust historical corn yield to the same base. Agricultural experts predict 1.5% annual yield increase. In Figure 3.1 below, Li (2014) shows corn yield trends in Iowa, indicating that genetic gain is almost an annual constant number and increases greatly after 2000. According to Li (2014), genetic gain is 2.5 bushels/acre per year from 1980 to 2000 and 4.67 bushels/acre per year after 2000.

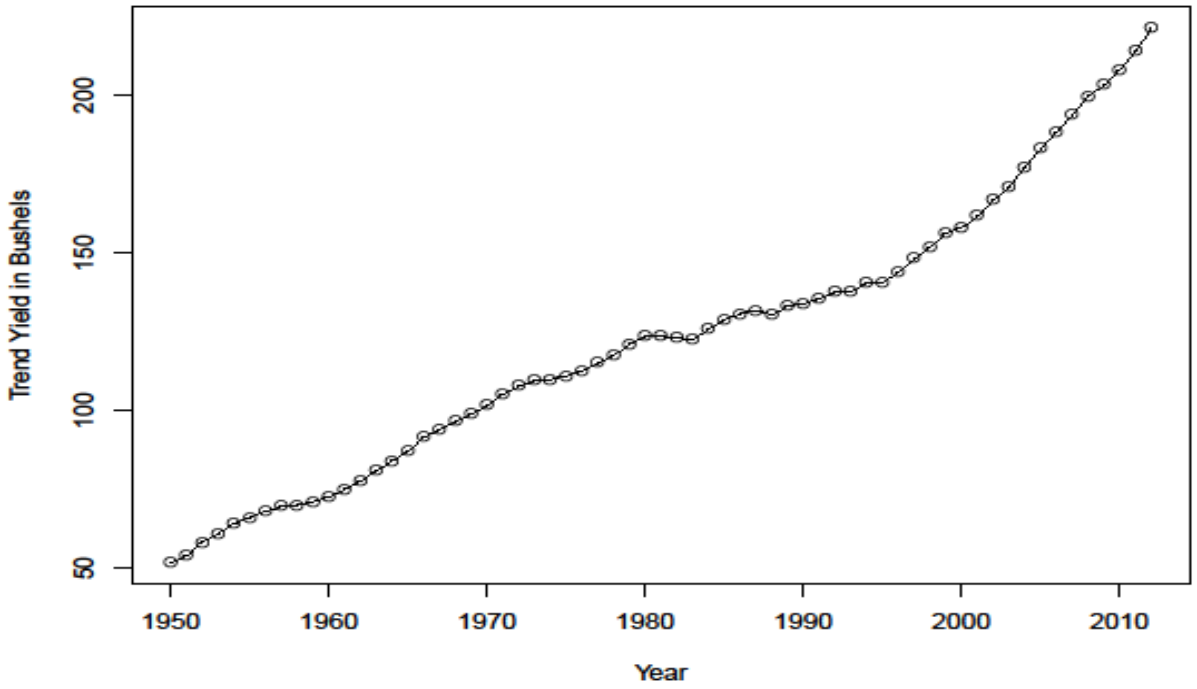


Figure 3.1: Corn yield trends in Iowa, 1950–2013

Note: Figure 3.1 taken from Li (2014).

We try both adjustments in our model; however that raises a few more concerns. Which year should the yield be adjusted to? Will that influence the prediction results? If all yields are adjusted to the 2015 base year, will the 2015 prediction be better than other years? To answer these questions, we train our model with corn yield adjusted to both 2013 and the 2015 base year. The results show that there is no evidence for such concerns.

3.2 Predictor Variables

Three types of input variables closely associated with corn yield are available - hourly weather data, soil quality data, and soil moisture data. Hourly weather data comes from Weather Underground, a professional weather data company, and is representative of a 19×19 mile area, which is more accurate than the commonly used weather station data. This is also our data

contribution to the literature. We utilize weather data from April to October because the corn growing season in the Corn Belt falls close to this period.

We choose raw data of three weather variables—precipitation, wind, and temperature. Precipitation is an important factor for corn growth. Enough rainfall in the growing season guarantees good yield, but floods ruin corn plants. High wind speed can damage corn crops by uprooting plants. Maximum, minimum, and mean daytime temperature strongly influence yield. Research shows that most plants do not grow any faster above 30°C. It is widely believed that temperatures between 50°F and 86°F (10°C and 30°C) are best for crop growth, hence they use a variable called Growing Degree Days (GDD) to measure the temperature effect on crop growth. GDD, which measure of heat accumulation to predict plant development rates, is calculated as follows:

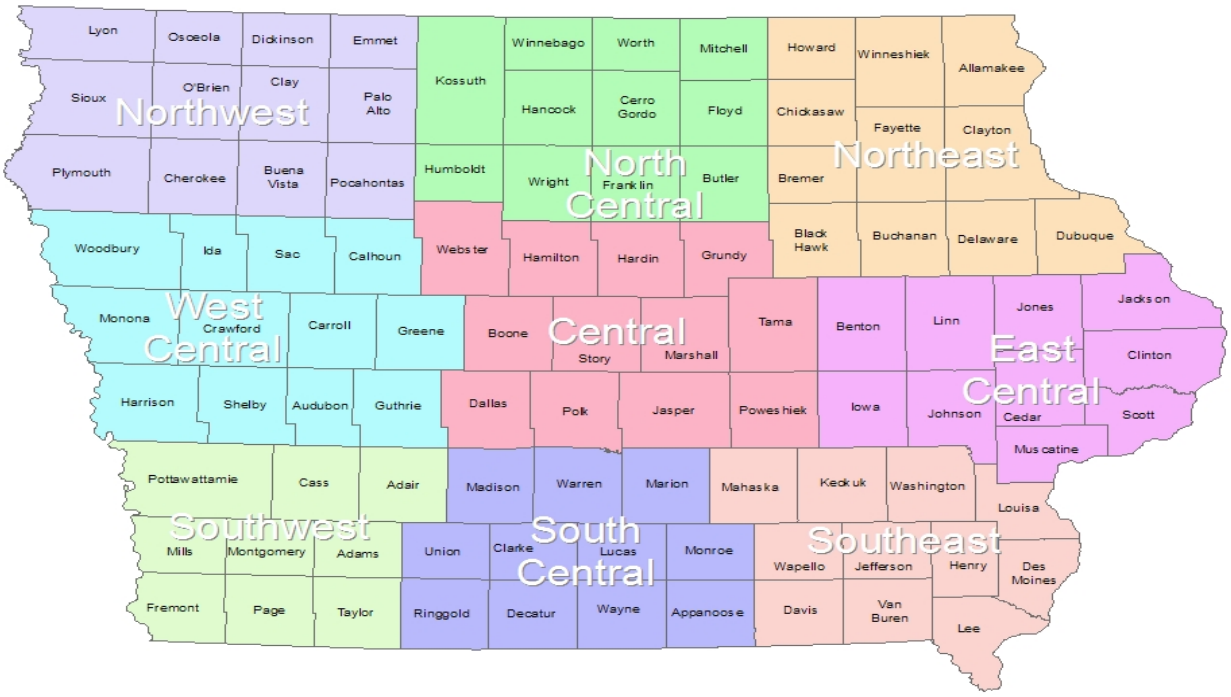
$$GDD = \frac{T_{\max} + T_{\min}}{2} - T_{base}$$

where $T_{\max}$ = min(86°F, daily maximum temperature),

$T_{\min}$ = max(50°F, daily minimum temperature), and T_{base} is the base temperature required to trigger optimum growth (50°F for corn). Accumulated GDDs during the growing season are an important factor in yield.

Soil moisture has critical impact on corn yield. The Palmer Drought Severity Index (PDSI) is a long-term cumulative measure of water availability in the soil that spans from -10 (dry) to +10 (wet), with zero being normal moisture conditions. PDSI uses temperature data and a physical water balance model to capture the basic effect of global warming on drought. PDSI is measured monthly by the National Oceanic and Atmospheric Administration at the Crop Reporting District

(CRD) level (see Figure 3.2). We match the counties within each district and assign the PDSI value to those counties (i.e., counties in the same CRD have the same PDSI value).



U.S. Department of Agriculture Crop Reporting Districts

Figure 3.2: Iowa Crop Reporting District map

Soil quality data comes from Dr. Hendricks at Kansas State University. Data is collected from the gSSURGO database (a database for storing gridded soil survey results) and aggregated to the county level using only areas classified as cropland according to the National Land Cover Database. Thus, the data covers the whole Corn Belt at the county level. The data has over 100 variables, and each variable is a constant number for each county, since soil quality is not considered to change much over time. We pick 14 variables from the data, which we think are

most related to corn yield (see Table 3.1). Rootznaws and droughty are considered the most two significant soil variables from expert knowledge.

Table 3.1: Soil Quality Variables List

Variables	Explanation
ffd	frost free days, number of days between the last freezing day in Spring and first freezing day after that
sandtotal	total sand (mineral particles 0.05mm to 2.0mm in diameter)
silttotal	total silt (mineral particles 0.002mm to 0.05mm in diameter)
claytotal	total clay (mineral particles less than 0.002mm in diameter)
om	weight percentage of organic matter (decomposed plant and animal residue)
bulkDensity	the oven-dry weight of soil material less than 2 mm
lep	linear expression of the volume difference of soil fabric and oven dryness
caco3	the quantity of Carbonate (CO3) in the soil expressed as CaCO3
ec	the electrical conductivity of an extract from saturated soil paste
soc0_150	soil organic carbon stock estimate (SOC) in standard layer
Rootznaws	root zone available water storage (mm)
droughty	soil droughty vulnerability determined by earthy major components
sand	percentage of sand contained in the soil
share_cropland	cropland share of the whole county land

3.3 Variable Selection and Data Preprocessing

The time series of input variables can be expressed both in hourly and daily format (weekly or monthly variable sequence does not this take into account as it averages too much information). Each county for each year is a sample record with output value—yield and the corresponding input time series falling into the growth period of April to October. Thus, the length of the input time series $\{x_t\}$ would be t=5136 for hourly input vectors or t=214 for daily input vectors. However, for hourly inputs vectors, there are too many parameters needed to estimate with only 3267 training samples; and, such long time series require large memory and computer resources for calculation,

which would be time consuming and lose information since computer memory is limited. Therefore, we take the daily input sequence $\{x_t\}$ with t=214.

The next question is variable selection. Even with the limited raw data we have, there are still 28 candidate input variables. Besides the 14 soil quality variables and PDSI, there are also max/min/mean day temperature, total daily rainfall, daily average wind speed, max day rainfall, up-to-date accumulated rainfall, and GDD. Since July has been proven to be the most important month for corn growth, rainfall and max temperature in July should be related. The ratio of acres planted to corn divided by the average acres planted may also influence corn yield as farmers may take less effort in farm management if they plant much less corn than average in a specific year. Moreover, two interaction terms (max temperature*soil droughty and max*PDSI) are also included. The idea is that for high temperatures the soil should be more humid than for low temperatures.

First, we trained the model with all 28 variables, but our prediction turned out to be an almost horizontal line. We then calculated the correlation between input variables and eliminated the highly correlated ones. We create a correlation matrix with each cell being the correlation coefficient. Correlation coefficient is computed by $\rho_{x_i,x_j} = \frac{\operatorname{cov}(x_i, x_j)}{\sigma_{x_i}\sigma_{x_j}} = \frac{E[(x_i - \mu_{x_i})(x_j - \mu_{x_j})]}{\sigma_{x_i}\sigma_{x_j}}$. If the absolute value of ρ_{x_i,x_j} is close to one, it means that these two variables are highly correlated. However, this method did not work with soil quality data as soil quality data is a constant number for each county, and correlation between constant numbers is pointless. Therefore, we use the minimum redundancy maximum relevance (mRMR) method to select the soil quality variables.

Minimum redundancy maximum relevance (mRMR) is a feature selection method first introduced by Peng, Long, and Ding (2005). It is computed in two parts:

Minimize Redundancy:

$$\min D_{I,} \qquad D_I = \frac{1}{|S|^2} \sum_{X_i, X_j \in S} I(X_i, X_j)$$

Maximize Relevance:

$$\max R_{I,} \qquad R_I = \frac{1}{|S|} \sum_{X_i \in S} I(X_i, Y)$$

where S is the set of features, and $I(X_i, X_j)$ is the mutual information between features i and j. Formally, the mutual information of two discrete random variables X and Y is defined as:

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right)$$

where $p(x, y)$ is the joint probability function of X and Y. Mutual information is widely used as a measure of the mutual dependence between the two variables X and Y. Maximum relevance means maximizing the average mutual information between features and the target. Past feature selection usually stops at identifying the best m features with the most individual mutual information to target Y. In fact, different features may share redundant information on the target. Thus, redundancy is another important factor to be considered in feature selection. In order to characterize the most relevant subset of features whose size is limited to a given factor, we need to balance the tradeoff between relevance and redundancy. Thus, the mRMR criterion is a combination of the two above measures and defined as:

$$mRMR = \max_{S_m \subset S} \left[\frac{1}{|S|} \sum_{X_i \in S} I(X_i, Y) - \frac{1}{|S|^2} \sum_{X_i, X_j \in S} I(X_i, X_j) \right]$$

Induction is used for searching the mRMR candidate feature set. Suppose we already have S_{m-1}, in the feature set with $m-1$ features, then the task is to select the m^{th} feature from the set $X \setminus S_m$; that is, to optimize the condition:

$$\max_{x_j \in X \setminus S_{m-1}} \left[I(x_j, Y) - \frac{1}{m-1} \sum_{x_i \in S_m} I(x_j, x_i) \right]$$

Thus, we get different sequential feature sets $S_1 \subset S_2 \subset \ldots \subset S_{n-1} \subset S_n$. After comparing all the sequential feature sets, we find the range of $k, 1 \le k \le n$, called Ω, within which the classification error is constantly small. Within Ω, we find the smallest classification error $e^* = \min e_k$. The optimal size of the candidate feature set, n^*, is chosen as the smallest k that corresponds to e^*. We then apply backward or forward selection to add or remove features from S_{n^*}, with the selection based on the criterion to make the largest classification error reduction. Finally, we can rank the ordered features. The higher the rank is, the more relevance the feature has. Typically, mRMR works with categorical target variables, hence we need to transform the yield variable into a categorical format at the start of the mRMR selection.

Referring to the results of the correlation matrix and mRMR feature selection, after trial and error with different combinations of input variables and expert knowledge, we figure the 10 "best" input variables for corn yield prediction: each of max/min/mean temperature, total daily rainfall, wind speed, soil root space for holding water, soil droughty, PDSI, cumulative rainfall, and GDD.

Even when using a daily input series, 3267 training samples are not enough. Thus, another contribution we make to the literature is finding a creative way to generate more training samples. By picking two or three counties from the same CRD in Iowa and using the average of their yield and input variables respectively, we create a new sample. There are nine CRDs in Iowa; hence, the

total number of training samples combined with combination samples increases to 70,026. This number is computed through $3267+33\times(\sum_{i=1}^{9}C(k_i,2)+\sum_{i=1}^{9}C(k_i,3))$, where k_i is the number of counties in each CRD, where $\{k_i\}=\{12,11,11,12,12,10,9,11,11\}$ for Iowa. These combination samples should be reasonable since PDSI is also collected at the CRD level and all other data are also averaged number for county area (the most precise data point should be each farmland, which is not available).

Ten input variable sequences are stored in the format of a 3D tensor cube (see Figure 3.3 for an example of a 3D tensor cube). This is a key step to make our data fit into the model. All kinds of input data should be converted into 3D tensor format for LSTM training in computers. Figure 7 is an example of 3D tensor cube. X-axis indicates the number of input variables, Y-axis is the length of the time series, and Z-axis is the number of samples. Hence the dimension of our 3D tensor cube is $10\times214\times70026$. The matrix we face is a sample of input variables $\{x_t\}$ with the dimensions 10×214, where x_t = (mean temperature, rainfall, windspeed, PDSI, rootznaws, droughty, cumulative GDD, cumulative rainfall, max temperature, min temperature) for t=1,…,214, which is the row vector of the matrix. PDSI is monthly data, so it repeats once for each day in each month, while rootznaws and droughty repeat 214 times since they are constant.

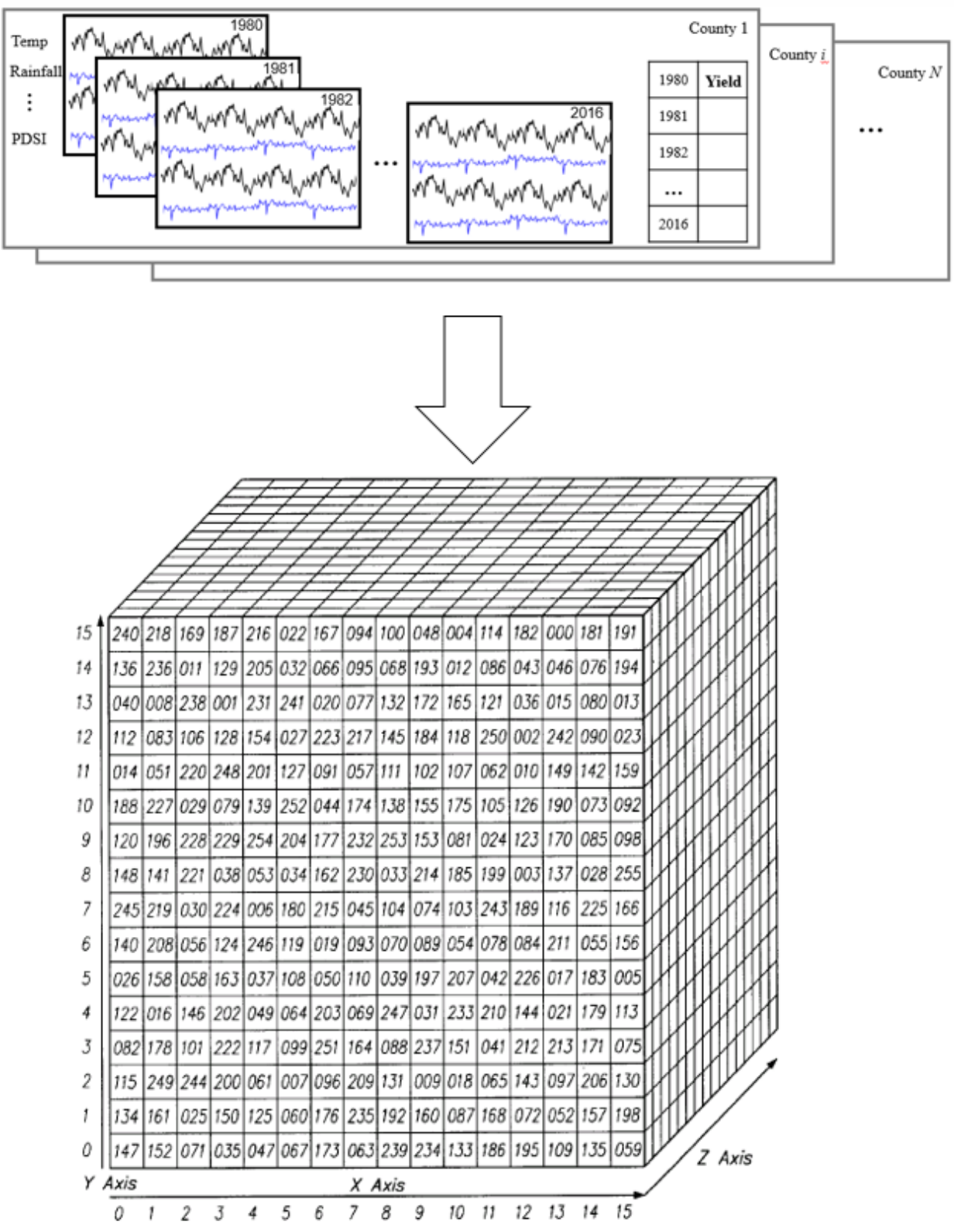


Figure 3.3: 3D tensor cube example

CHAPTER 4. METHODOLOGY

Chawla et al. (2016) use Bayesian Network and data from 2005 to 2009 to predict Iowa's 2010 county-level corn yield. We follow this method and extend their work using more predictors and 35 years of county-level data across five Corn Belt states (Illinois, Indiana, Iowa, Minnesota, and Nebraska). As BN works with categorical response variable, corn yield is needed to be discretized at the beginning. We use ChiMerge, an algorithm introduced by Randy (1992), to convert numeric values into discrete values. The network is trained through Genie introduced by Druzdzel (1999). Though BN could successfully predict the yield in the right category, the results (see the Appendix) show that the performance is bad for extreme cases. BN has obvious disadvantage in yield prediction. First, BN works with categorical data, and while corn yield is continuously variable, discretizing the yield into intervals will make the accuracy too low. Moreover, the way Chawla et al. (2016) calculate the actual yield number, which is sums the expected average, does not distinguish the difference in weather predictors. This method leads to yields in the same category having the same calculated actual yield predictor. Thus, in order to get higher accuracy in corn yield prediction, we look for a machine learning method to work with continuous time series data.

First, we tried an unsupervised method. The inspiration came from Akintayo and Sarkar (2015). There are also several applications for reference, such as Liu (2017) and Jiang (2017). We use their method for classification of a slow time epoch to compute the probability representing the similarity of weather conditions between testing and historical years in different counties. Our problem shares similar structure with this paper. The input variable data is hourly time series data for each year in each county as described in the data section. Due to the format of data, our task fits a classification problem. Each input time series data is first transited into symbolic sequences

using statistical similarity-based discretization and maximally bijective discretization. Then the probability can be obtained by the following expression:

$$\Pr(\tilde{S} \mid S^i) = \prod_{m=1}^{|Q|} \frac{(\tilde{N}_m)!(N_m^i + |\Sigma| - 1)!}{(\tilde{N}_m + N_m^i + |\Sigma| - 1)!} \times \prod_{n=1}^{|\Sigma|} \frac{(\tilde{N}_{mn} + N_{mn}^i)!}{(\tilde{N}_{mn})!(N_{mn}^i)!}$$

where $\tilde{S}$ is the testing symbolic sequence and S^i is the historical symbolic sequence. Each symbol string S^i is represented as $s_1^i s_2^i \ldots s_{N_i}^i$, so we take the neighboring two symbols as a state. This follows a D-Markov machine where $|Q|$ is the number of states, $|\Sigma|$ is the column number of the Markov transition matrix, and N_{mn}^i is the number of times a symbol σ_n in S^i is emanated from the state q_m. Then the number of occurrences of the state q_m in the state sequence is given by $N_m^i = \sum_{n=1}^{|\Sigma|} N_{mn}^i$, where $\tilde{N}_m$ and $\tilde{N}_{mn}$ represent the same number in the sequence $\tilde{S}$. More details of the formulation can be found in Akintayo and Sarkar (2015). We match our target year of a specific county with a training year by computing the probabilities for the testing sequence with all historical sequences and picking the highest year. For example, Madison County 2015 has the highest probability with Wapello County 1996. Then we take the yield of Wapello County 1996, adjusted with genetic gain, as our predicted yield for Madison County 2015. However, the method fails as we cannot use different weights to indicate the contribution of the input variables to the yield through time. For example, weather conditions in July are widely believed to have the most significant influence on corn yield; however, this method does not have the capability to capture this important expert knowledge. Though the method has fatal flaws, we still think that it is a good way to refine the pattern and characteristics of the input variables and may be a promising method for data preprocessing of our final model. We believe the improvement of our model can be discovered from here.

Finally, we turn to a supervised method—RNN. RNN is a family of neural networks for processing sequential data that has become very popular in deep learning for text prediction and speech recognition. RNN is a black-box method, which is powerful in handling nonlinearity and interaction relation but falls short of causal explanation. Corn yield depends on a complex set of economical, meteorological, agricultural, and financial inputs, which makes prediction very difficult. The advantage of RNN in learning complex interdependent relations between inputs and outputs makes us believe that it will perform well in crop yield prediction.

This paper is the first to apply RNN to crop yield prediction and shows the power and possibility of RNN in this area. The most difficult part is how to fit our problem into the format of RNN. Even though RNN was created for time series data, previous applications only focus on the prediction of the next following points in the same time series. This paper indicates a novel way that RNN can be used to solve crop yield prediction.

4.1 Recurrent Neural Network

Figure 4.1 shows the representation of a regular RNN. The left side is the simple representation and the right side is the same model unfolded. The network consists of three layers: input, hidden, and output, where $x^{(t)}$ is the input sequence, $y^{(t)}$ is the output sequence, and $h^{(t)}$ is a series of hidden states. The number of hidden layers is not constrained to one. In the deep learning recurrent neural networks, the number of hidden layers can reach eight or more. Adding hidden layers can help to study the more complex structure of the model, but also requires more data. U, V, W are shared weights that we need to learn, and f is an activation function where $h^{(t)} = f(Ux^{(t)} + Wh^{(t-1)})$. The corn yield prediction problem could not fit into a regular RNN, so we use the many-to-one RNN model here. The many-to-one RNN model is suitable when there is sequence input with one output, thus it is perfectly match with our data described in the data section.

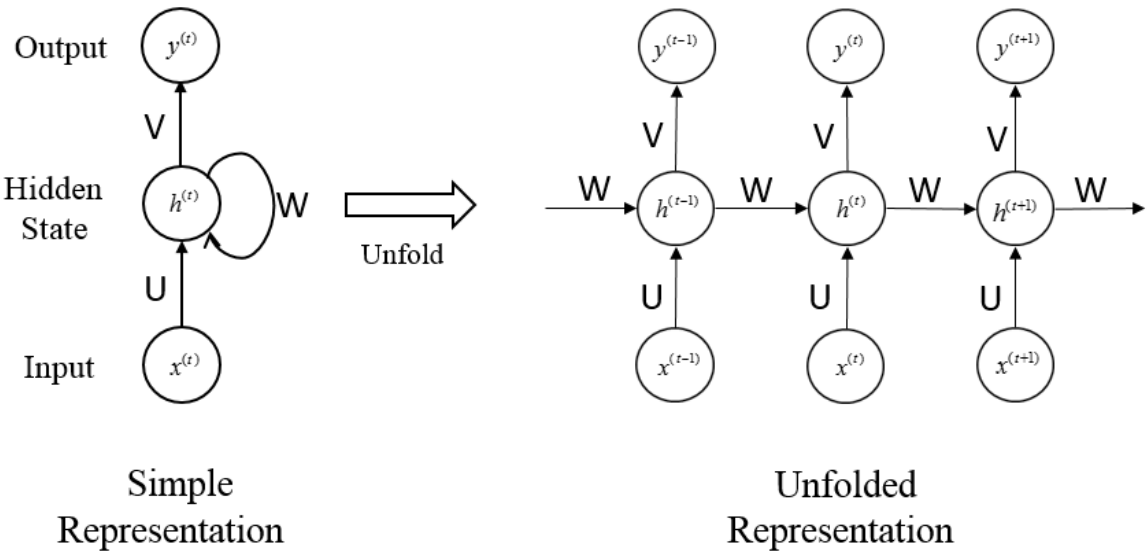


Figure 4.1: Recurrent Neural Network representation

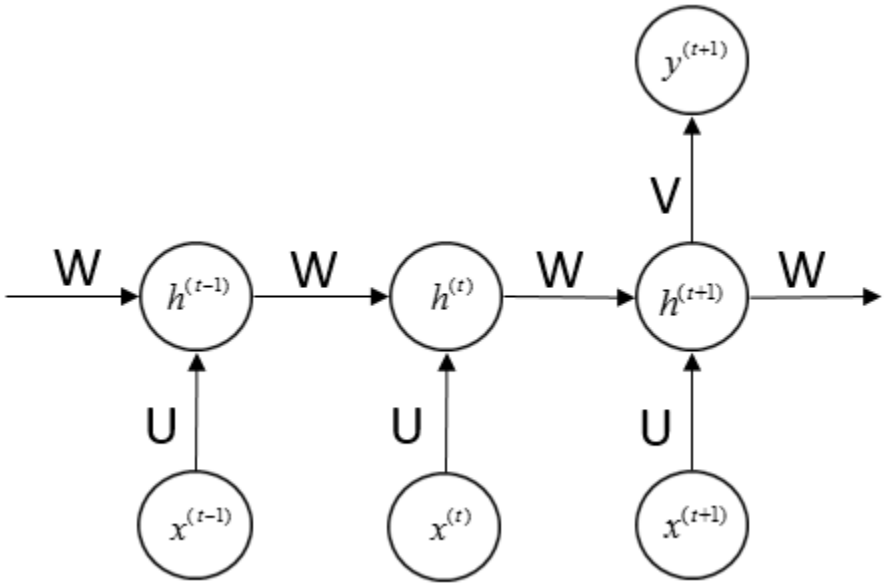
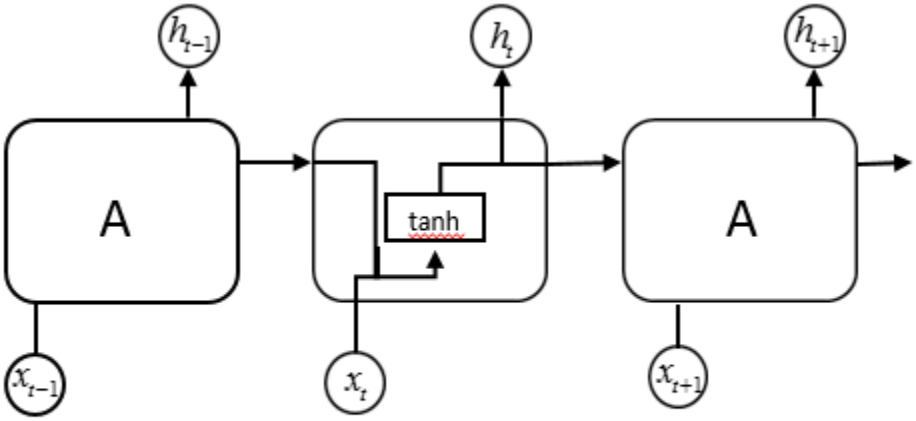


Figure 4.2: Many-to-one RNN

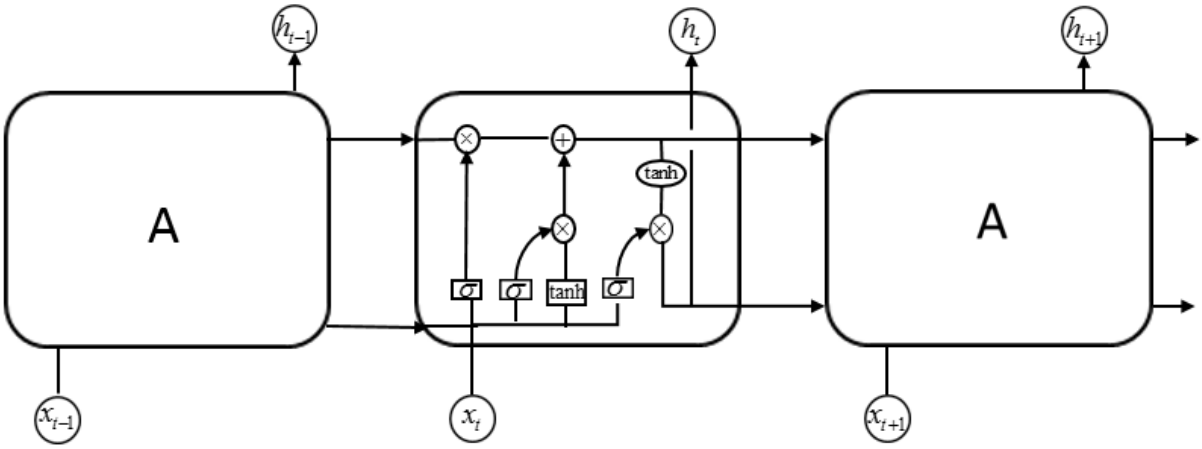
4.2 Long Short-Term Memory (LSTM)

The mathematical challenge of learning long-term dependencies in recurrent networks is called the "vanishing gradient problem." As we propagate the gradient back in time, the magnitude quickly decreases. That is to say, as the input sequence gets longer, it becomes more difficult to capture the influence from the first stage. The gradients to the first several input points vanish and are approximately equal to zero (rarely the gradients will explode with much damage to the optimization). Therefore, a special RNN model called Long Short-Term Memory (LSTM) was developed. LSTM uses the identity function with a derivative of one. As a result, the back propagated gradient remains constant instead of vanishing or exploding when passing through. Figure 4.3 shows the difference in the framework between regular RNN and LSTM, where tanh is a commonly used activation function (can be any other functions such ReLU or sigmoid). Clearly LSTM has a more complex structure to capture the recursive relationship between the input and hidden layer. We call the cell between the input and hidden layer LSTM cell.

LSTM adds a new sequence $\{c_t\}$ called cell state to RNN. Cell state is a space specifically designed for storing past information (i.e., the memory space) that mimics the way the human brain manipulates information when making decision. The left part of the cell in Figure 4.4 is the forget gate layer, which makes the decision whether past information stored in the cell state should be discarded or not. The middle is the input gate layer, which decides whether new information from the input should be added or not. The operation is executed to update old cell state c_{t-1} to c_t. This is when old information is dropped and new information added. We can get the output as h_t by operating the right part, which is the same process as regular RNN.



Regular RNN



Long Short Term Memory

Figure 4.3: Comparison of regular RNN and LSTM

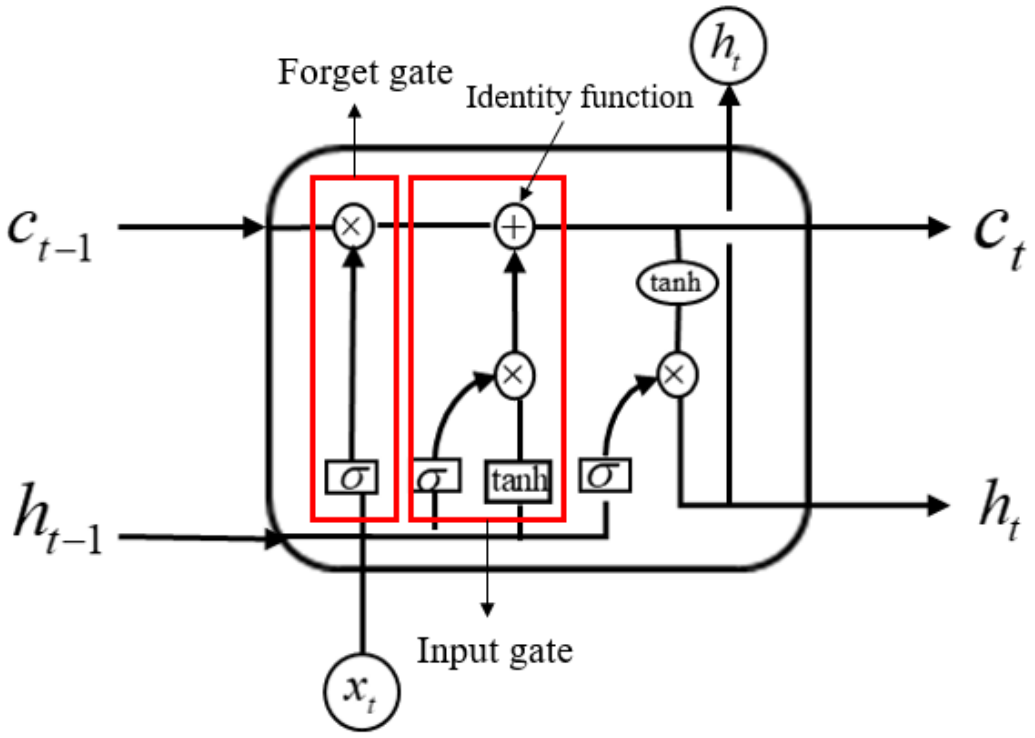


Figure 4.4: Details of an LSTM cell

In conclusion, the behavior of the memory cell is determined by three gates: input i_t, output o_t, and forget f_t. The updated equations are as follows:

$$i_t = sigmoid(W_i h_{t-1} + U_i x_t + b_i)$$

$$f_t = sigmoid(W_f h_{t-1} + U_f x_t + b_f)$$

$$o_t = sigmoid(W_o h_{t-1} + U_o x_t + b_o)$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$h_t = o_t \odot \tanh(c_t)$$

where all $U \in \mathbb{R}^{d \times d}, W \in \mathbb{R}^{d \times k}, b \in \mathbb{R}^d$ are learnable parameters and the operator $\odot$ denotes the element-wise multiplication.

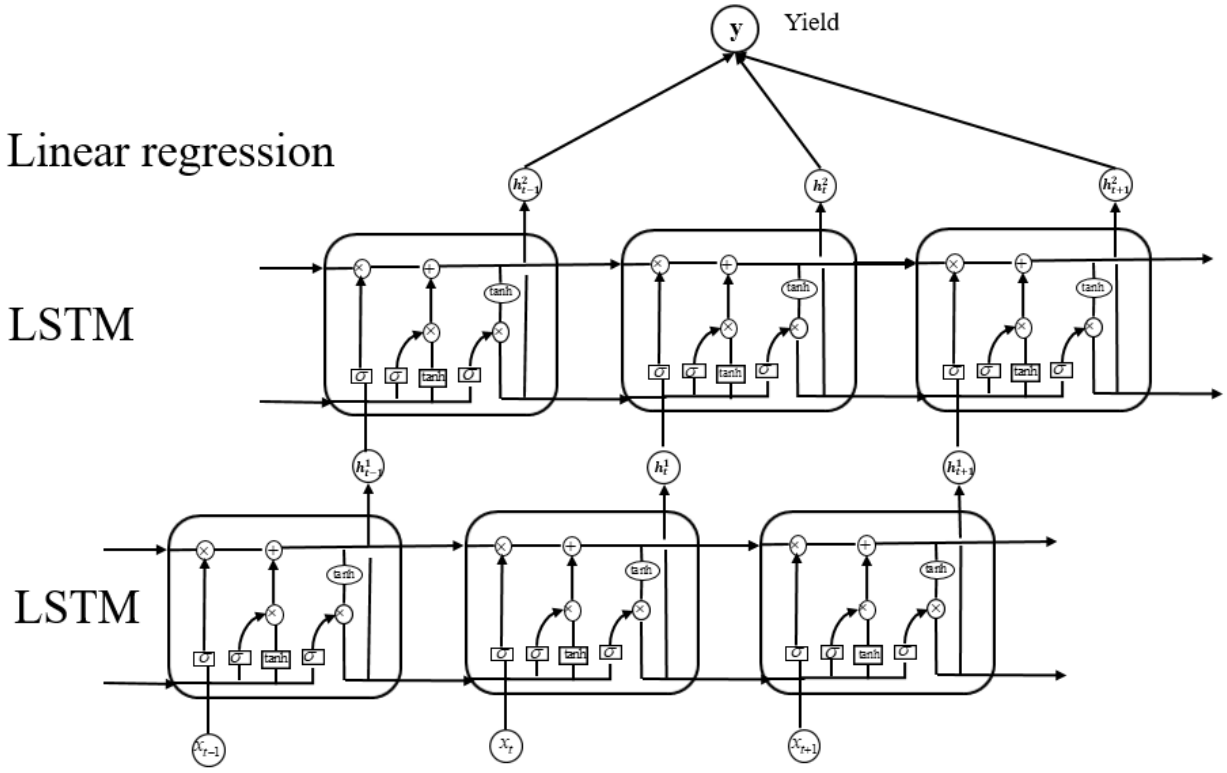


Figure 4.5: Structure of final LSTM model

Figure 4.5 shows the structure of our final LSTM model used for county-level corn yield prediction in the Corn Belt.

4.3 Training of LSTM

Loss function is a measure of how good a prediction model does in terms of being able to predict the expected outcome. The loss function we picked for our LSTM model is the mean squared error (MSE). The target of training the prediction model is to find parameters that could achieve the minimum point of the loss function, thus turning it into an optimization problem. The algorithm to learn the recurrent neural network is gradient descent and back-propagation through time (BPTT). Gradient descent is one of the most popular algorithms to perform optimization. It is an efficient algorithm to search for the local minimum of the loss function. The BPTT algorithm is used to compute the gradient for the equation $h^{(t)} = f(Ux^{(t)} + Wh^{(t-1)})$ and the loss function. The

core idea behind BPTT is the composite function chain rule. The nodes of our computational graph include the parameters U, V, W and constant terms as well as the sequence of nodes indexed by t for $x^{(t)}$ and $h^{(t)}$. Once the gradients on the internal nodes of the computational graph are obtained, we can obtain the gradients on the parameter nodes. The parameters are shared across time steps. Given a starting point, calculating the gradient of that point and searching in the direction of the negative gradient is the fastest way to search for a local minimum. Then we can update the parameters with iterations of the gradient descent optimizer by searching for a smaller local minimum.

Our LSTM model was learned using the Keras Python package on top of Theano backend. We assigned a linear relation between the hidden and output layers. There are several choices for gradient descent optimizer in Keras. We tried both Stochastic Gradient Descent (SGD) and RMSprop optimizer. Our final LSTM model used RMSprop optimizer as this optimizer is usually a good choice for RNNs. Besides the parameters that we need to learn from the data, there is also another kind of parameter specified manually for LSTM models, called a hyperparameter. A model hyperparameter is a configuration that is external to the model whose value cannot be estimated from data. Hyperparameter searching is an important process before the commencement of the learning process. The choice of the hyperparameter influences the learning result.

The hyperparameters that we decided manually for our LSTM model include the number of hidden nodes within each hidden layer, batch size, dropout rate, learning rate, momentum, and decay rate. Batch size is the number of training examples utilized in one iteration of SGD or RMSprop optimizer. The higher the batch size, the more memory space needed. Dropout is a technique where randomly selected neurons are ignored during training. A 0.2 dropout rate means that one in five hidden neurons will be randomly excluded from each updated cycle. Dropout could

make the network less sensitive to the specific weights of neurons, and in turn solve the overfitting problem. Learning rate, momentum, and decay rate are important parameters for SGD optimizer. They would decide the speed of convergence of the network. The learning rate is how quickly a network abandons old beliefs for new ones. With a large learning rate, we take huge jumps to reach the bottom. There is also a possibility that we will overshoot the global minima (bottom) and end up on the other side of the pit instead of the bottom. Thus, we will never be able to converge to the global minima, but instead wander around. However, it will take too much time to converge if the learning rate is too small. Hence, it is often useful to reduce the learning rate as the training progresses, which is what the decay rate is used for. Momentum is an argument in SGD optimizer to obtain faster convergence. RMSprop optimizer is similar to the SGD optimizer with momentum. It uses a moving average of squared gradients to normalize the gradient itself. We only need to define the learning rate for RMSprop optimizer.

Table 4.1: Hyperparameters Choice Set List

Hyperparameters	Choice Set	Optimizer
Hidden nodes	[8, 16, 32, 64, 128, 214]	SGD, RMSprop
Batch size	[16, 64, 128, 512, 1024]	SGD, RMSprop
Dropout rate	[0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7]	SGD, RMSprop
Learning rate	[1e-07, 1e-06, 1e-05, 1e-04, 0.001]	SGD, RMSprop
Momentum	[0.0001, 0.001, 0.01,0.05, 0.1]	SGD
Decay rate	[0.0001, 0.001, 0.01,0.05, 0.1]	SGD

Nobody has the ability to know the best value for a model hyperparameter of a given problem. We may use rules of thumb, copy values used on other problems, or search for the best value by trial and error. What we did was assign a set of numbers by experience for these hyperparameters and let the machine randomly pick one value in the set for each hyperparameter.

The choice set is listed by each hyperparameter in Table 4.1. Usually after searching for over 300 models with different combinations of hyperparameter settings, we can find the 'best' model and the corresponding 'best' hyperparameters.

4.4 Model Settings

There are still some questions about the model settings. Should we use one, two, or more hidden layers? Will more related input variables improve the prediction? Will more created training samples generated with combinations described in the data section improve the prediction? We tried all these different settings and discuss the results in the next section.

CHAPTER 5. RESULTS AND DISCUSSION

Our initial model starts with an hourly input vector $\{x_t\}$ where t=5136 and yield is adjusted to both 2013 and 2015 base with 1.5% yearly increase. We have 3267 samples and one hidden layer. The input variables are hourly temperature, rainfall, wind speed, PDSI, soil root space for holding water, soil droughty, accumulative rainfall, and GDD by hour. Figure 18 shows the prediction results for this original model. The left part is the results for yield adjusted to the 2013 base while the right is adjusted to 2015. The black line is our prediction and the red line is true yield. There is also a picture of the absolute error between the prediction result and the true yield. Agricultural experts are convinced that it will be amazing if a model without farm management data can make over 8% of the county-level yield prediction fall into the +/-20 region of the true yield. Therefore we include the absolute error as a standard line to judge the performance of the model. After comparing the results between the left and right part, we conclude that whichever year the yield is adjusted to, there will be hardly any influence on prediction results. Therefore, we uniformly adjust the yield to 2013 base for all models.

Due to the cons of hourly input vectors introduced in section 3.3, we trained the initial model with daily input vectors. All the model settings are the same as the hourly input vector model except two more input variables—max and min day temperature—are added. Absolute error in Figure 19 indicates that the performance of this model is good. However, the prediction line does not match the true yield curve well. We try to improve the model with three changes in the settings: (*a*) using two hidden layers instead of one; (*b*) adding more training samples created with the combination method (two+three counties average); and, (*c*) including more input variables (variable selection from the 28 variables introduced in section 3.3). Table 5.1 shows that the 10 "best" input variables with two hidden layers and combination samples has the smallest MSE.

Figure 5.7 indicates the performance of this "best" model. Nevertheless, the fluctuation of prediction is still less than the true yield. Does this average trend exist because of including too many combination samples? To be sure, we trained the "best" model again with only two county combination samples added, which totals 19734 samples. And we also use the constant genetic gain adjustment with yield data.

Table 5.1: Comparison of Different LSTM Model Settings

Number of input variables	Number of samples	Hidden layers	Mean squared error
10	3267	1	255.404
	70026	1	211.6265
	3267	2	233.2547
	70026	2	191.0535
15	70026	2	361.9132
16	70026	2	Very large
28	70026	2	Very large

Table 5.2: Details of Different Selected Sets of Input Variables

Number	Input variables	Selection criterion
10	Refer to section 3.3	mRMR, expert knowledge and trial and error
15	Mean temperature, rainfall, wind speed, PDSI, rootznaws, droughty, accumulative gdd, acre_share, frost free days, total clay, organic matter, electrical conductivity, max rainfall, rainfall in July, max temperature in July	Correlation matrix
16	All input variables excluding the twelve soil quality variables except rootznaws and droughty	mRMR
28	Refer to section 3.3	all input variables

Table 5.3: Comparison of State-Level Predictions of Different LSTM Model Settings

year	yield	NASS	1 layer	1 layer com	2 layers	2 layer com
2013	164.00	169	157.59	160.80	165.39	161.15
2014	178.00	183	185.27	180.21	177.62	185.12
2015	192.44	189	188.26	185.39	180.65	184.71

Table 5.4: Comparison of State-Level Predictions with Two "best" Models

year	yield	NASS	percentage adjustment	constant adjustment
2013	164.00	169	171.33	165.57
2014	178.00	183	179.13	184.13
2015	192.44	189	192.22	190.50
2016	203.04	199	189.13	195.45

Figure 5.8 shows the prediction with two combinations of samples and constant adjustment. We can see that our prediction perfectly captures the fluctuation of the yield. Tables 5.3 and 5.4 compare the prediction results at the state level. NASS regularly report their state-level yield prediction every August, September, October, and November. Here we compare with their November prediction (final prediction). For 2013, constant adjustment has the best prediction, for 2014 and 2015, percentage adjustment perfectly predicts the yield while NASS and constant adjustment has almost the same performance. However, for 2016, our predictions are much lower than the actual yield, which indicates that the genetic gain adjustment may need to be higher for year 2016.

We also observed that the absolute predicting error was very large for some counties. Do these counties have common properties? Are they neighboring counties? Hence we marked the best and worst five counties in yield prediction for each year in the map. The results are presented in Figures 5.1-5.4. The best five counties were filled with green, while the worst five counties were

filled with red. After comparing these four figures, there is no obvious trend or commonality among these counties. This indicates that we did not miss any important explanatory variables. Prediction accuracy was not influenced by the geographical position of the county. The reason for bad county-level prediction is complex. One explanation is that maybe our prediction is more close to the true yield. The actual yield data downloaded from the USDA website is not the true yield at the county level. The USDA only collect corn yield data at the state level, and they distribute the data into county level according to planting area information. Another possible issue is that our LSTM model could not capture extreme cases such as drought or flood. Thus, if extreme conditions occurred, the prediction for that specific county could be inaccurate.

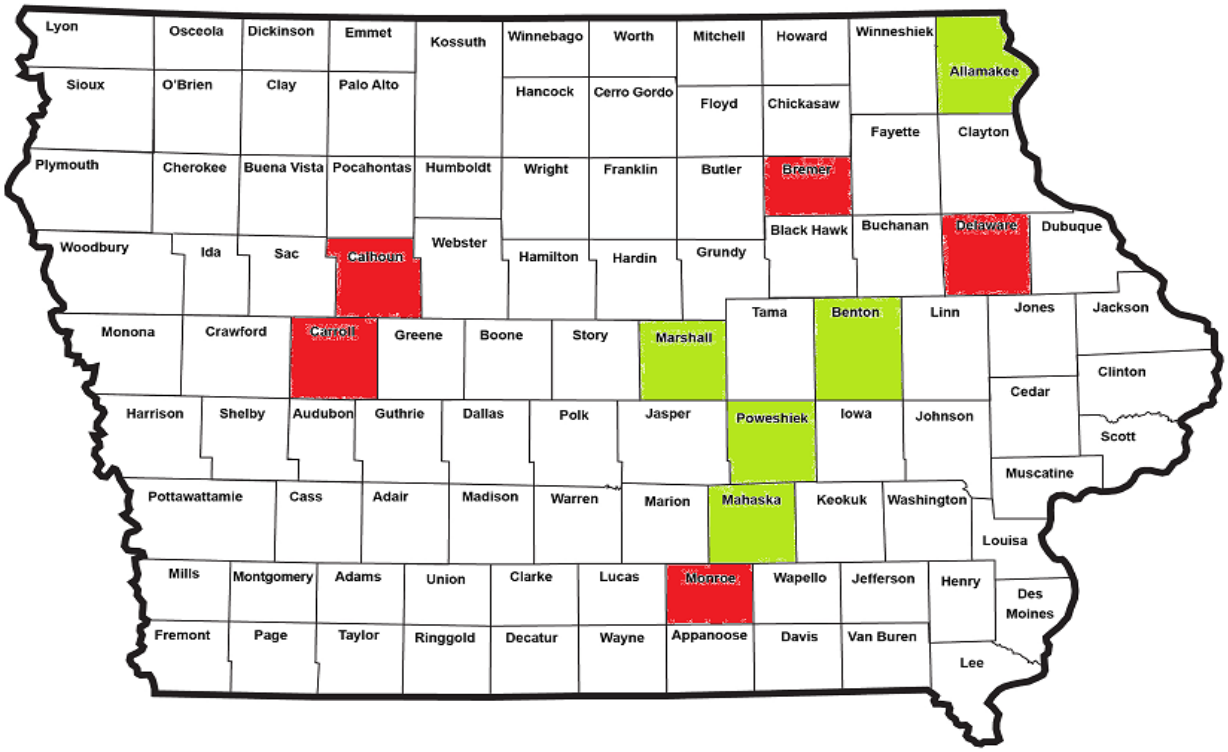


Figure 5.1: Five best and worst counties, 2013

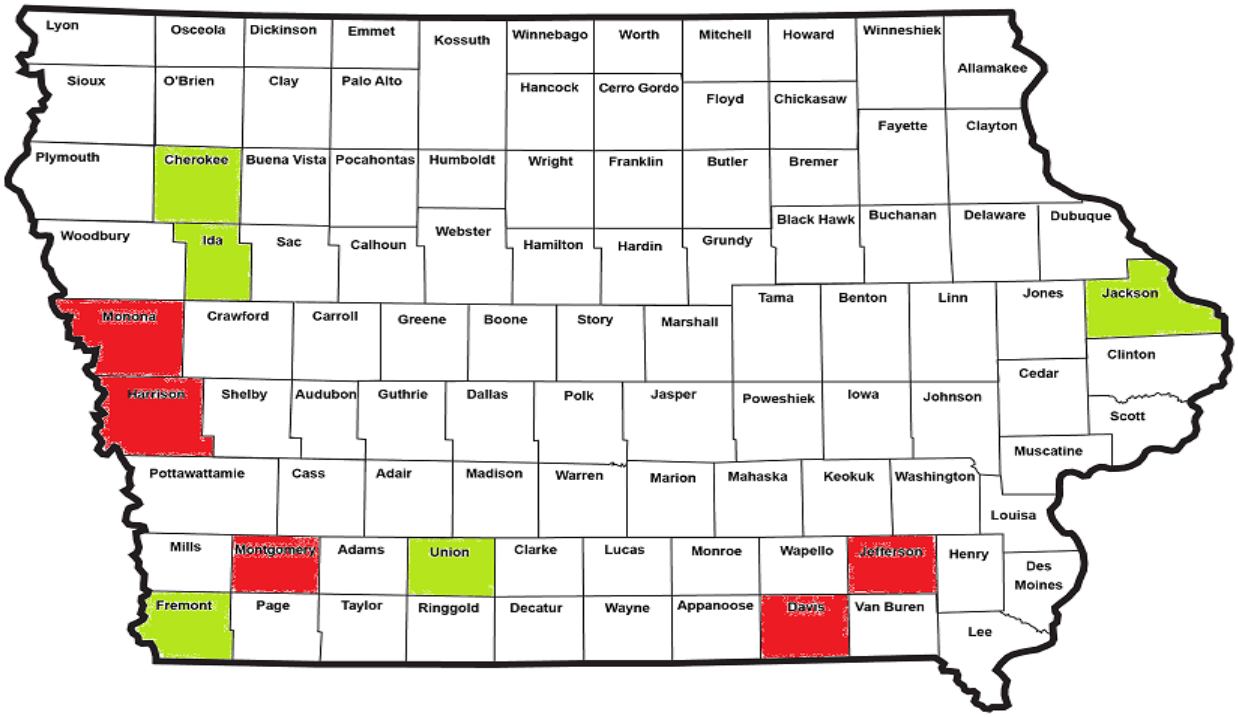


Figure 5.2: Five best and worst counties, 2014

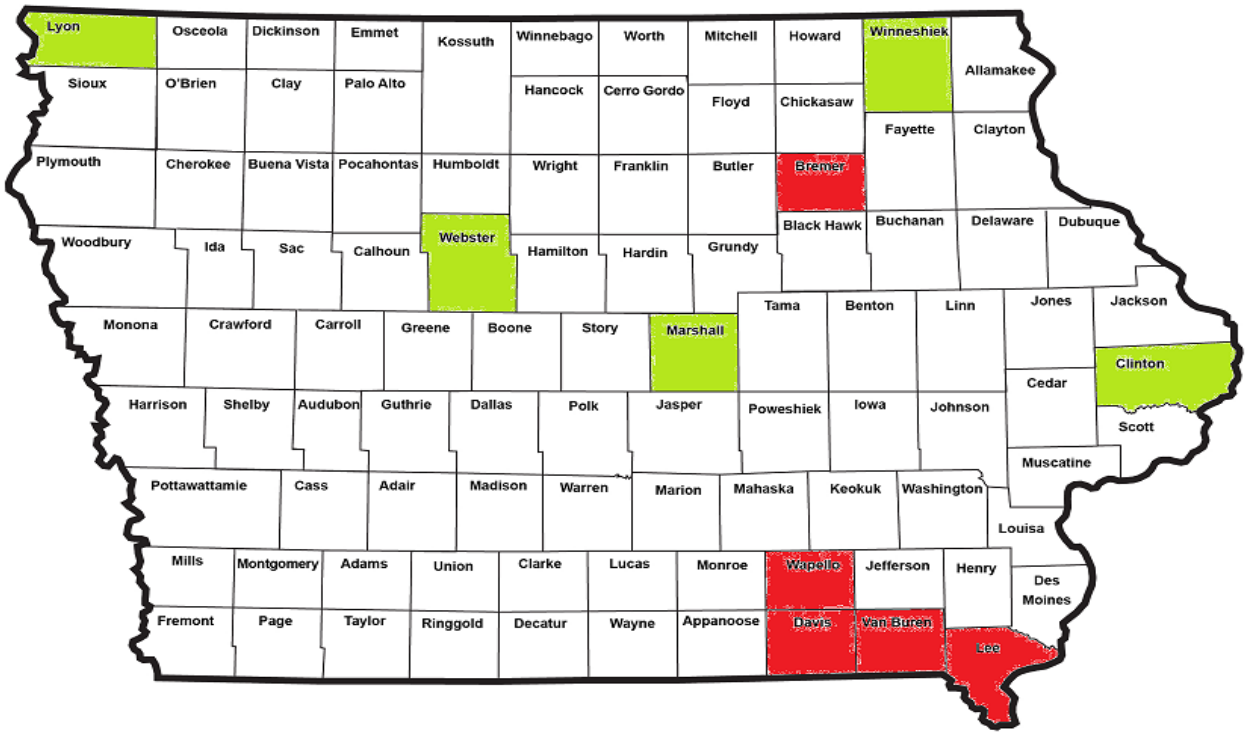


Figure 5.3: Five best and worst counties, 2015

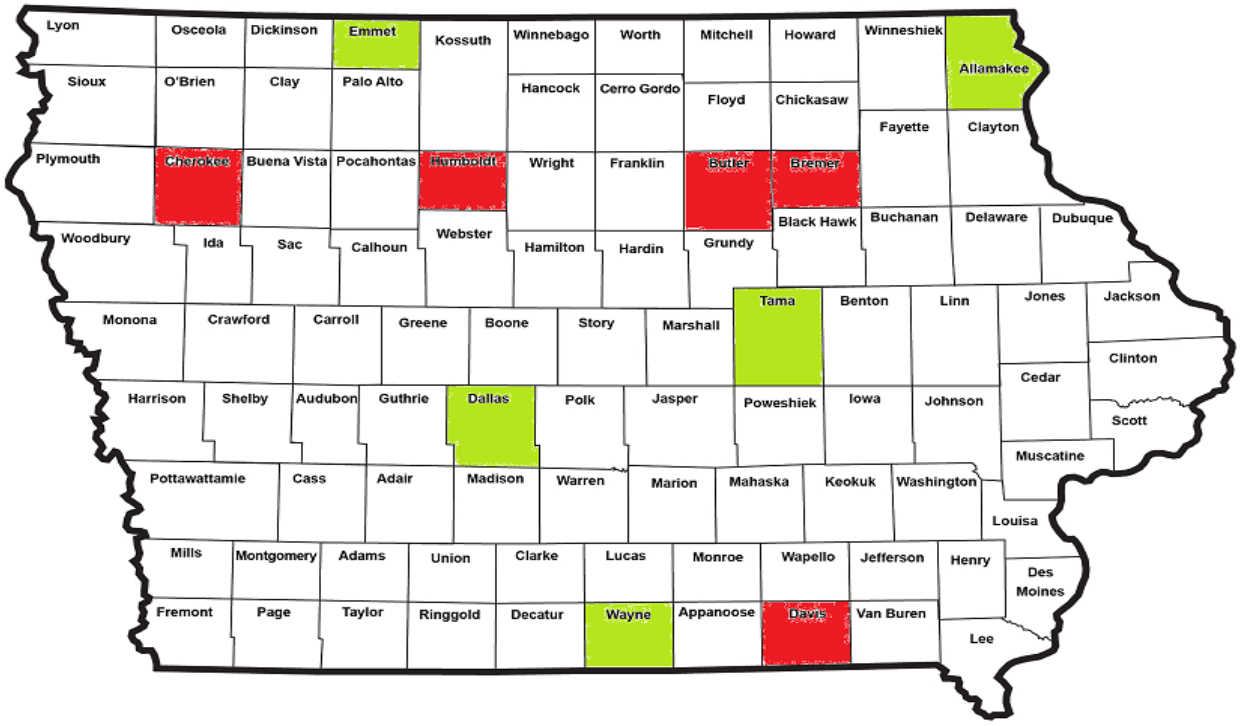
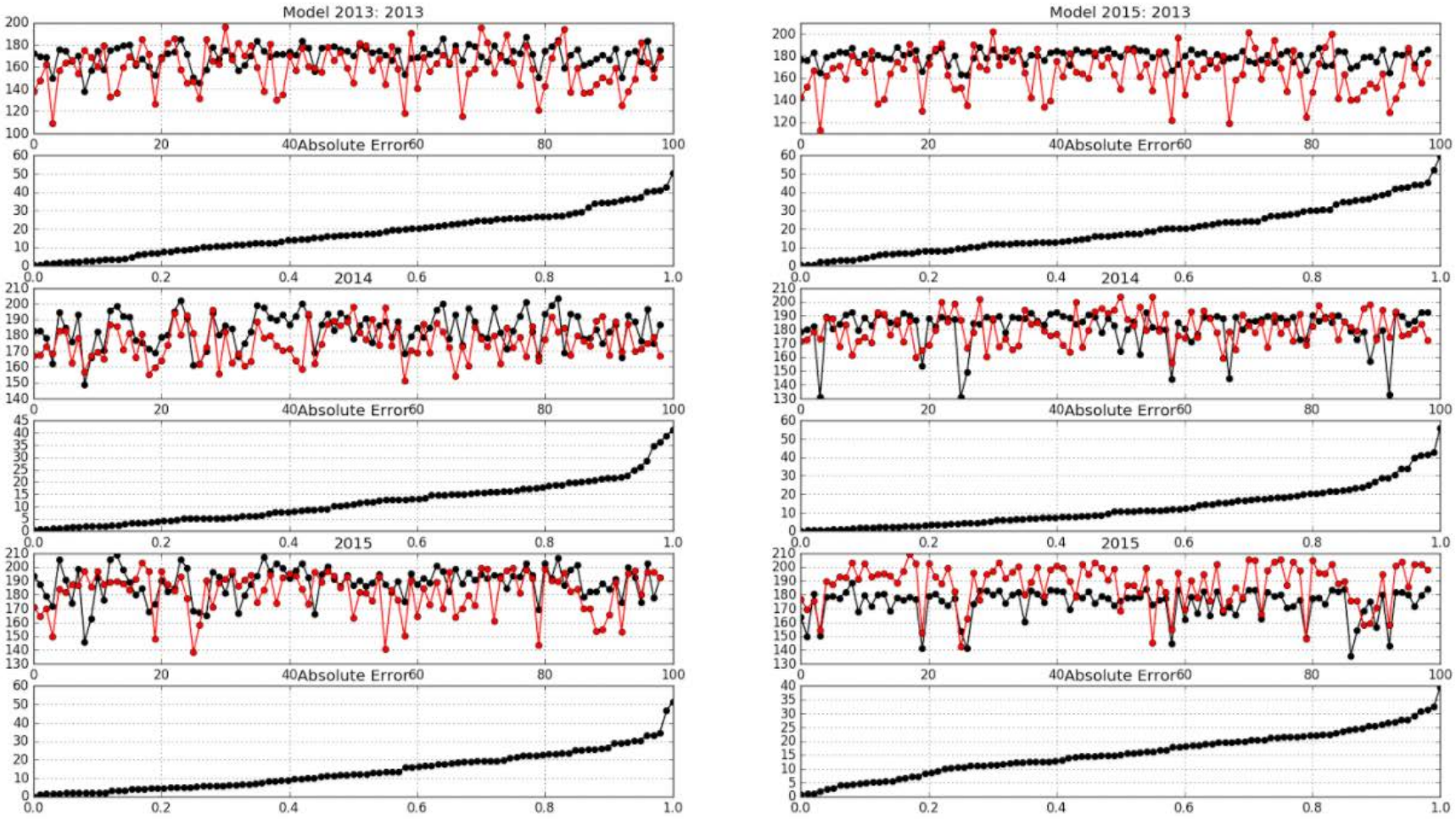
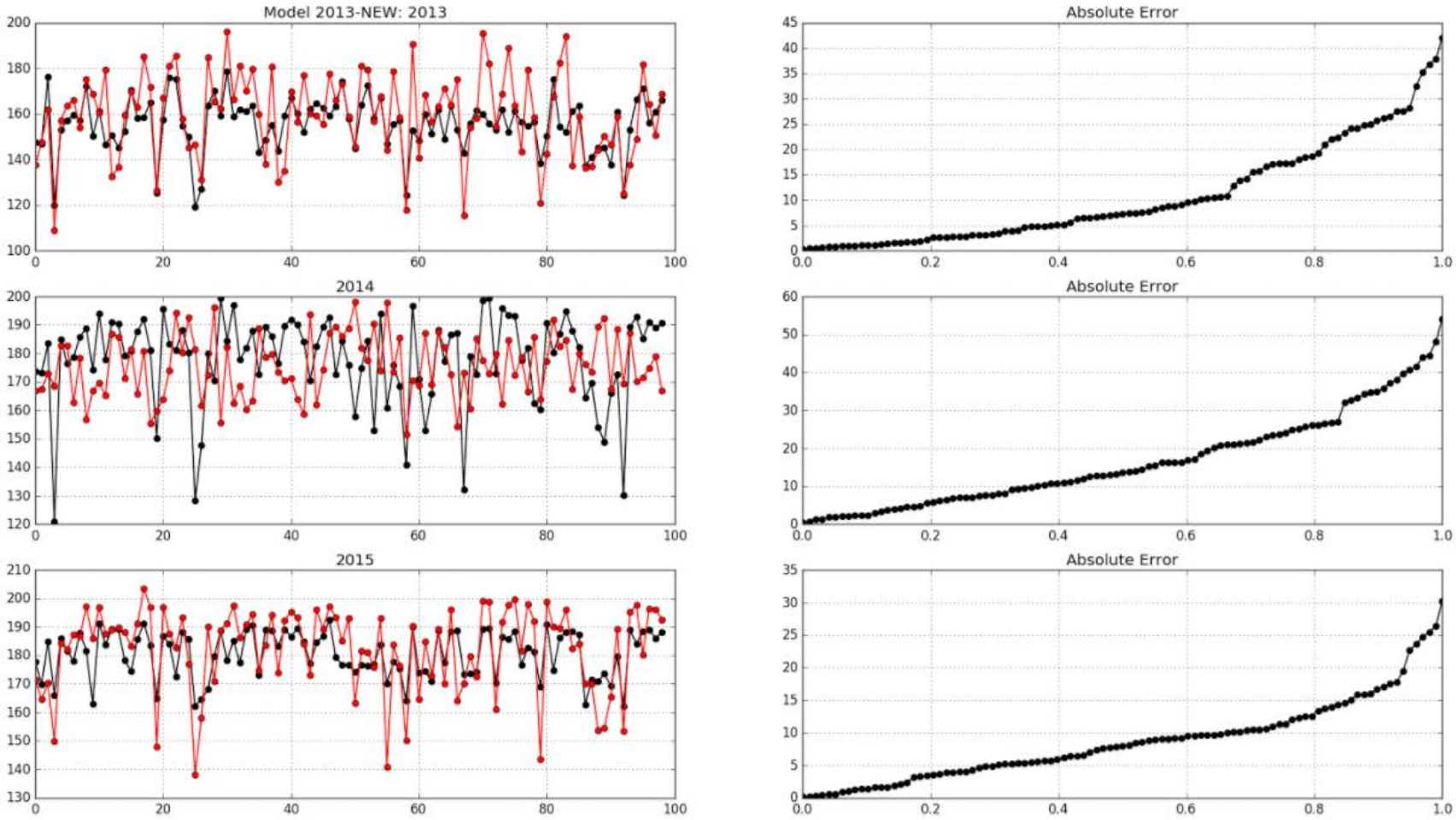


Figure 5.4: Five best and worst counties, 2016



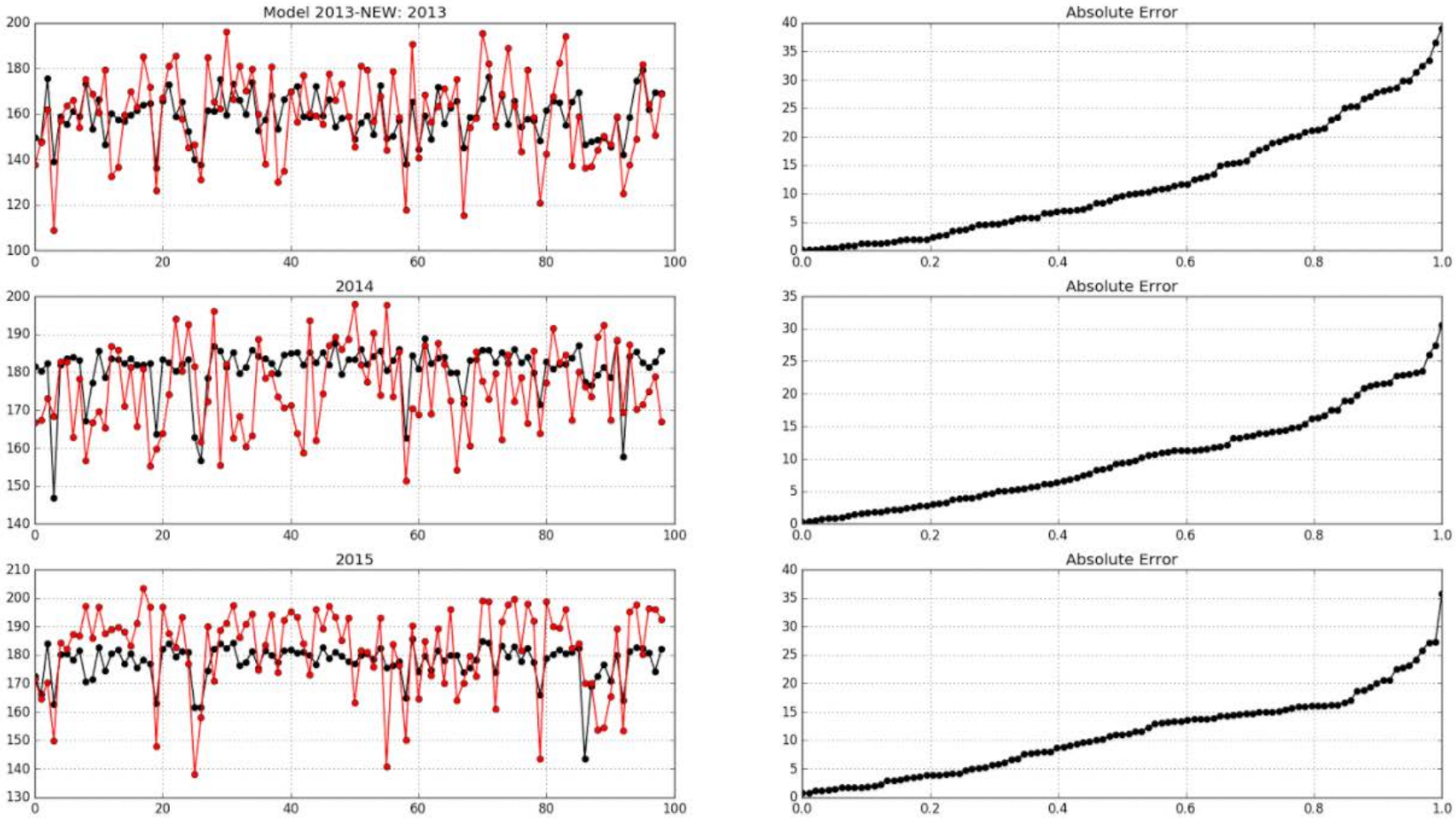
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 5.5: Prediction results for hourly input vectors with initial model



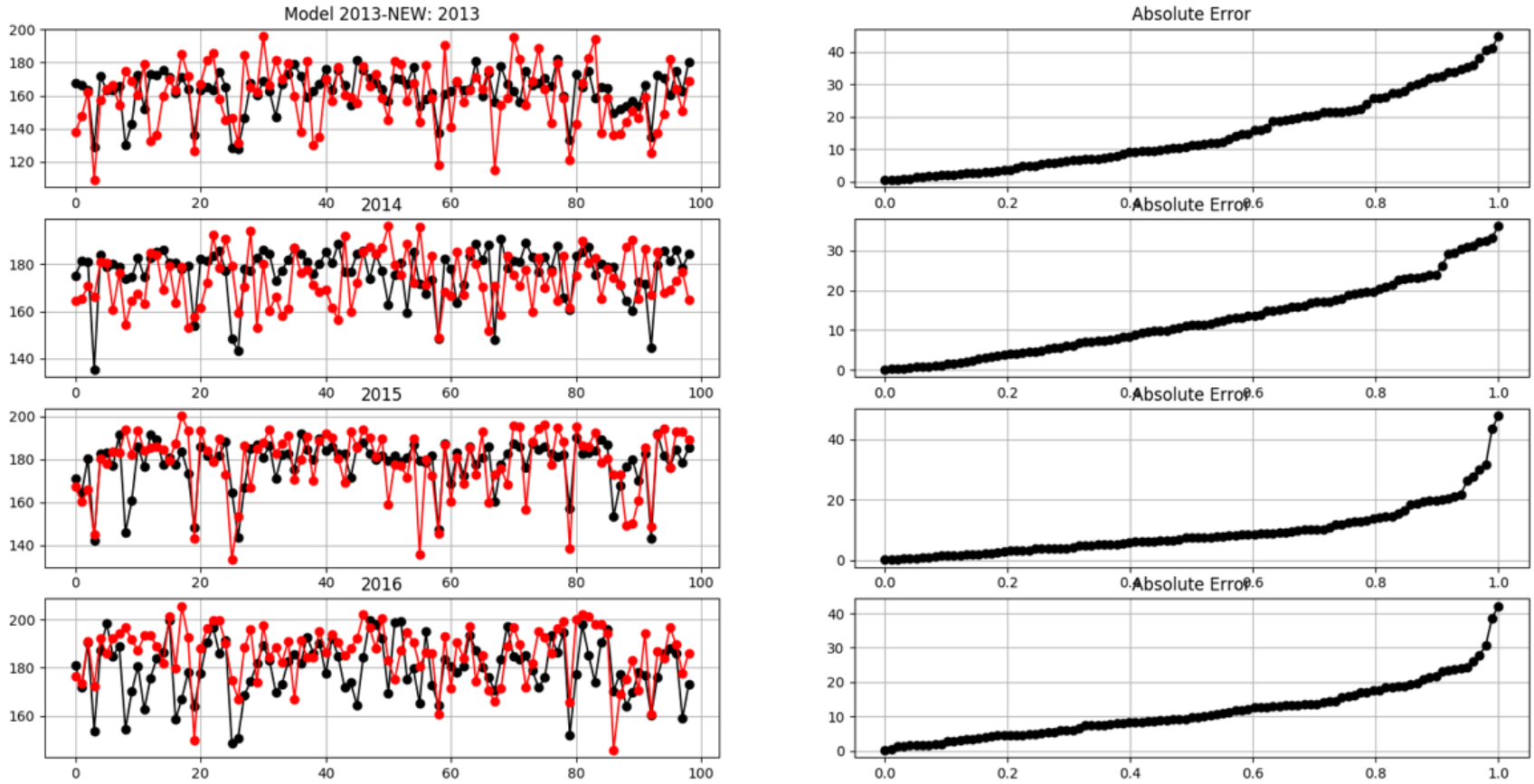
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 5.6: Prediction results for daily input vectors with initial model



Note: Red line is the actual yield, black line is the LSTM prediction

Figure 5.7: Prediction results for two hidden layers LSTM with combination samples



Note: Red line is the actual yield, black line is the LSTM prediction

Figure 5.8: Prediction results for the “best” Iowa LSTM model with constant adjustment

CHAPTER 6. EXTENSIONS

6.1 Prediction of Corn Belt

We show in section 5 that our LSTM performs well in Iowa. Will it also perform well in other states in the Corn Belt? How should we expand the model to other states in the Corn Belt? We have three choices: (*a*) apply the "best" Iowa model to all other states; (*b*) train models separately for each state; and, (*c*) train the "best" model with all data from the Corn Belt. We pick Illinois to test option one since Illinois shares the most corn condition similarities with Iowa. Figure 22 shows that the Iowa model does not predict the Illinois yield well, which indicates that the Iowa model could not fit the other states.

Next, we collect all the data from the 10 states in the Corn Belt to build up a comprehensive model. We still apply the best setting of the model. The comprehensive model consists of two hidden layers, 10 explanatory variables, and 214 days of weather data series. We set two different models under two kinds of adjustment of yield data—percentage and constant. There are 28617 records for training. Since the number of samples is large enough, we did not create combination records here. We test the comprehensive model with Iowa and Illinois data respectively since these two states are the most important corn planting states. Figures 6.3 and 6.4 present the prediction result. Both figures show the same problem with the model containing all candidate variables—our prediction varies less around the average line. We feel the comprehensive model may perform well for aggregated state-level yield prediction as it is more averaged, but it failed to capture the most fluctuation trend; thus, the comprehensive model is not good choice since our first goal is to get accurate county-level yield prediction. Even though we include all data from different states, many unknown variables correlated with each state did not get included in the model. LSTM then only keeps the common information and throws away all other noises, leading to the result that

our big Corn Belt model learns less information than does a separate model for each state. Therefore, the best choice is to train models separately for each state.

Model settings for each state are the same as the Iowa model. In order to guarantee the result is convictive, we search the best model with smallest MSE among over two hundred trained models for each state. Since we are uncertain about which yield adjustment works better, two models under different adjustment were trained for each state. According to the MSE, percentage adjustment worked much better for most of the states except South Dakota. Even though the MSE of the South Dakota model with constant adjustment is still very large, it is significantly smaller than the percentage adjustment. Two county combination samples are still added for each model. The map of agricultural CRDs for each state is attached in the Appendix. Among all the remaining nine states, Missouri is a special case. Missouri is divided into nine CRDs according to the official documentation; however, PDSI for Missouri is recorded by six areas. This information can be found in *Missouri Drought Plan*, Water Resources Report Number 69. Therefore we added combination samples referring to the drought plan instead of the nine CRDs. Another difference from the Iowa case is that areas for planting corn kept changing for many states. Unlike Iowa and Illinois, where the whole state is used for planting corn, only small parts of the state is used for corn production in Kansas, Missouri, and South Dakota. The area is also changing due to climate change—several counties used to plant corn but abandoned it and new counties started planting corn in recent years. Due to this data missing problem, we only include counties that have full yield information for the testing period (2013–2016). Therefore, only 35 of 105 counties in Kansas, 38 of 114 in Missouri, 40 of 83 in Michigan, and 35 of 66 in South Dakota have been predicted. For training data, we include all the available data that the yield is not equal to zero. Detailed prediction results are attached in the Appendix. As +/- 20 error is good performance standard, MSE

should not be much greater than 400. The county-level prediction for Kansas and South Dakota is worse as the MSE is higher than 700. There are only 8822 samples for training in South Dakota, hence the bad result may come from the lack of data. Kansas is a very special state, though we have 14398 historical records, only 35 counties are available for testing. The planting area in Kansas has changed too much, leading to a huge amount of information noises existing in the training data. Therefore, the prediction performance is below expectation.

Table 6.1: Model Information for All Corn Belt State Except Iowa

State	No. of samples	No. of counties	mean squared error	
			percentage	constant
Illinois	21150	102/102	354.18	414.64
Indiana	17094	70/92	329.20	461.21
Kansas	14398	35/105	700.81	1980.37
Michigan	10500	40/83	295.46	755.30
Minnesota	15232	68/87	260.29	427.30
Missouri	32818	38/114	434.64	987.56
Nebraska	19589	71/93	463.06	1156.16
Ohio	14398	71/88	316.53	713.38
South Dakota	8822	35/66	888.70	723.29

We also aggregate the prediction to state level for each state to compare with the USDA prediction in November. In order to make the comparison straight and clear, we only present the mean absolute error (MAE) and mean absolute percentage error (MAPE) of the prediction for 2013–2016. MAPE is calculated by $M = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$, where A_t is the actual yield, F_t is the forecast value, $n = 4$ here. According to the results listed in Table 6.2, all the MAPEs are less than 7% and most of them are even less than 3%, which is really great prediction. Half of the predictions beat the USDA prediction, while the MAPE of Kansas, Missouri, and South Dakota was about

5%–7%, which are the states that only include less than 40 counties for testing. Even though the result is worse than USDA for these states, the prediction accuracy is still better than other research using machine learning methods for corn yield prediction. When we aggregate the results to the whole Corn Belt in Table 6.3, the result are good. The prediction error is even less than 1 bushel/acre and also beats the USDA prediction.

Table 6.2: State-Level Prediction Comparison

State	LSTM_NOV		NASS_NOV	
	MAE (bu/ac)	MAPE (%)	MAE (bu/ac)	MAPE (%)
Iowa	4.31	2.29	4.37	2.41
Illinois	2.85	1.56	3.50	1.92
Indiana	2.96	1.74	3.75	2.27
Kansas	10.14	6.95	4.25	3.02
Michigan	6.09	3.79	3.25	2.03
Minnesota	4.24	2.38	4.50	2.75
Missouri	9.92	5.88	3.25	2.06
Nebraska	3.41	1.95	2.50	1.39
Ohio	3.42	2.01	4.00	2.56
South Dakota	7.76	5.29	6.75	4.46

Table 6.3: Corn-Belt-Level Prediction Comparison

Year	Yield	LSTM_NOV	NASS_NOV
2013	161.78	162.9	164.29
2014	175.60	175.44	178.49
2015	173.1	173.4	173.04
2016	181.09	179.34	181.54
MAE (bushels/acre)		0.8325	1.4475
MAPE (%)		0.48	0.87

State-Level Prediction Results in detail:

Table 6.4: State-Level Prediction for Illinois in November

Illinois (Percentage): 102/102

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	178	175.1381	180	80%	96%
2014	200	196.3075	200	72.5%	91%
2015	175	170.1886	168	68%	88%
2016	197	196.9654	202	63%	82%

Table 6.5: State-Level Prediction for Indiana in November

Indiana (Percentage): 70/92

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	177	172.3266	174	81%	97%
2014	188	186.8123	186	84%	97%
2015	150	152.5922	156	60%	81%
2016	173	176.3785	177	70%	86%

Table 6.6: State-Level Prediction for Kansas in November

Kansas (Percentage): 35/105

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	126	124.0206	130	34%	60%
2014	149	129.7874	157	40%	63%
2015	148	136.5575	148	77%	91%
2016	142	134.0558	147	74%	86%

Table 6.7: State-Level Prediction for Michigan in November

Michigan (Percentage): 40/83

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	155	153.8289	156	72.5%	87.5%
2014	161	145.6939	166	65%	92.5%
2015	162	155.11	167	85%	95%
2016	157	155.9927	155	67.5%	90%

Table 6.8: State-Level Prediction for Minnesota in November

Minnesota (Percentage): 68/87

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	159	158.8651	164	76%	94%
2014	156	160.8041	165	72%	91%
2015	188	179.035	187	85%	93%
2016	193	196.067	190	84%	94%

Table 6.9: State-Level Prediction for Missouri in November

Missouri (Percentage): 38/114

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	136	138.7733	133	66%	84%
2014	186	165.8375	181	63%	82%
2015	142	145.904	145	74%	89%
2016	163	150.1406	165	58%	82%

Table 6.10: State-Level Prediction for Nebraska in November

Nebraska (Percentage): 71/93

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	169	163.2297	169	66%	82%
2014	179	178.0822	181	70%	85%
2015	185	183.7879	187	62%	86%
2016	178	183.7312	184	68%	86%

Table 6.11: State-Level Prediction for Ohio in November

Ohio (Percentage): 71/88

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	174	170.3723	174	75%	96%
2014	176	168.7132	177	79%	97%
2015	153	151.3659	163	62%	80%
2016	159	157.844	164	76%	92%

Table 6.12: State-Level Prediction for South Dakota in November

South Dakota (Constant): 35/66

Year	Yield	LSTM_NOV	NASS_NOV	+/- 20	+/- 30
2013	137	150.7581	145	69%	86%
2014	148	154.2152	151	54%	71%
2015	159	162.5958	162	54%	77%
2016	161	168.4958	148	34%	60%

6.2 Early Prediction

NASS reports their yield prediction annually in August, September, October and November. November is the final prediction when all the information for corn growth is known since corn is usually harvested in October. Hence, any prediction that is prior to the harvest time is defined as the early prediction. Accurate early prediction is important—it will not only influence corn storage decisions, but also the corn futures market. USDA reports their estimate of total corn harvest acreage June 30 every year. Starting in July, combined with early corn yield prediction, corn production estimates are made until harvest in October. Many factors influence futures price, but estimated corn production is the most influential factor from July to October.

6.2.1 Early Prediction in Iowa

We trained three early prediction models for Iowa with data from August, September, and October, respectively, to compare with results from NASS (i.e. y=122, 153, and 183 for the 3D tensor cube in section 3.3). Figure 6.1 summarizes the prediction results of NASS and our LSTM model. All our models are trained with almost 700 hyperparameter sets, which means we should reach an optimal model. The grey line is the actual yield, the red line is our prediction, and the blue line is the USDA prediction. Most of our early prediction results are better than the USDA prediction, which indicates the strong power of the LSTM method in making early prediction with limited data. We also notice a trend that predictions in November are usually worse than in October.

This is because we could not know the accurate harvest date for each year and data after the harvest date is included in the training data, thus resultant redundant and noisy information has been learned by the model and influences the final prediction.

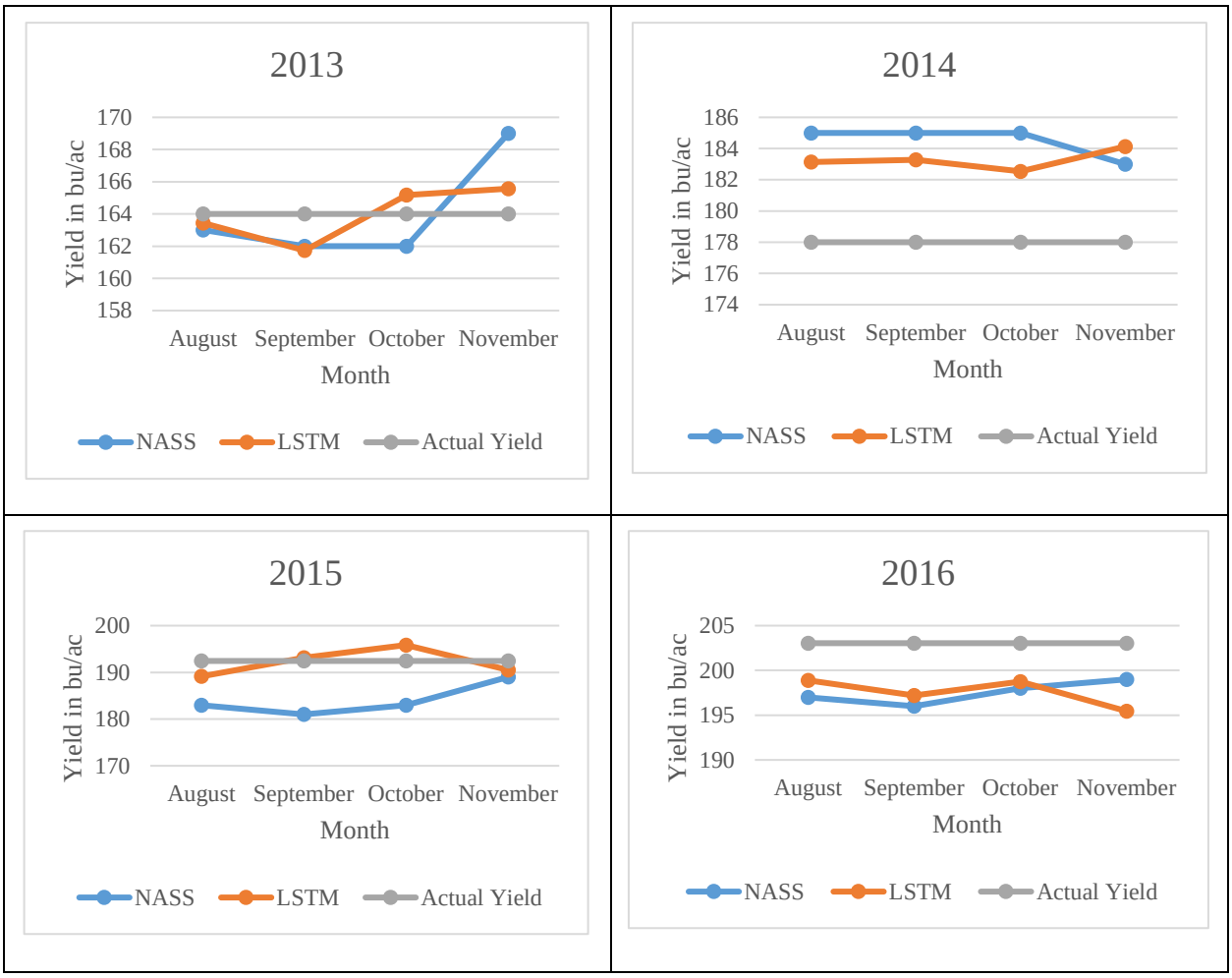


Figure 6.1: Yield prediction comparison between NASS and LSTM

6.2.2 Early Prediction in Corn Belt

For the nine Corn Belt states other than Iowa, we only made early predictions in August. Still, we compared our results at the state (Table 6.13) and Corn Belt level (Table 6.14), respectively, with the USDA prediction. Half of the MAPEs are less than 5% and all MAPEs are less than 9%. Six of the ten states beat the USDA results. This indicates that LSTM still works

well in early prediction around the whole Corn Belt. The results at the Corn Belt level are only 1.339 bu/acre away from the actual yield, which means that we are able to predict the yield as early as August with only 0.78% difference between our prediction and USDA statistics.

LSTM is also available for daily prediction. It would be too much work to train separate models for each day to make daily predictions. Therefore, we can use the best model from sections 5 and 6.1 for each state, input known explanatory variables data, and fill unknown data with expected values. Weather and soil humidity data after the date on which the daily prediction is made is the unknown data. The expected input values can be either past 10-year average or from the professional prediction in weather channels if it is available (IBM's Weather Underground, Weather Channel and Accuweather are good resources, weather prediction is usually available for 30 days in the future and accurate for only one week).

Table 6.13: Comparison of State-Level Early Prediction in August

State	LSTM_AUG		NASS_AUG	
	MAE	MAPE (%)	MAE	MAPE (%)
Iowa	3.29	1.75%	6.12	3.23%
Illinois	4.55	2.43%	7.75	4.14%
Indiana	9.22	5.59%	10.50	6.11%
Kansas	9.87	6.78%	5.25	3.86%
Michigan	6.00	3.76%	2.75	1.74%
Minnesota	6.75	4.03%	8.00	4.72%
Missouri	9.93	5.82%	10.75	6.47%
Nebraska	5.38	3.06%	6.25	3.56%
Ohio	9.44	5.61%	5.50	3.51%
South Dakota	13.12	8.91%	6.25	4.03%

Table 6.14: Comparison of Corn-Belt-Level Early Prediction in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	161.78	160.3323	157.3596
2014	175.60	173.0495	172.4956
2015	173.1	172.6832	172.8598
2016	181.09	180.1483	180.7805
MAE (bushels/acre)		1.339	2.019
MAPE (%)		0.78	1.20

State Level Early Prediction Results in August in detail:

Table 6.15: State-Level Prediction for Iowa in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	164	163.44	163
2014	178	183.13	185
2015	192.44	189.14	183
2016	203.04	198.89	196

Table 6.16: State-Level Prediction for Illinois in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	178	168.98	165
2014	200	193.42	188
2015	175	173.98	172
2016	197	195.43	200

Table 6.17: State-Level Prediction for Indiana in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	177	171.85	166
2014	188	181.72	179
2015	150	165.5	158
2016	173	182.96	187

Table 6.18: State-Level Prediction for Kansas in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	126	123	116
2014	149	133.17	145
2015	148	132.9	152
2016	142	136.45	145

Table 6.19: State-Level Prediction for Michigan in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	155	150	158
2014	161	148.61	161
2015	162	157.24	165
2016	157	158.87	152

Table 6.20: State-Level Prediction for Minnesota in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	159	169	166
2014	156	164	168
2015	188	185.4	184
2016	193	186.57	184

Table 6.21: State-Level Prediction for Missouri in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	136	132.47	130
2014	186	159.69	160
2015	142	147.18	150
2016	163	158.31	166

Table 6.22: State-Level Prediction for Nebraska in August

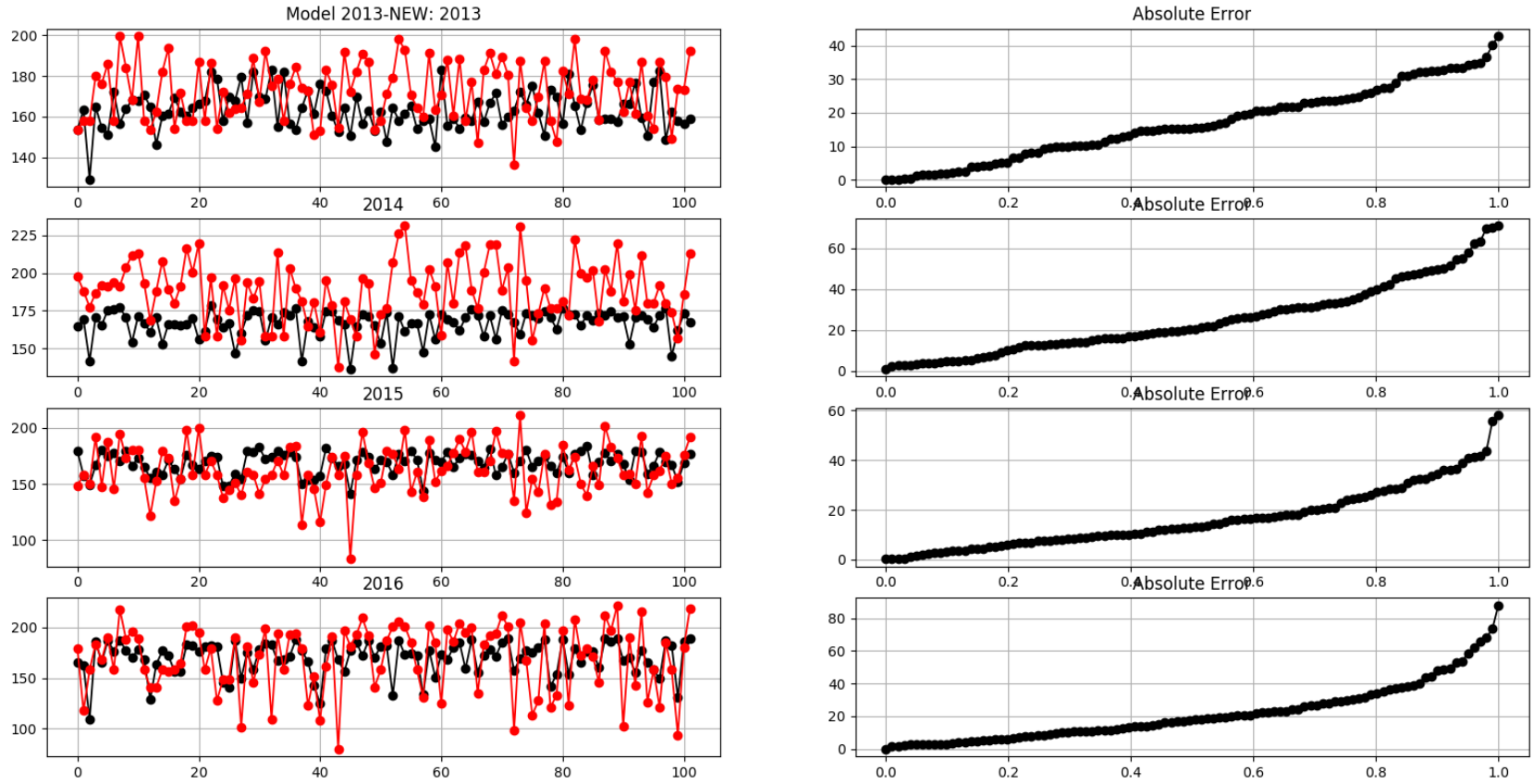
Year	Yield	LSTM_AUG	NASS_AUG
2013	169	159.36	161
2014	179	175.6	173
2015	185	178.8	187
2016	178	175.73	187

Table 6.23: State-Level Prediction for Ohio in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	174	166.61	172
2014	176	158.38	177
2015	153	160.05	168
2016	159	164.71	163

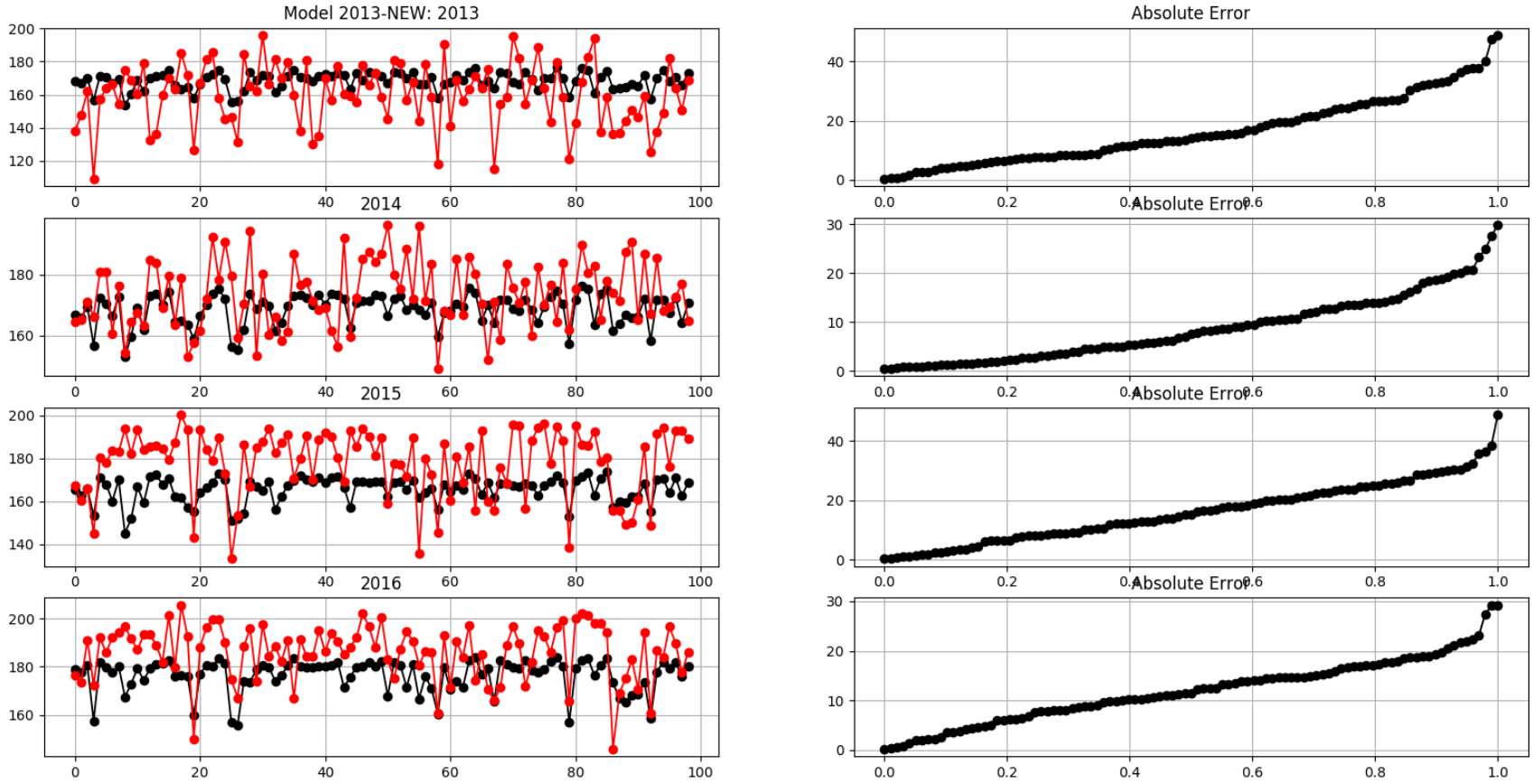
Table 6.24: State-Level Prediction for South Dakota in August

Year	Yield	LSTM_AUG	NASS_AUG
2013	137	156.5	138
2014	148	163.34	139
2015	159	167.47	160
2016	161	170.18	147



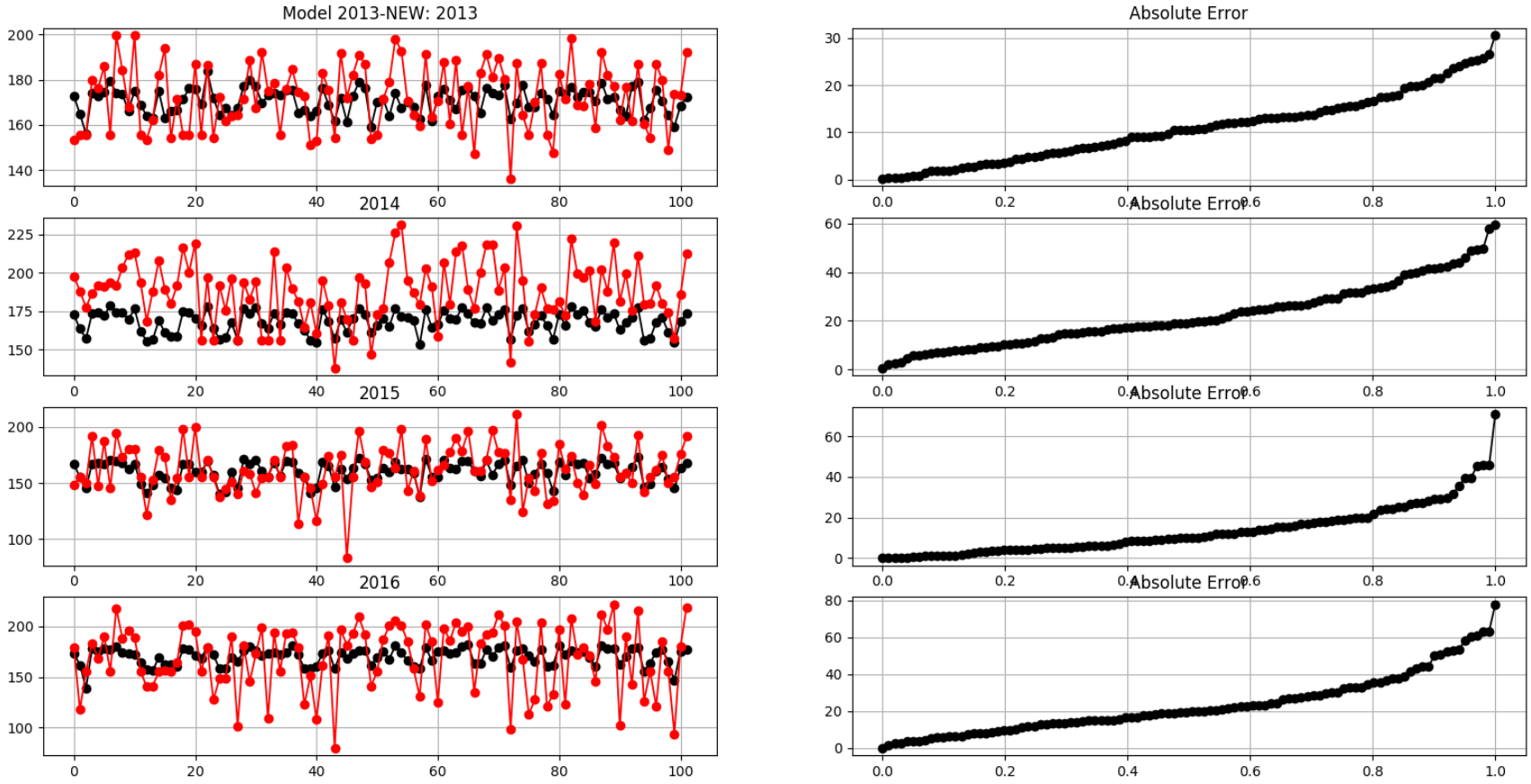
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.2: Iowa model performance in Illinois



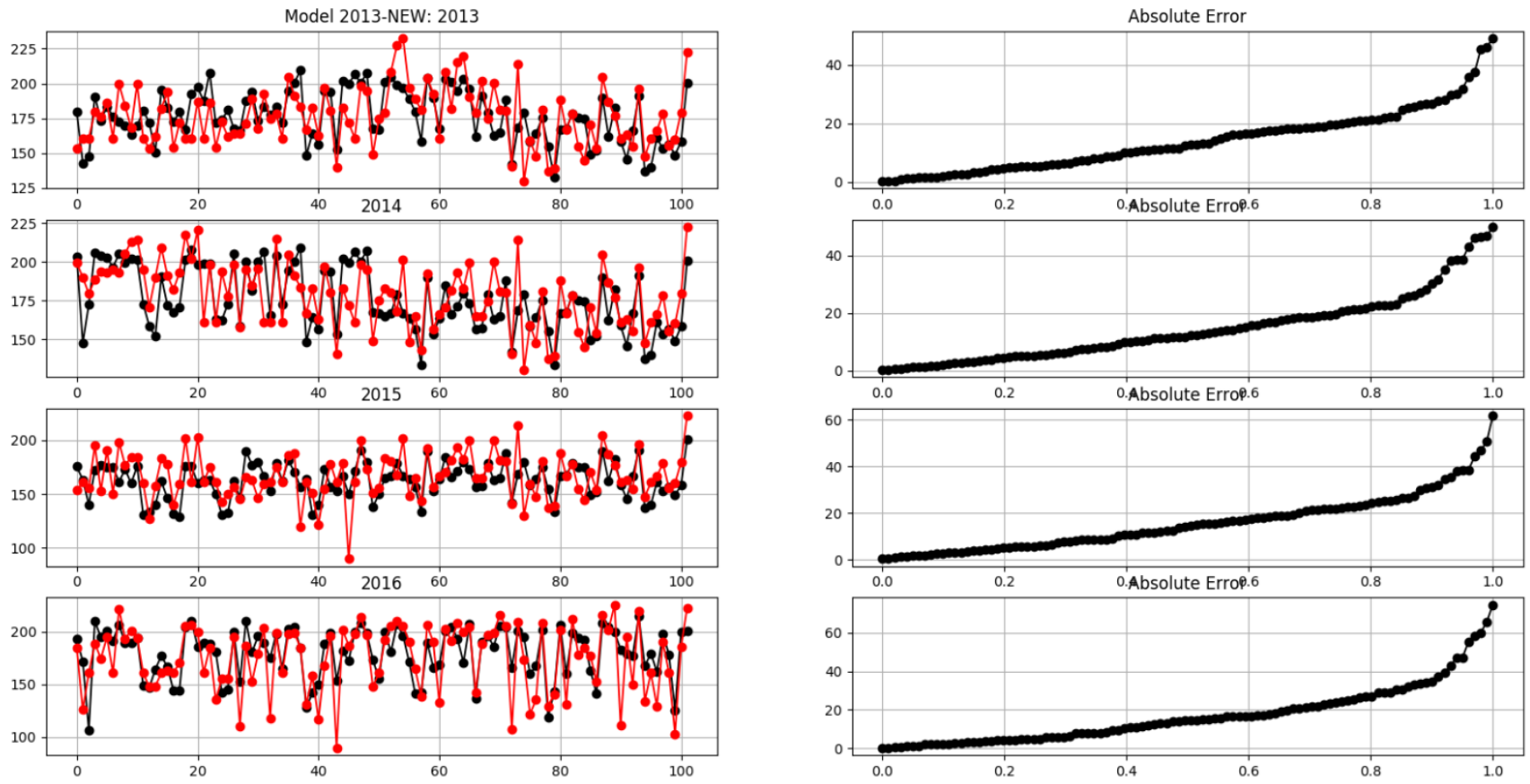
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.3: Comprehensive model performance in Iowa



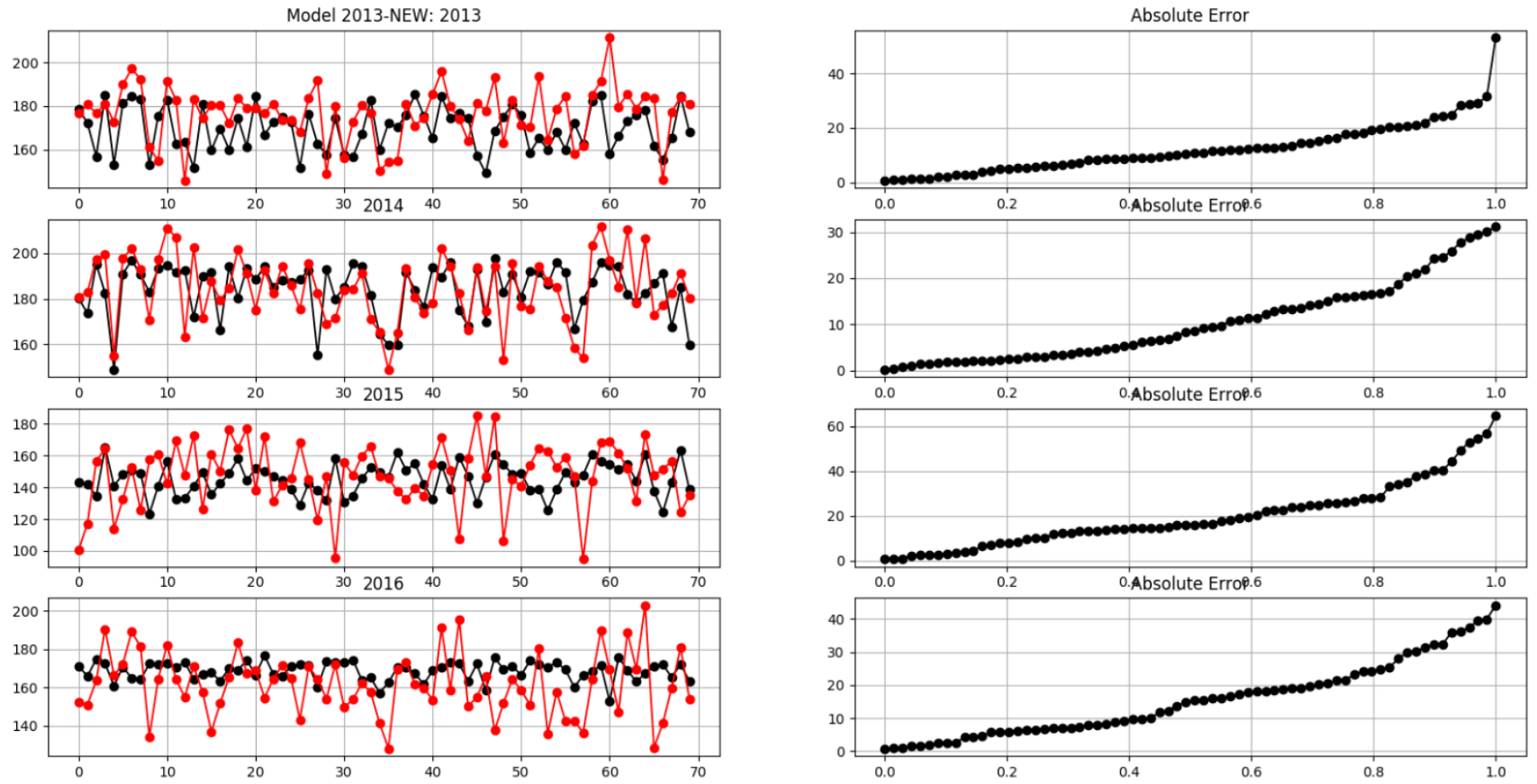
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.4: Comprehensive model performance in Illinois



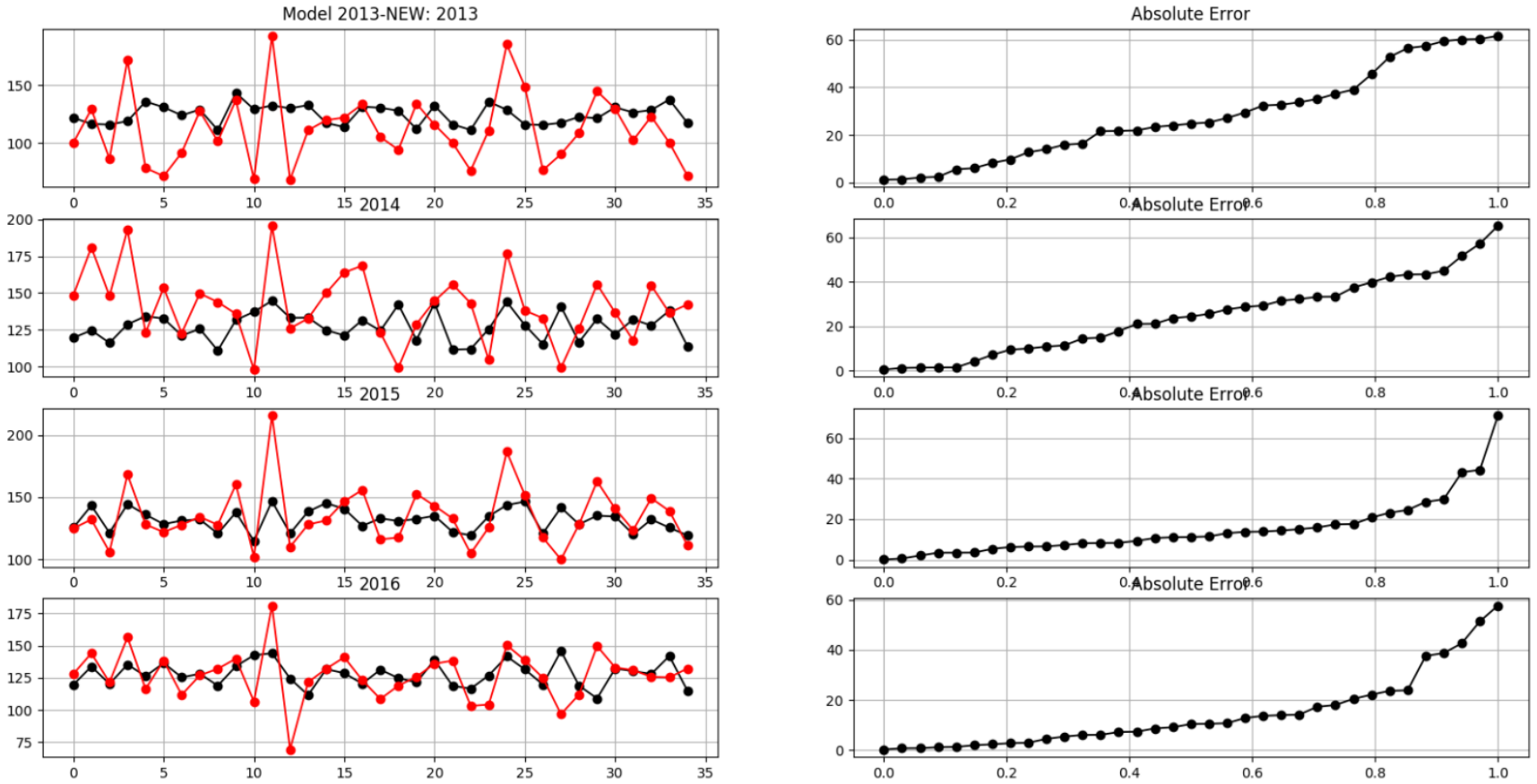
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.5: Prediction results for the “best” Illinois LSTM model with percentage adjustment



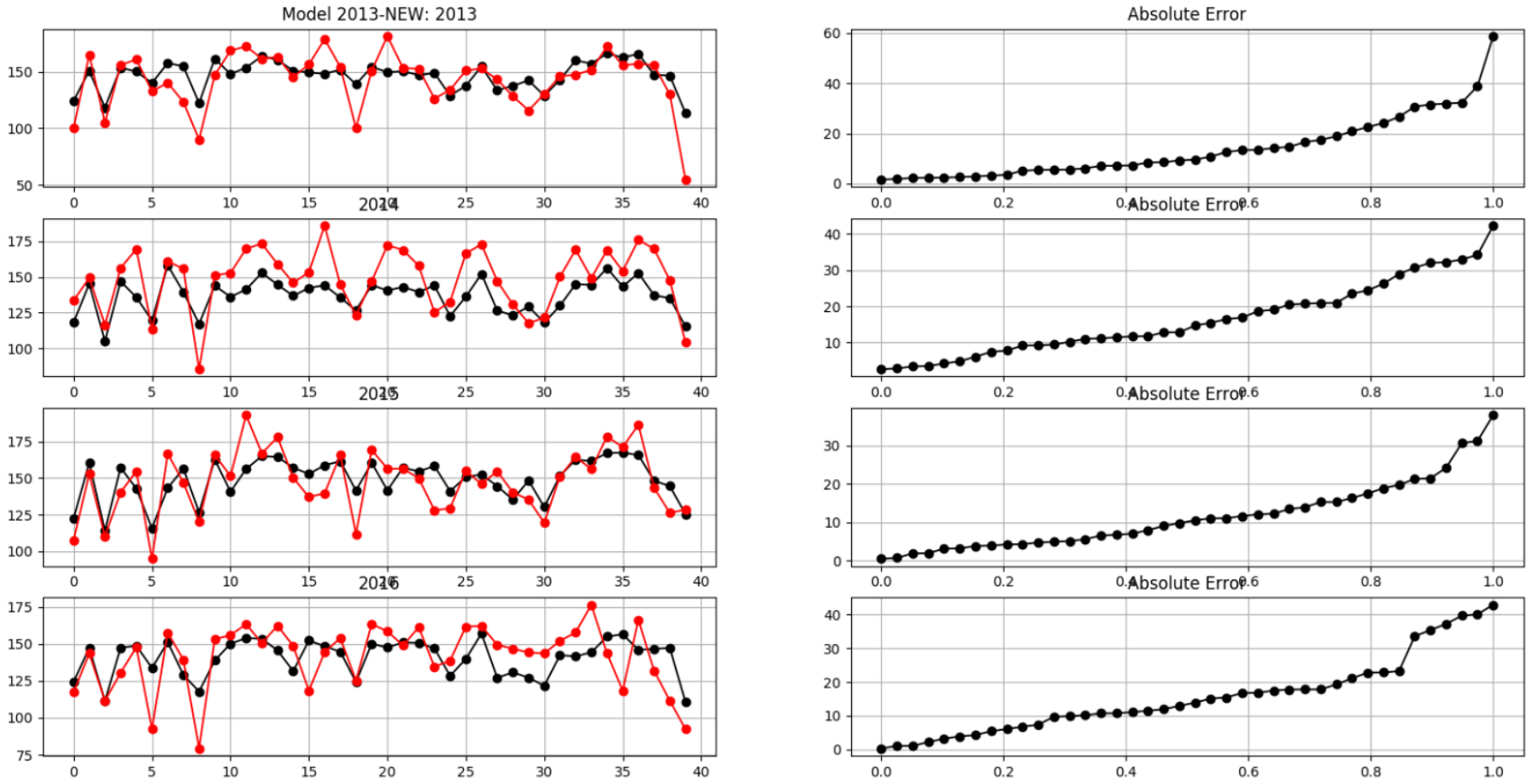
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.6: Prediction results for the “best” Indiana LSTM model with percentage adjustment



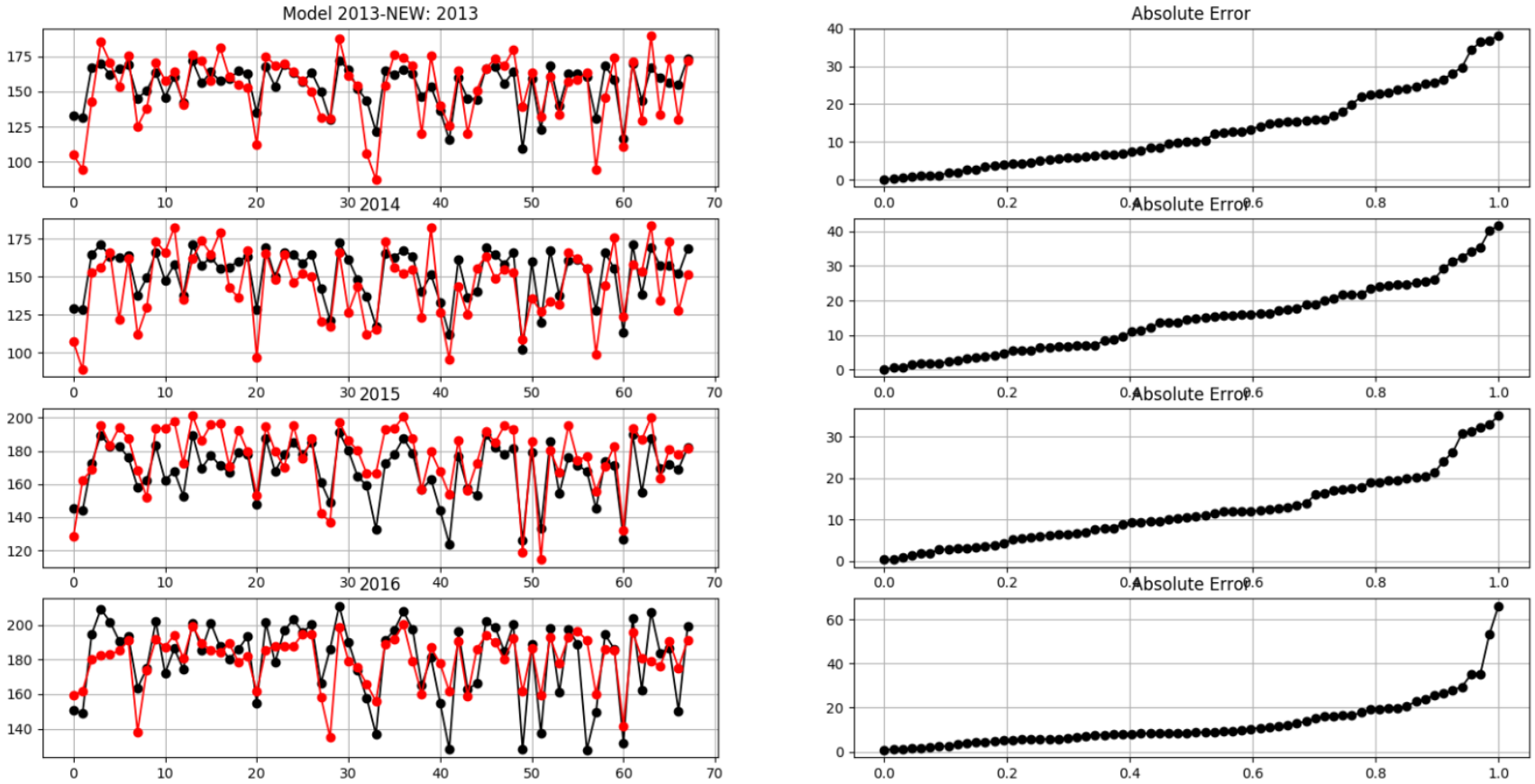
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.7: Prediction results for the “best” Kansas LSTM model with percentage adjustment



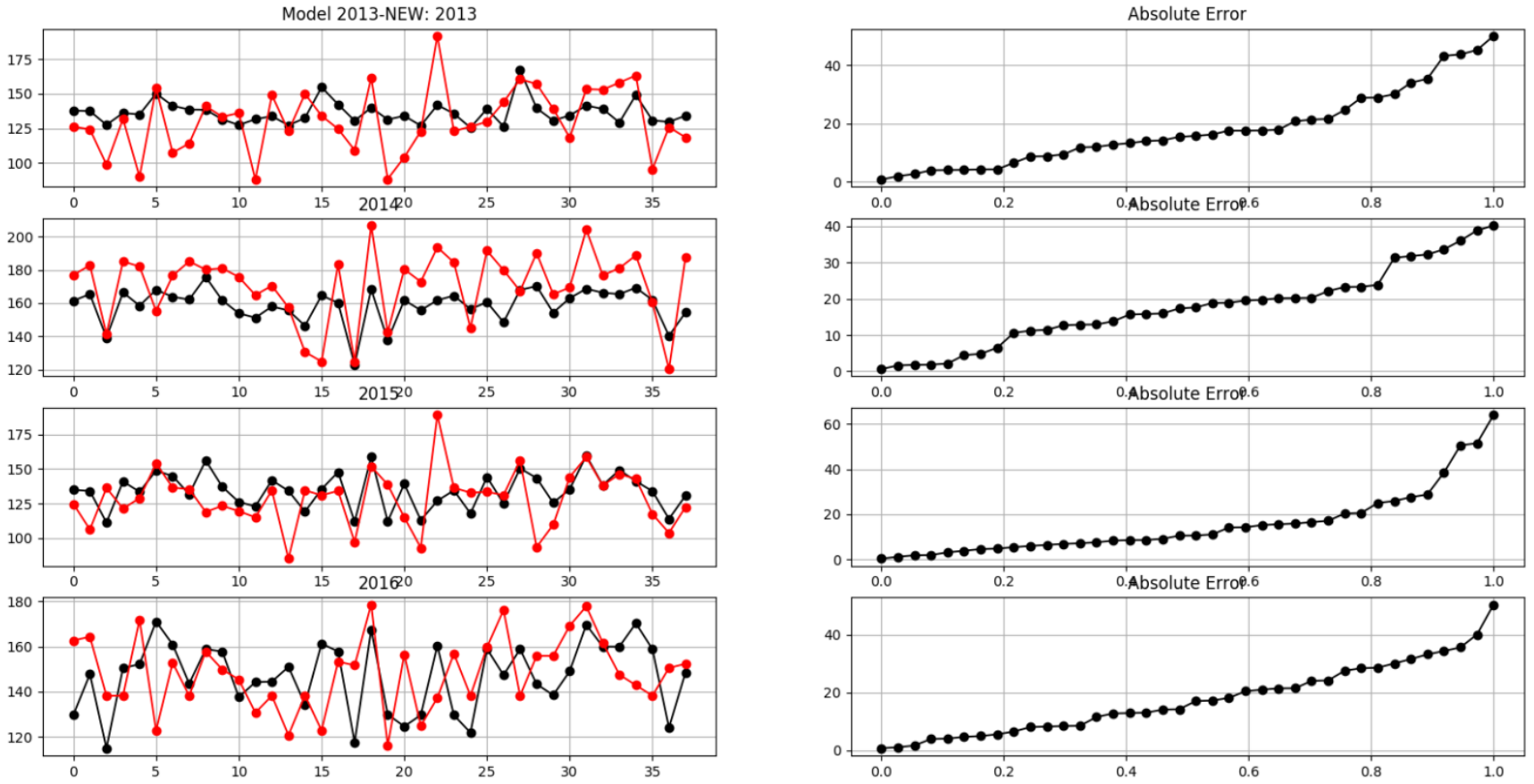
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.8: Prediction results for the “best” Michigan LSTM model with percentage adjustment



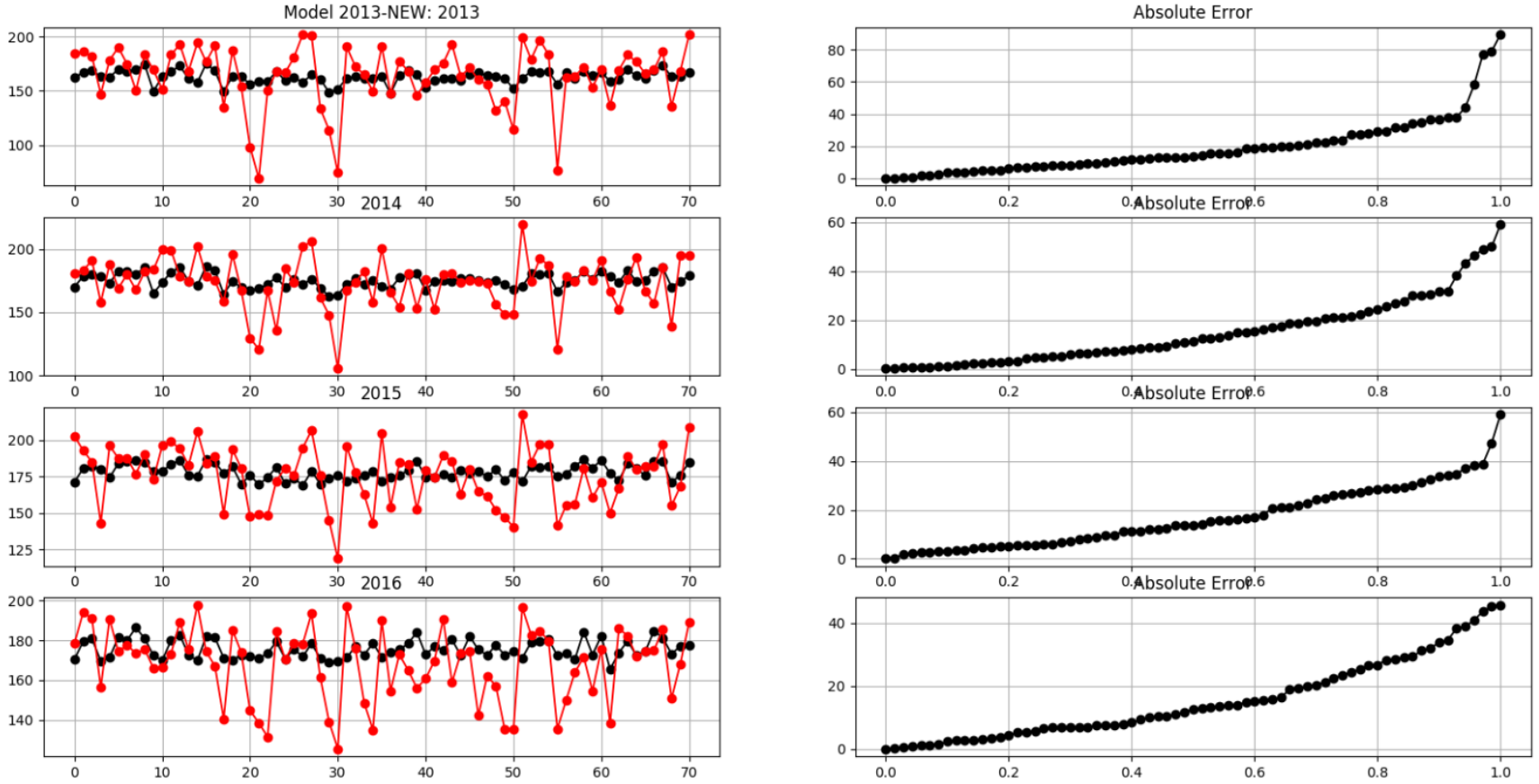
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.9: Prediction results for the “best” Minnesota LSTM model with percentage adjustment



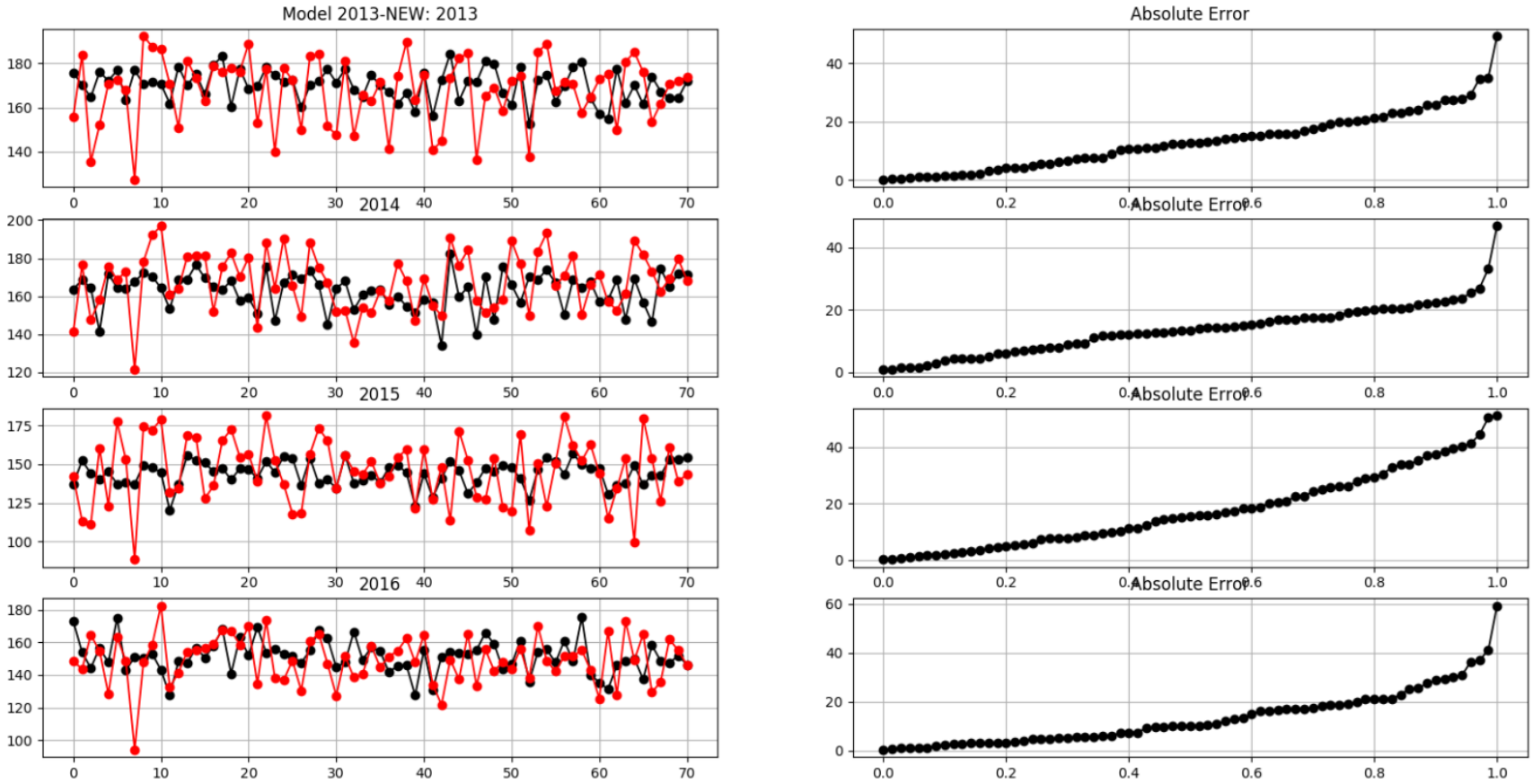
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.10: Prediction results for the “best” Missouri LSTM model with percentage adjustment



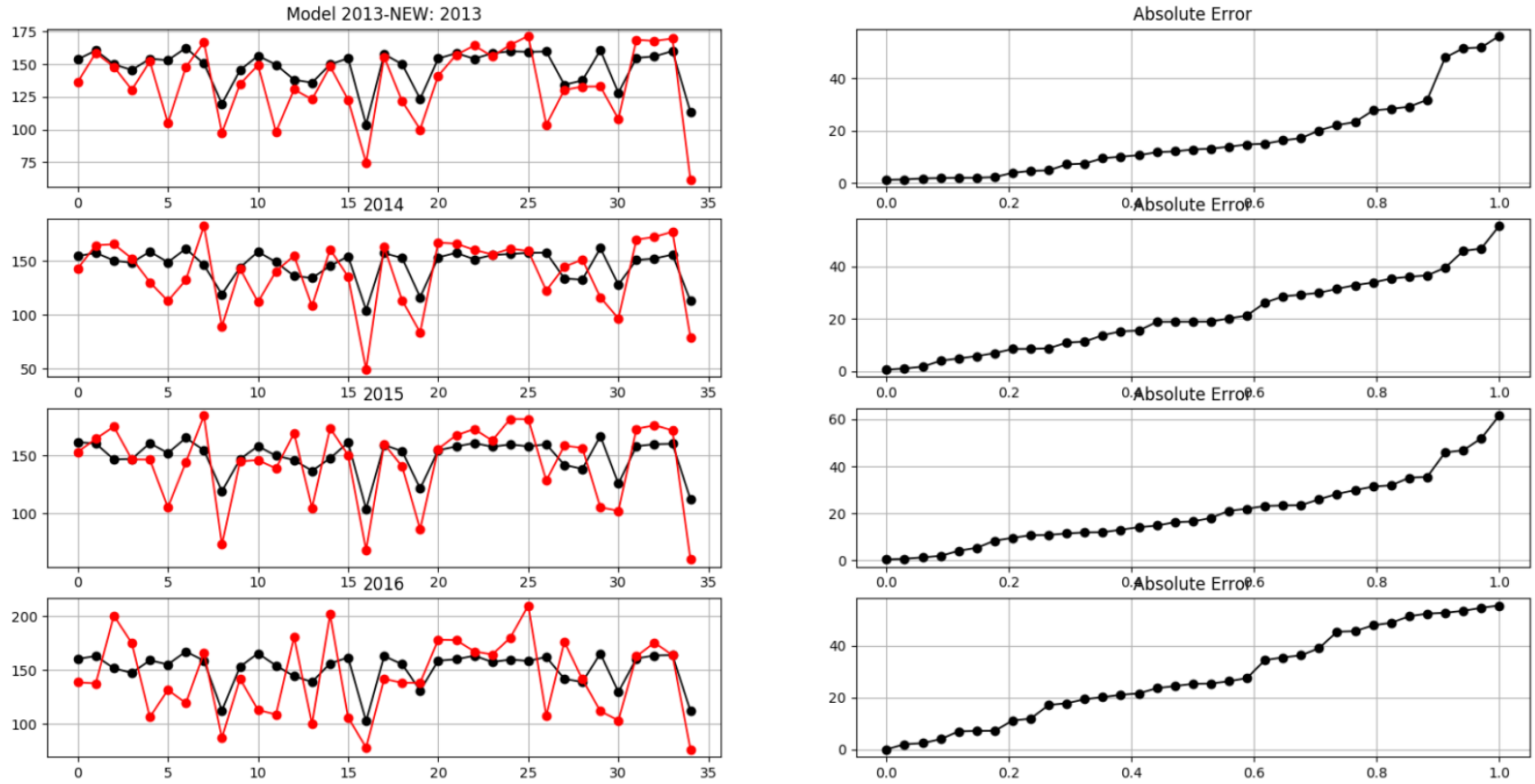
Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.11: Prediction results for the “best” Nebraska LSTM model with percentage adjustment



Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.12: Prediction results for the “best” Ohio LSTM model with percentage adjustment



Note: Red line is the actual yield, black line is the LSTM prediction

Figure 6.13: Prediction results for the “best” South Dakota LSTM model with constant adjustment

CHAPTER 7. CONCLUSION

This paper describes the prediction of county-level corn yields in the Corn Belt area using the deep learning method LSTM. We develop a novel way to apply LSTM in crop yield prediction and are convinced that LSTM can be a powerful option for crop yield modelling. Our prediction is only 0.83 bushel/acre lower than actual corn yields in the Corn Belt. Eighty percent of our LSTM model predictions fall in the +/-20 bushel/acre accuracy region. Results show that our LSTM model can provide good early prediction and accurate Corn Belt county-level corn yield prediction without farm management and genetic data. Our LSTM models for county-level corn yield prediction in the Corn Belt area is a good supplement and improvement to the USDA prediction. This supplement and improvement will contribute to eliminating the information asymmetry problem that arises from the success of private companies in crop yield prediction.

REFERENCES

Anandkumar, A., Hsu, D., Javanmard, A. and Kakade, S.M., 2013. "Learning linear Bayesian networks with latent variables." *Proceedings of 30th International Conference on Machine learning* 28: 1-9.

AgriData. NRCS Soil Map Data. Available at https://www.agridatainc.com/Home/Products/Surety/NRCS%20Online%20Soil%20Maps

Almaraz, J.J., Mabood, F. and Zhou, X.M., 2008. "Climate change, weather variability and corn yield at a higher latitude locale, Southwestern Quebec." *Climatic Change* 88: 187-197.

Akintayo, A., and Sarkar, S. 2015. "A Symbolic Dynamic Filtering Approach to Unsupervised Hierarchical Feature Extraction from Time-Series Data." Published paper, American Control Conference 15, Chicago, IL., July 1–3, 2015.

Basso, B., Ritchie, J.T., Pierce, F.J., Braga, R.P., Jones, J.W., 2001. "Spatial validation of crop models for precision agriculture." *Agricultural Systems* 68: 97–112.

Below, F. 2008. "The Seven Wonders of the Corn Yield World." *Institute of Ag Professionals.*

Charles, L.H. and Benjamin Cook, J., 2014. "Predicting Yield in the Corn Belt." *Stat 225, Harvard University*.

Chawla, V., Naik, H.S., Akintayo, A., Hayes, D., Schnable, P., Ganapathysubramanian, B., and Sarkar, S. 2016. "A Bayesian Network Approach to County-Level Corn Yield Prediction using Historical Data and Expert Knowledge." KDD 16 Workshop: Data Science for Food, Energy and Water, San Francisco (ISBN: 978-1-4503-2138-9) DOI: 10.1145/1235.

Che, Z., Sanjay, P., Cho, K., David Sontag and Yan Liu, 2017. "Recurrent Neural Networks for Multivariate Time Series with Missing Values." *ICLR 2017*

Climate Corp. Available at https://www.climate.com/

Corn Belt. Encyclopedia Britannica Online.

Deschênes, O., and Greenstone, M., 2007. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *The American Economic Review* 97: 354–385.

Druzdzel, M.J., 1999. "Smile: Structural modeling, inference, and learning engine and genie: a development environment for graphical decision-theoretic models." *AAAI '99/IAAI '99 Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence* (ISBN: 0-262-51106-1): 902-903.

Foote, R.J., and Bean, L.H., 1951. “Are Yearly Variations in Crop Yields Really Random?” *Agricultural Economics Research* Volume 3: 23–30.

Farnia, F., Kazerouni, A. and Babveyh, A. “Information-based Feature Selection.” *CS229*, Standford University

George, A.A., 1970. “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics* 84: 488–500.

Gerlt, S., W. Thompson and D.J. Miller. 2014. “Exploiting the relationship between farm-level yields and county-level yields for applied analysis.” *Journal of Agricultural and Resource Economics*, 39(2):253-270.

Gibson, L.R. Growing degree-day calculation. http://agron-www.agron.iastate.edu/Courses/agron212/Calculations/GDD.gtm.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. “Deep Learning.” *The MIT Press*.

Hinton, G., RMSprop. CSC321. *Lecture 6e Coursera Class.*

Historical Palmer Drought Indices, NOAA. Available at https://www.ncdc.noaa.gov/temp-and-precip/drought/historical-palmers/

Hochreiter, S., and Schmidhuber, J. 1997. “Long Short-Term Memory.” *Neural Computation* 9(8): 1735–1780.

Jiang, Z., Liu, C., Akintayo, A., Henze, G.P. and Sarkar, S., 2017. “Energy Prediction using Spatiotemporal Pattern Networks.” *Applied Energy* Feb 2017

Kaul, M., Hill, R.L., and Walthall, C. 2005. “Artificial Neural Networks for Corn and Soybean Yield Prediction.” *Agricultural Systems* 85: 1–18.

Kaylen, M.S., and Koroma, S.S. 1991. “Trend, Weather Variables, and the Distribution of U.S. Corn Yields.” *Review of Agricultural Economics*, *Oxford University Press on behalf of Agricultural & Applied Economics Association* 13: 249–258.

Kefaya, Q., 2014. “Modified Naive Bayes Based Prediction Modeling for Crop Yield Prediction.” *International Journal of Biological, Biomolecular, Agricultural, Food and Biotechnological Engineering*, 8(1):36-39.

Keras Documentation. Available at http://keras.io/

Kerber, R., 1992. “Chimerge: Discretization of numeric attributes.” *Proceedings of the tenth national conference on Artifical intelligence* 123-128.

Kim, N.L. and Lee, Y-W. 2016. “Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State.” *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 34(4): 383–390.

Léon, B., 1998. “Online Algorithms and Stochastic Approximations.” *Online Learning and Neural Networks*. *Cambridge University Press*.

Li, L., 2014. “The Impact of Weather and Weather Extremes on Corn Yields.” *Ph.D. thesis, Economics, Iowa State University.*

Liu, C., Ghosal, S., Jiang, Z. and Sarkar, S., 2017. “An unsupervised spatiotemporal graphical modeling approach to anomaly detection in distributed CPS.” *Grant No. CNS-1464279.*

Michael, H., and Peter, S., 2003. “Recurrent Neural Networks for Time Series Classification.” *Neurocomputing* 50(C): 223-235.

Missouri Department of Natural Resources, 2002. “Missouri Drought Plan.” *Water Resource Report Number 69.*

Mozer, M. C., 1995. “A Focused Backpropagation Algorithm for Temporal Pattern Recognition. Backpropagation: Theory, architectures, and applications.” *ResearchGate. Hillsdale, NJ: Lawrence Erlbaum Associates* 137–169.

NDAWN Center. Available at https://ndawn.ndsu.nodak.edu/help-corn-growing-degree-days.html/

NRC, 2010. “Advancing the Science of Climate Change.” *National Research Council*. The National Academies Press, Washington, DC, USA.

Newlands, N.K., and Townley-Smith, L. 2010. “Predicting Energy Crop Yield using Bayesian Networks.” *Proceedings of the Fifth IASTED International Conference, Computational Intelligence* 107–112.

Peng, H., Long, F., and Ding, C. 2005. “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(8).

Schlenker, W., and Roberts, M.J. 2009. “Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change.” *Proceedings of the National Academy of Sciences* 106: 15594–15598.

Tannura, M.A., Irwin, S.H. and Good, D.L., 2008. “Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt.” *Marketing and Outlook Research Report 2008–01*. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, February.

Tellus Labs, Kernel. Available at https://telluslabs.com/pages/kernel

Theano. Available at https://github.com/Theano/Theano

Thomas Reuters, Lanworth Crop Production Forecasts. Available at https://www.thomsonreuters.com/en.html

U.S. Geological Survey, Department of the Interior. Available at http://catalog.data.gov/dataset/national-land-cover-database-nlcd-land-cover-collection

USDA., NASS Quickstats. Available at http://quickstats.nass.usda.gov/

USDA., SSURGO. Available at https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053631

Weather Underground. Available at https://www.wunderground.com/history/

Yu, T., 2011. "Three essays on weather and crop yield." *Ph.D. thesis, Economics, Iowa State University.*

APPENDIX. ADDITIONAL MATERIAL

Tables A.1–A.9 are results for a Bayesian Network. State data means the BN is trained with state data only, full data means that BN is trained with all data gathered from Iowa, Indiana, Illinois, Nebraska, and Minnesota.

Table A.1: Confusion Matrix with State Data for Iowa

	Predicted yield (Bushels/acreer)					
Yield(Bush	Below 54	54-77	77-106	106-147	147-174	174 Above
Below 54	0	20	0	0	0	0
54-77	0	77	0	1	0	0
77-106	0	2	0	18	0	0
106-147	3	0	0	153	18	0
147-174	4	0	0	13	175	2
174 Above	0	0	0	0	8	1

Table A.2: Confusion Matrix with State Data for Illinois

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 56	56-80	80-107	107-147	147-173	173
Below 56	0	3	0	1	0	0
56-80	6	40	29	3	1	0
80-107	1	20	87	27	7	0
107-147	1	4	57	55	25	0
147-173	0	1	0	1	119	0
173 Above	0	0	0	0	16	0

Table A.3: Confusion Matrix with Full Data for Illinois

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 61	61-101	101-125	125-161	161-196	196
Below 61	0	3	0	0	1	0
61-101	0	32	34	4	9	0
101-125	0	27	87	11	17	0
125-161	0	20	49	38	35	0
161-196	0	0	0	1	120	0
196 Above	0	0	0	0	16	0

Table A.4: Confusion Matrix with State Data for Indiana

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 46	46-84	84-131	131-163	163-204	204
Below 46	55	36	6	0	0	0
46-84	31	46	28	0	0	0
84-131	2	16	111	1	7	0
131-163	0	0	20	2	13	1
163-204	0	0	8	4	26	1
204 Above	0	0	0	0	19	12

Table A.5: Confusion Matrix with Full Data for Indiana

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 61	61-101	101-125	125-161	161-196	196
Below 61	69	21	7	0	0	0
61-101	36	29	40	0	0	0
101-125	5	9	116	0	7	0
125-161	0	0	20	0	14	2
161-196	0	0	6	2	29	2
196 Above	0	0	0	2	29	0

Table A.6: Confusion Matrix with State Data for Minnesota

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 53	53-75	75-108	108-136	136-159	159
Below 53	11	11	0	0	0	0
53-75	9	20	17	2	0	0
75-108	1	14	61	26	2	0
108-136	1	1	20	88	14	0
136-159	0	0	3	36	43	0
159 Above	0	0	0	0	4	0

Table A.7: Confusion Matrix with Full Data for Minnesota

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 61	61-101	101-125	125-161	161-196	196
Below 61	10	12	0	0	0	0
61-101	0	32	14	2	0	0
101-125	0	17	57	24	6	0
125-161	0	1	23	79	21	0
161-196	0	0	3	33	46	0
196 Above	0	0	0	0	4	0

Table A.8: Confusion Matrix with State Data for Nebraska

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 54	54-85	85-106	106-139	139-164	164
Below 54	0	4	2	2	0	0
54-85	0	60	25	10	1	0
85-106	0	28	83	18	5	0
106-139	0	4	38	81	26	0
139-164	0	0	0	12	50	0
164 Above	0	0	0	0	1	0

Table A.9: Confusion Matrix with Full Data for Nebraska

	Predicted yield (Bushels/acreer)					
Yield(Bushels/acre)	Below 61	61-101	101-125	125-161	161-196	196
Below 61	0	6	2	0	0	0
61-101	0	54	34	7	1	0
101-125	0	27	86	16	5	0
125-161	0	6	43	74	26	0
161-196	0	0	0	12	50	0
196 Above	0	0	0	0	1	0

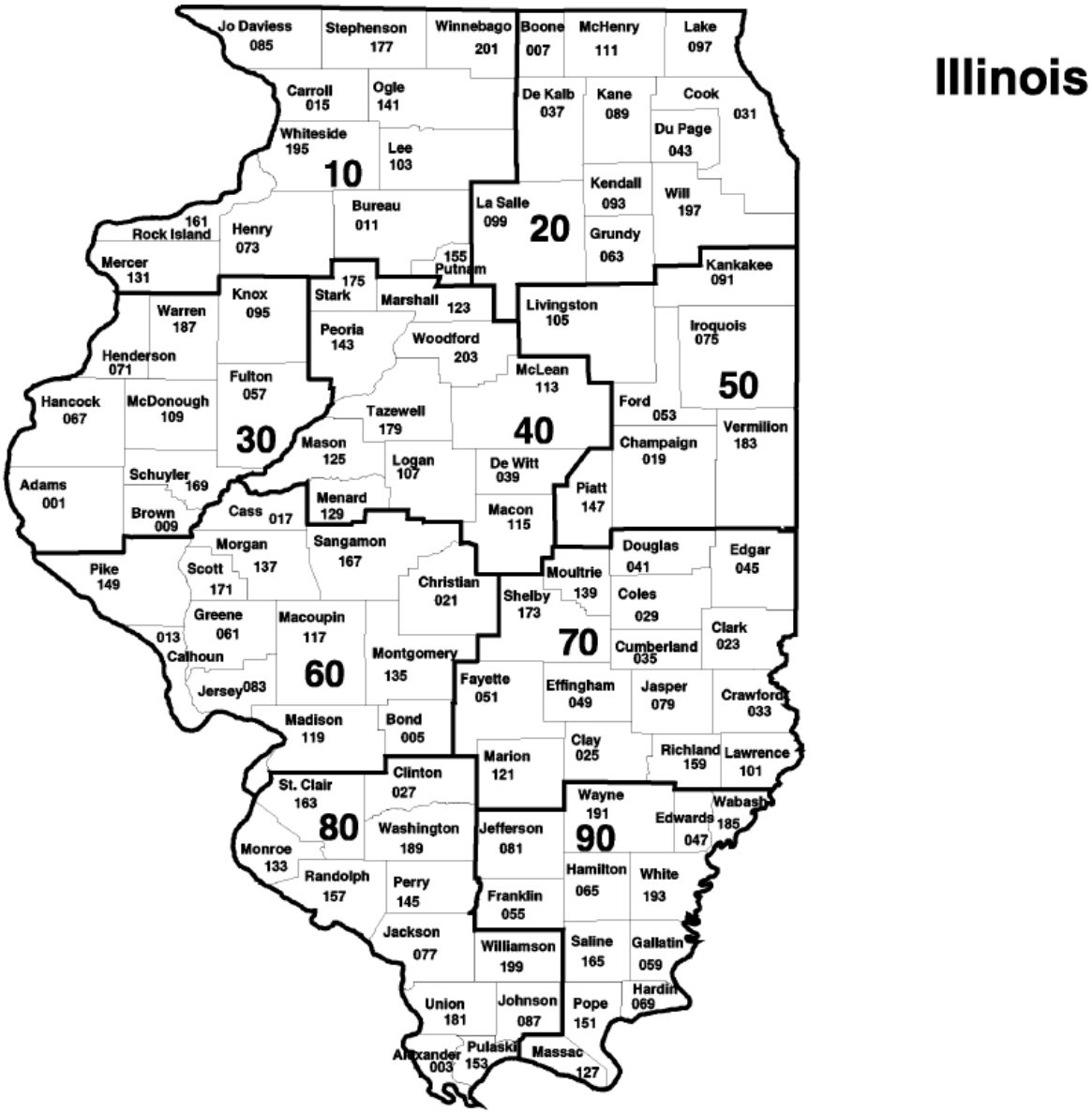


Figure A.1: Illinois crop reporting district map

Figure A.2: Indiana crop reporting district map.

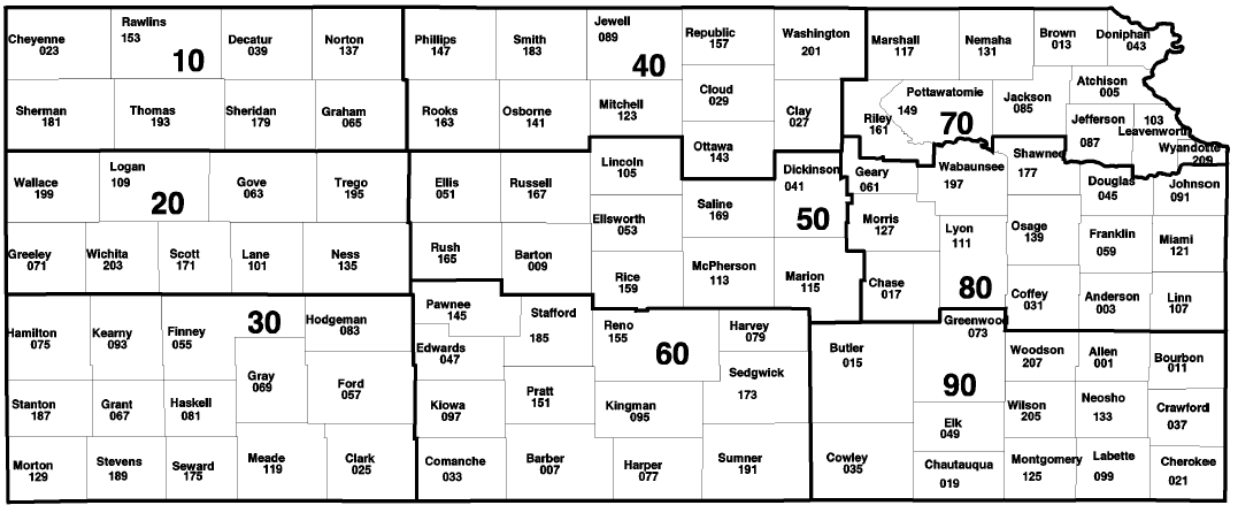


Figure A.3: Kansas crop reporting district map

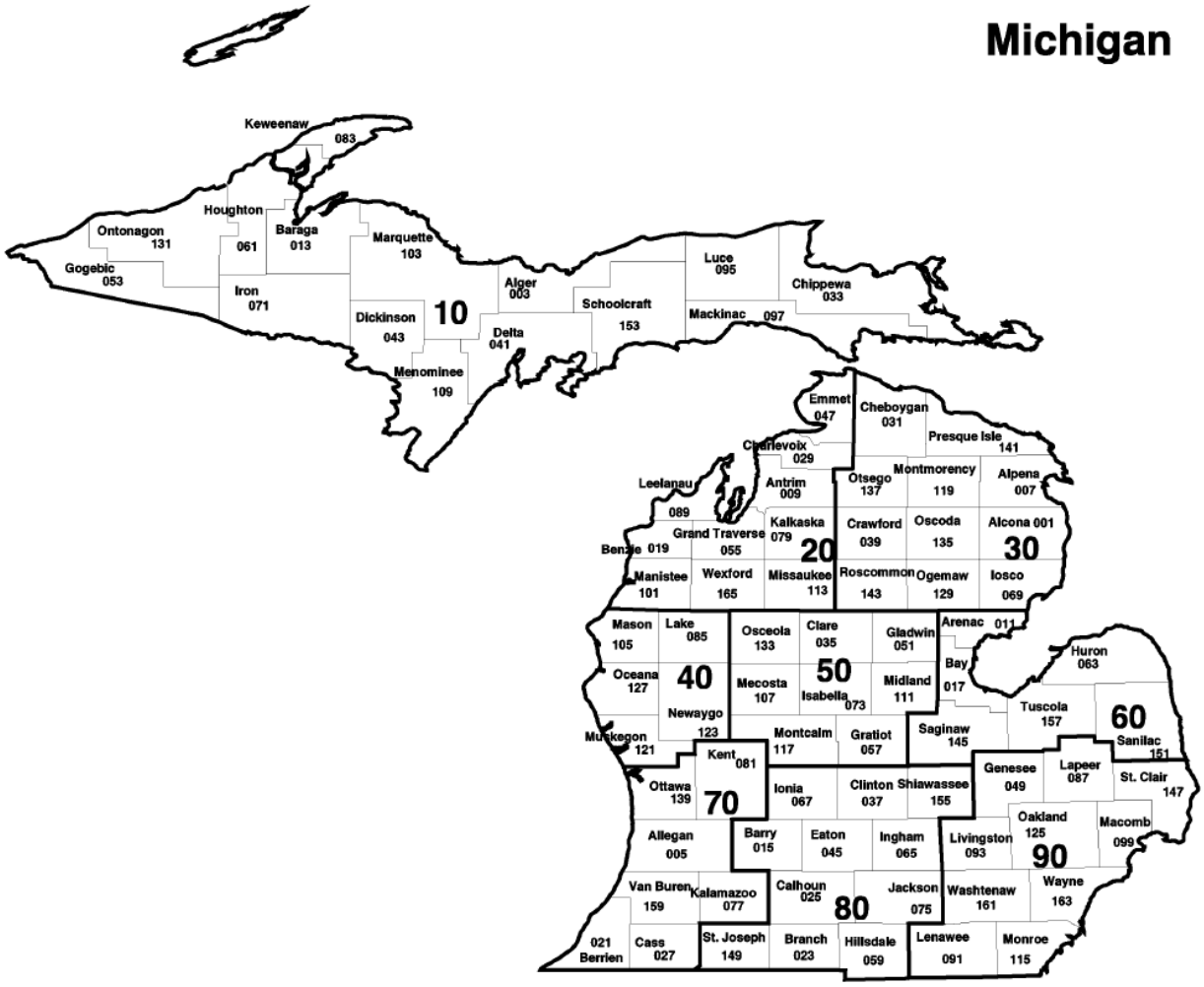


Figure A.4: Michigan crop reporting district map

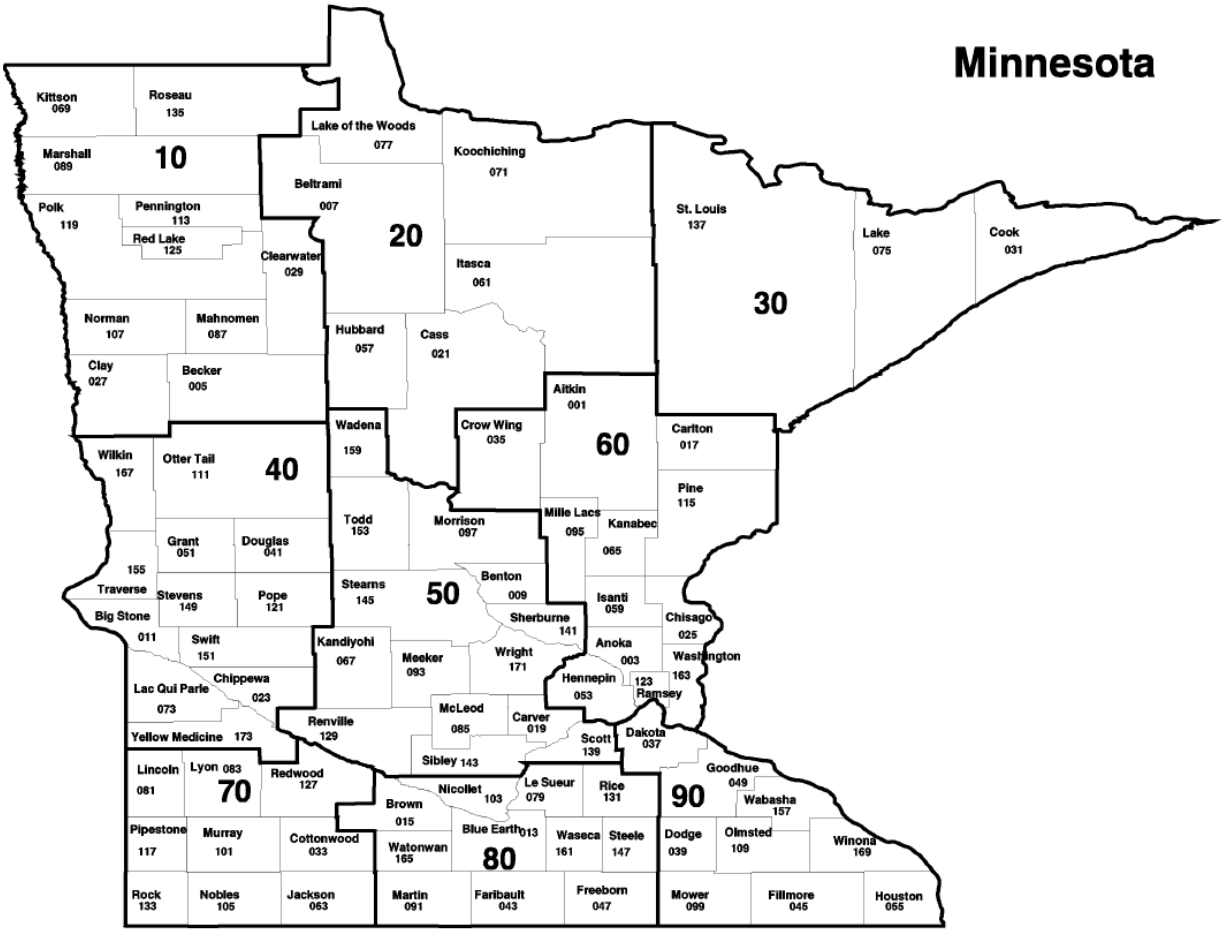


Figure A.5: Minnesota crop reporting district map

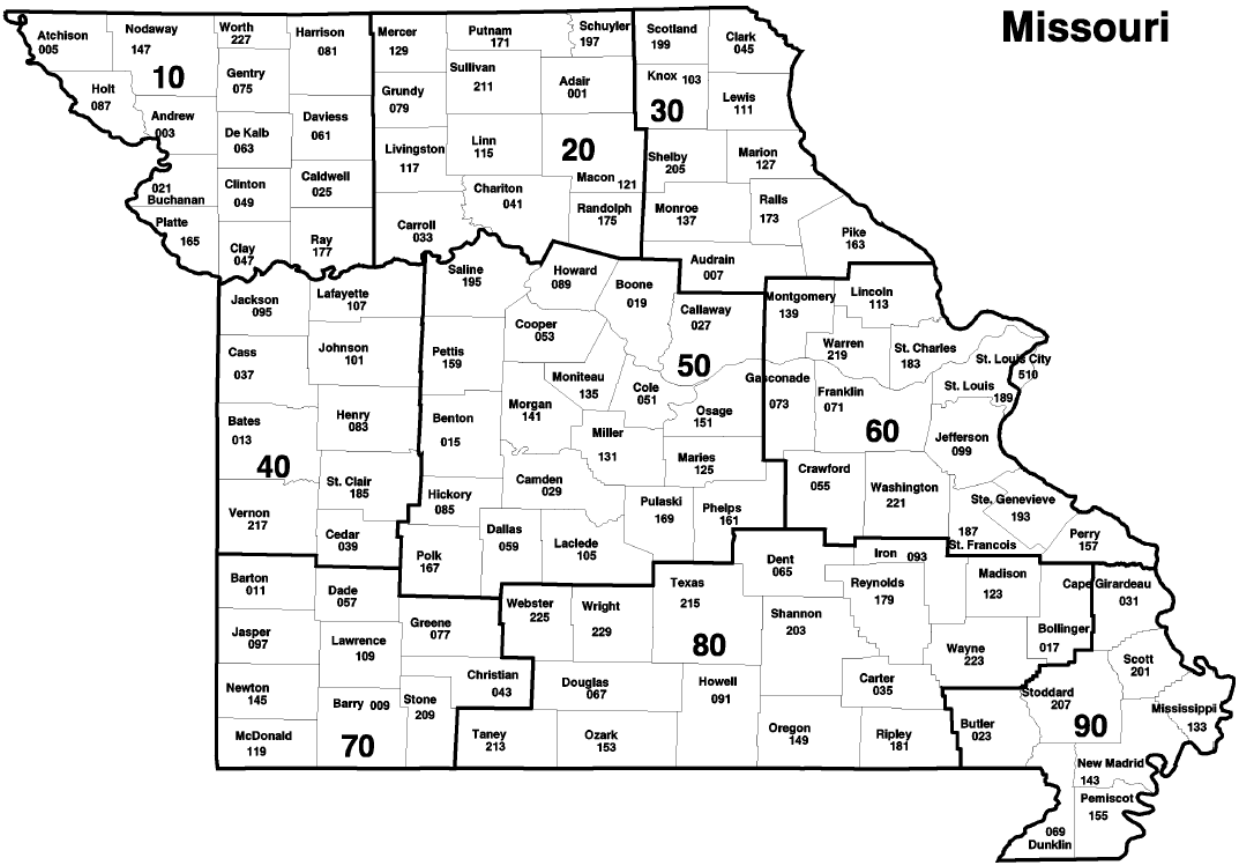


Figure A.6: Missouri crop reporting district map

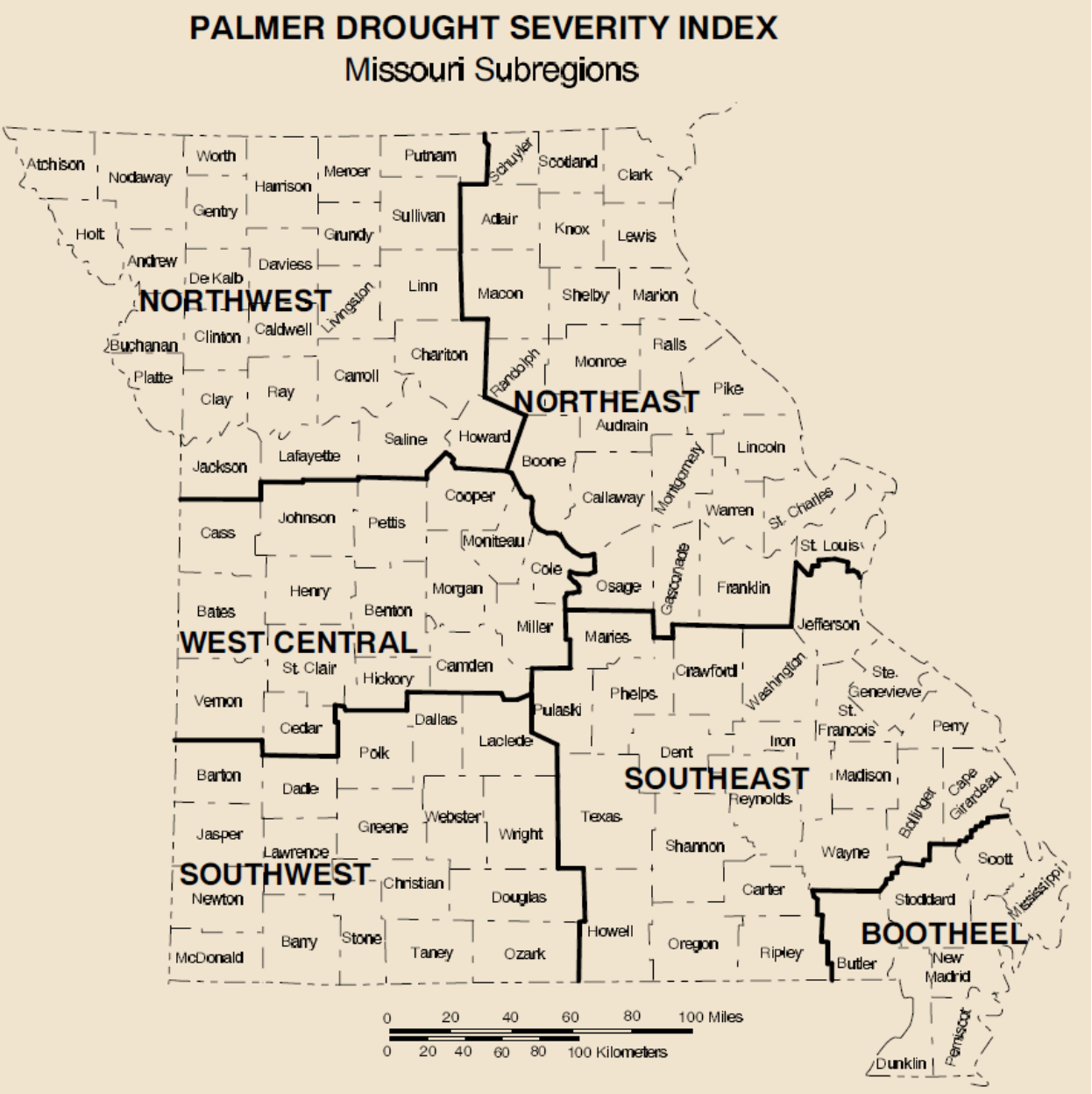


Figure A.7: Missouri PDSI sub-regions map

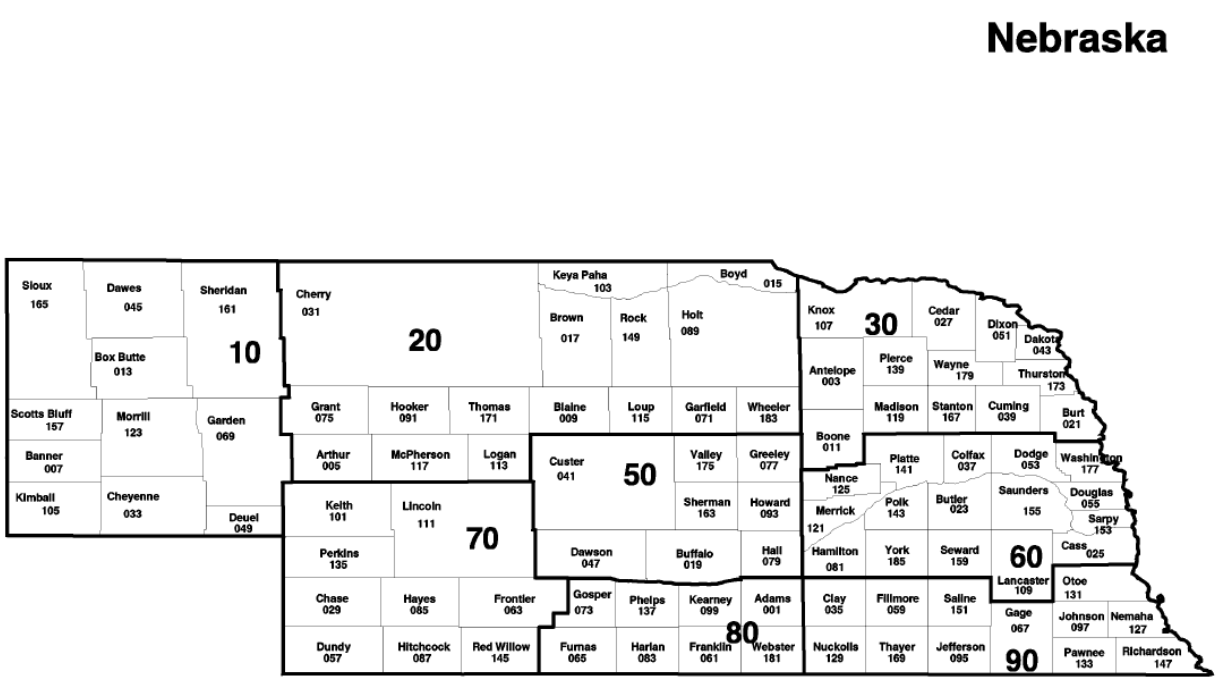


Figure A.8: Nebraska crop reporting district map

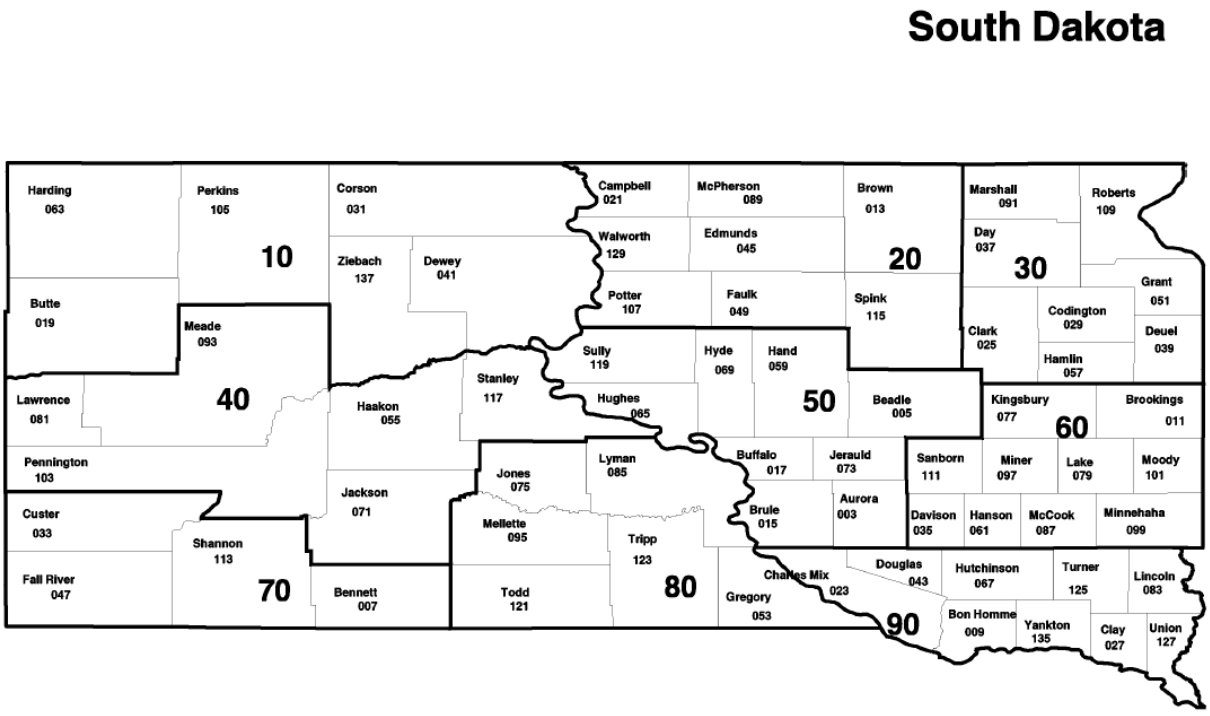


Figure A.9: South Dakota crop reporting district map

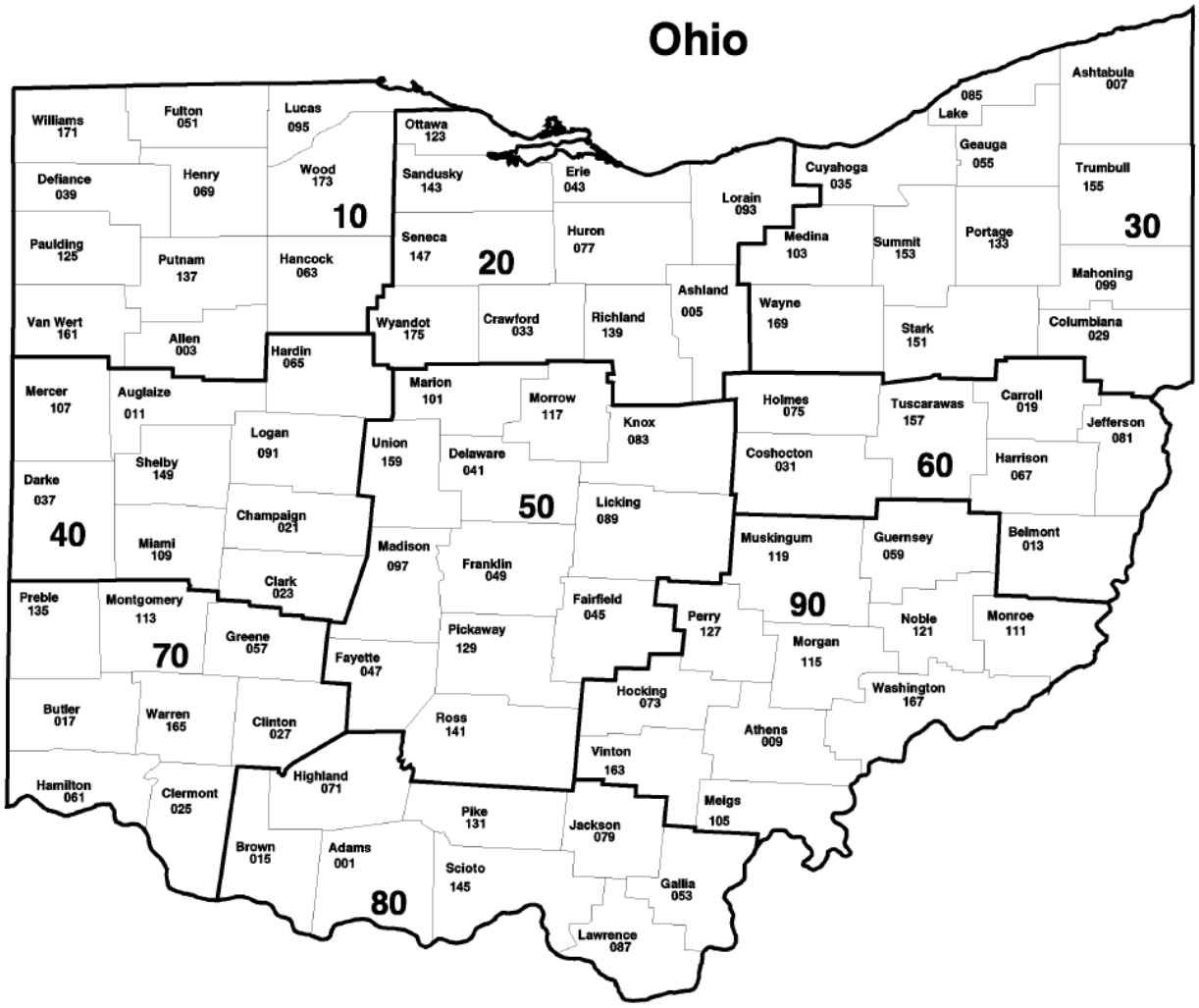
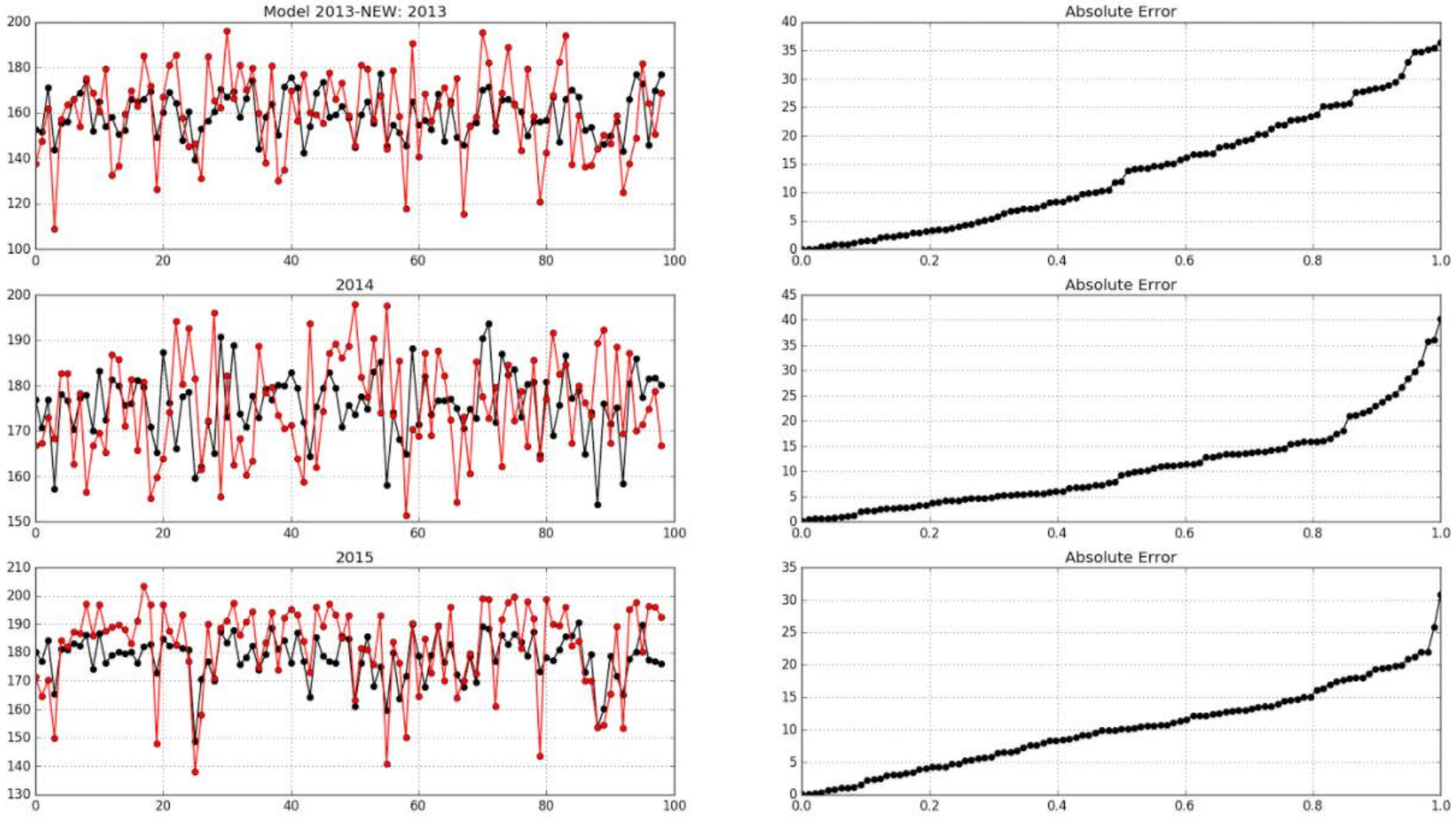
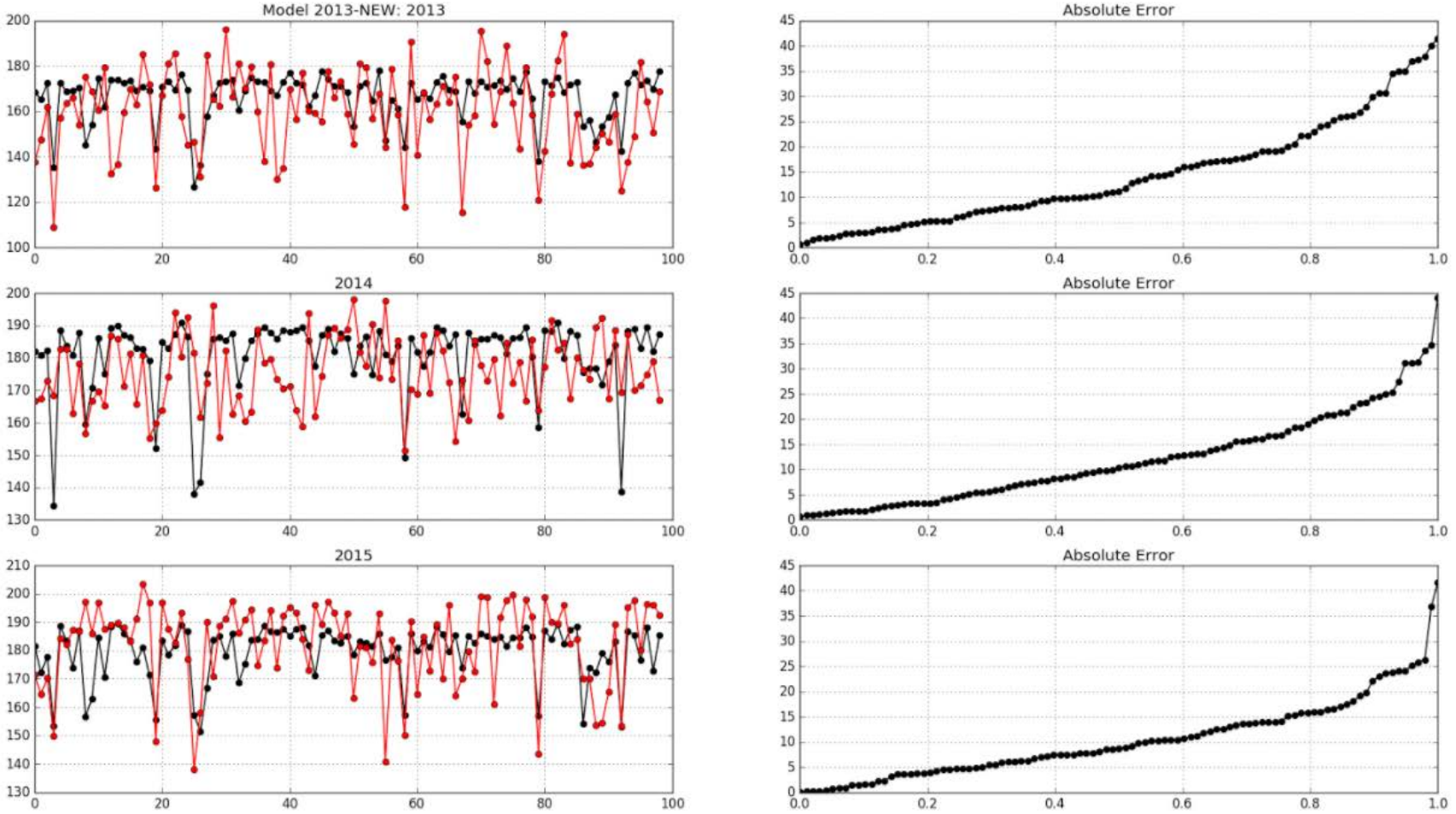


Figure A.10: Ohio crop reporting district map



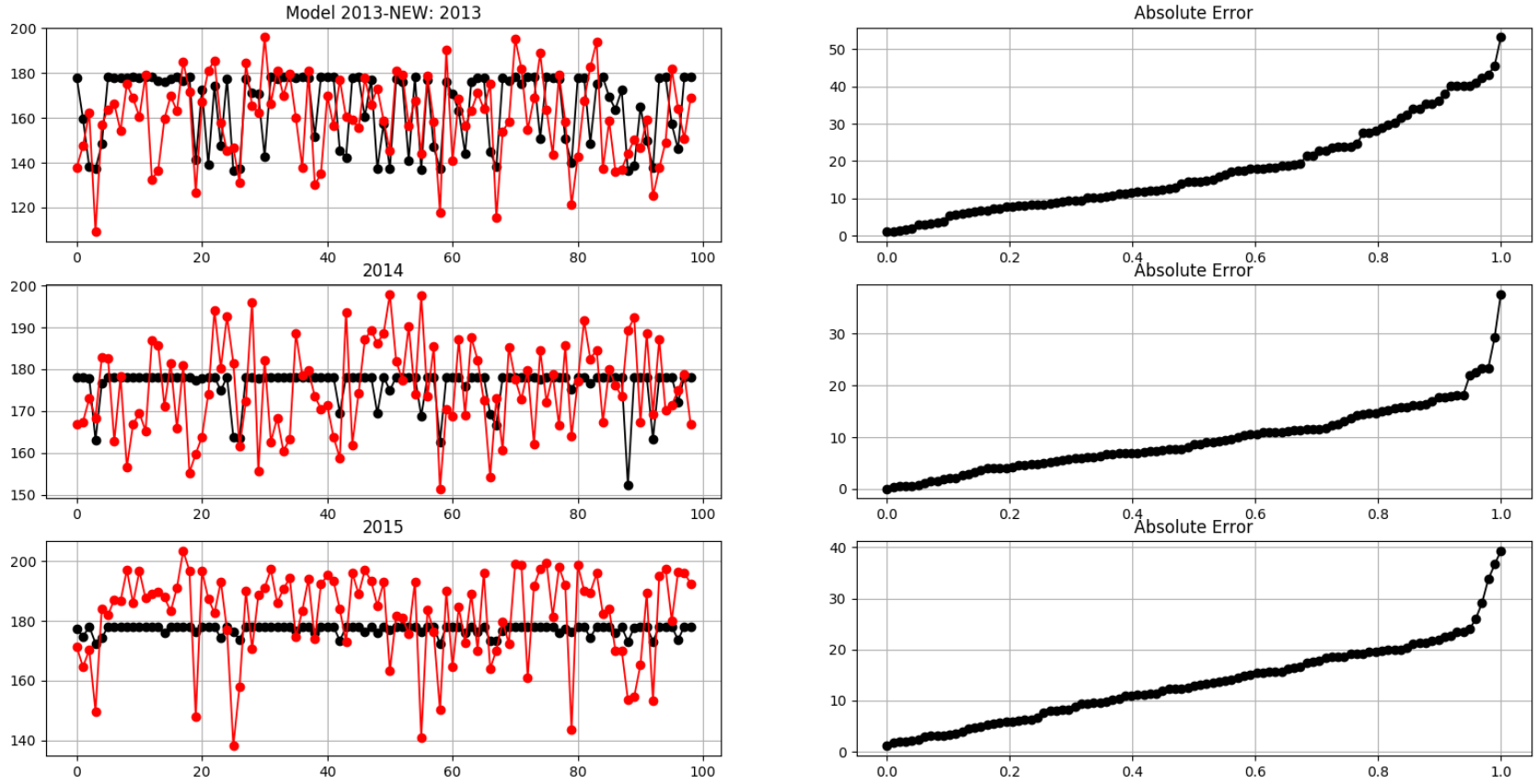
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.11: Prediction results for one layer LSTM with combination samples



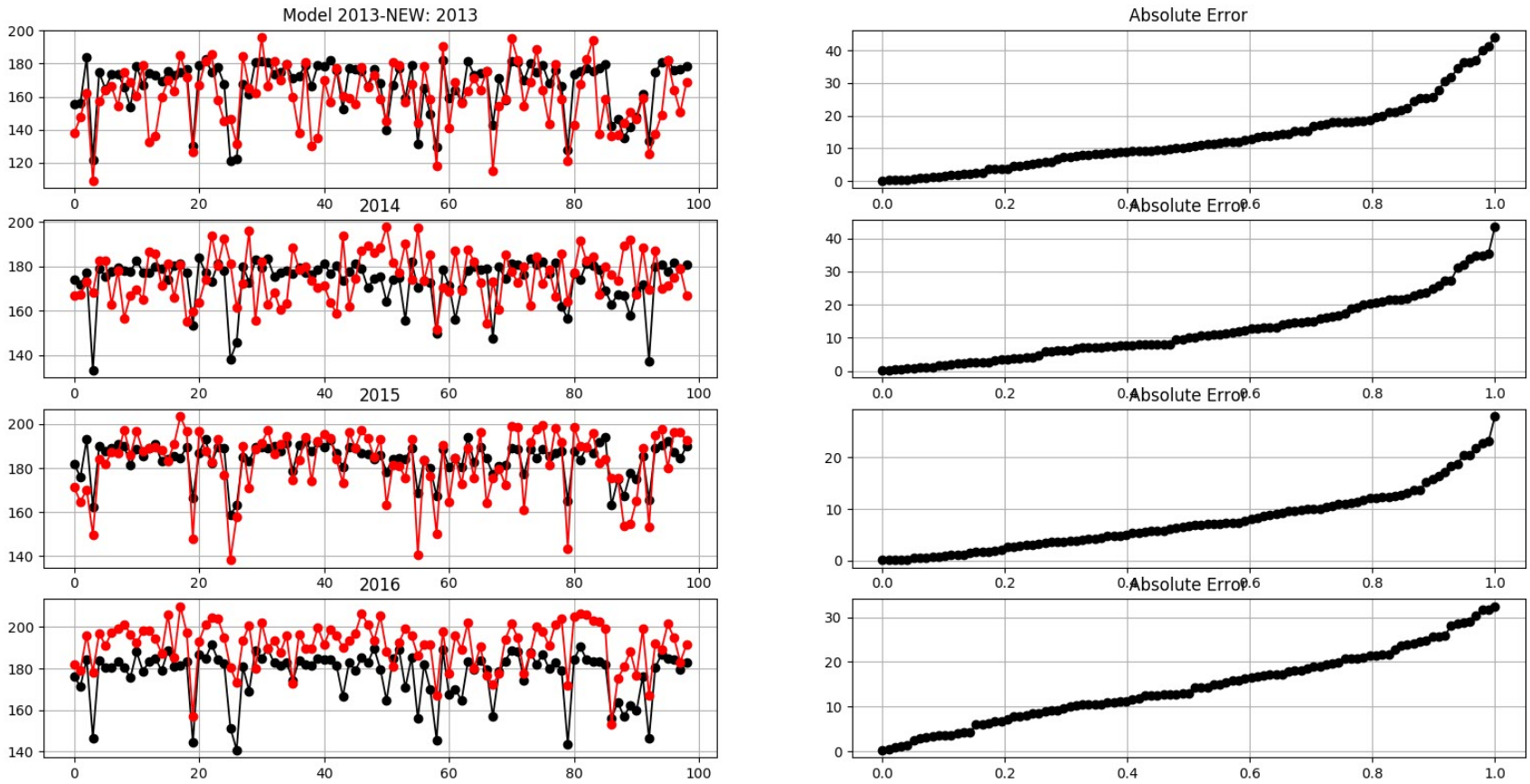
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.12: Prediction results for two-layers LSTM without combination samples



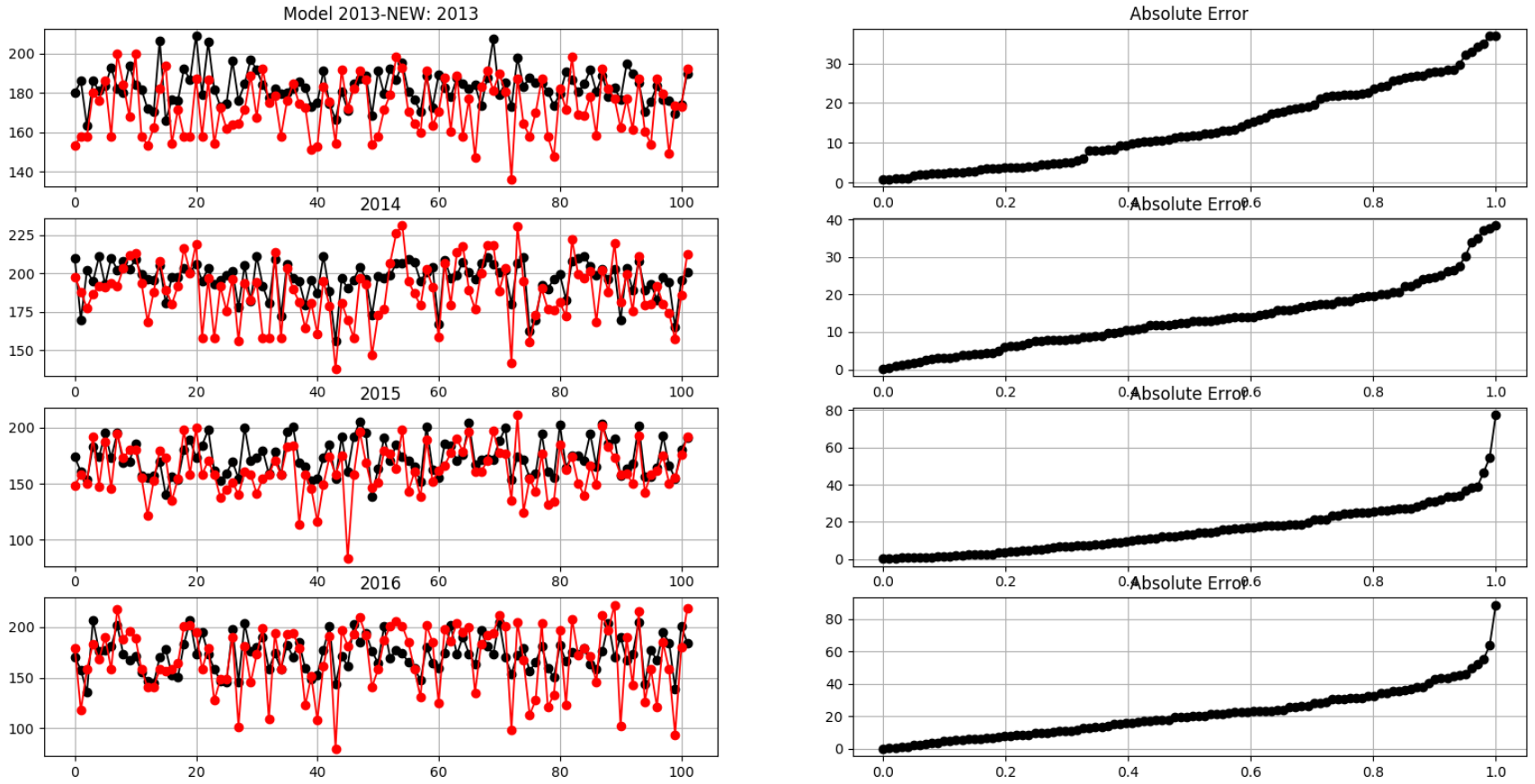
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.13: Prediction results with 28 input variable



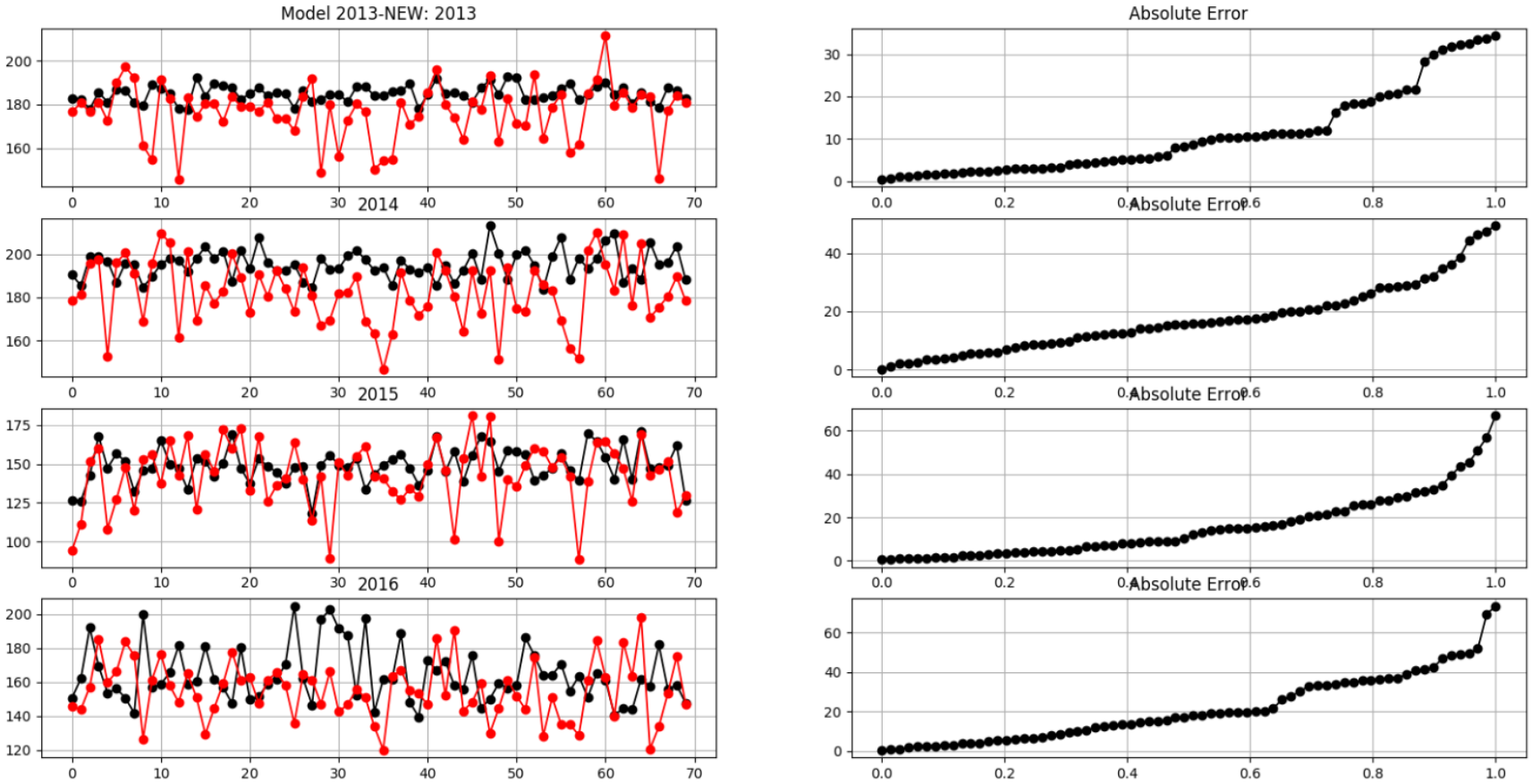
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.14: Prediction results for the “best” Iowa LSTM model with percentage adjustment



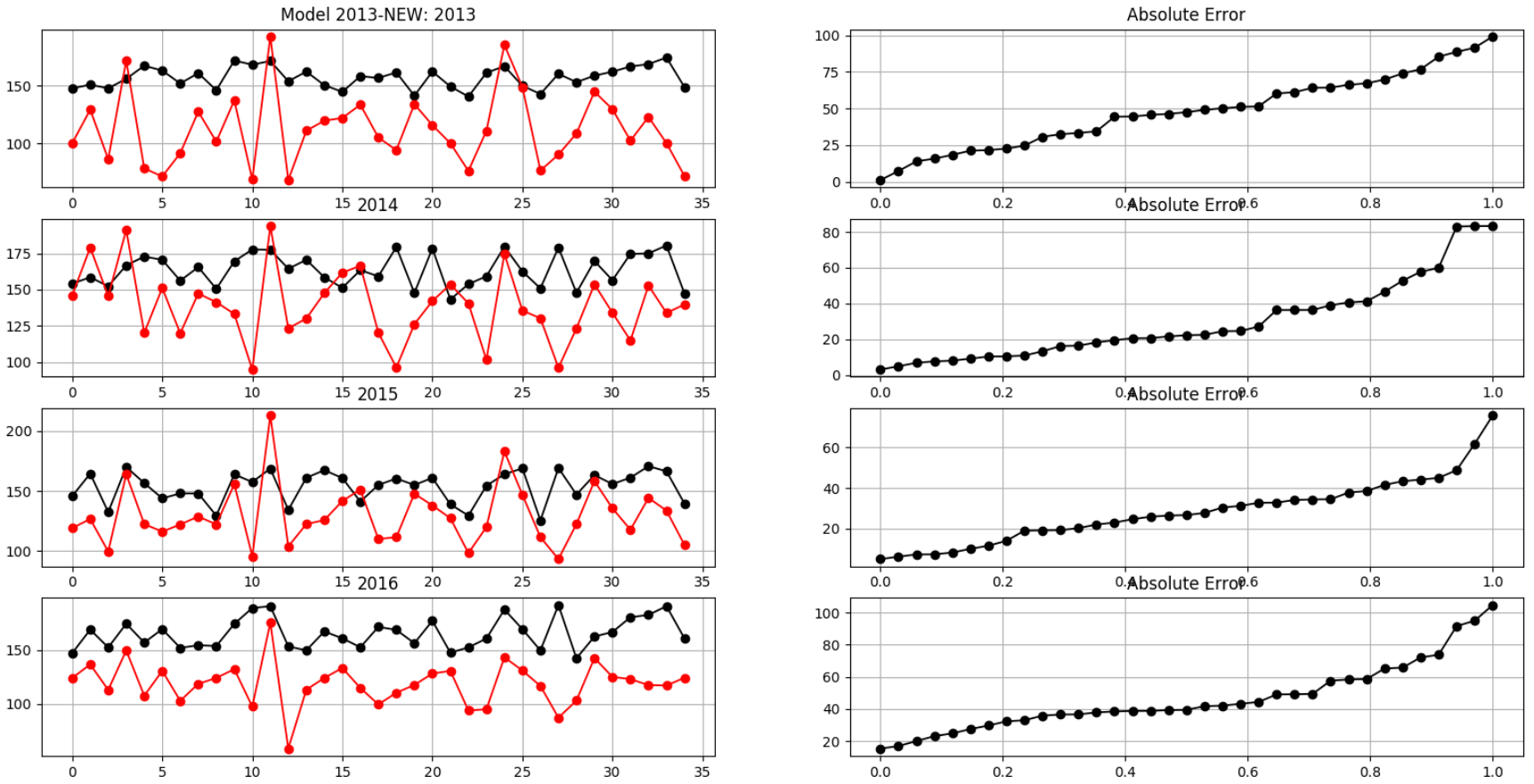
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.15: Prediction results for the “best” Illinois LSTM model with constant adjustment



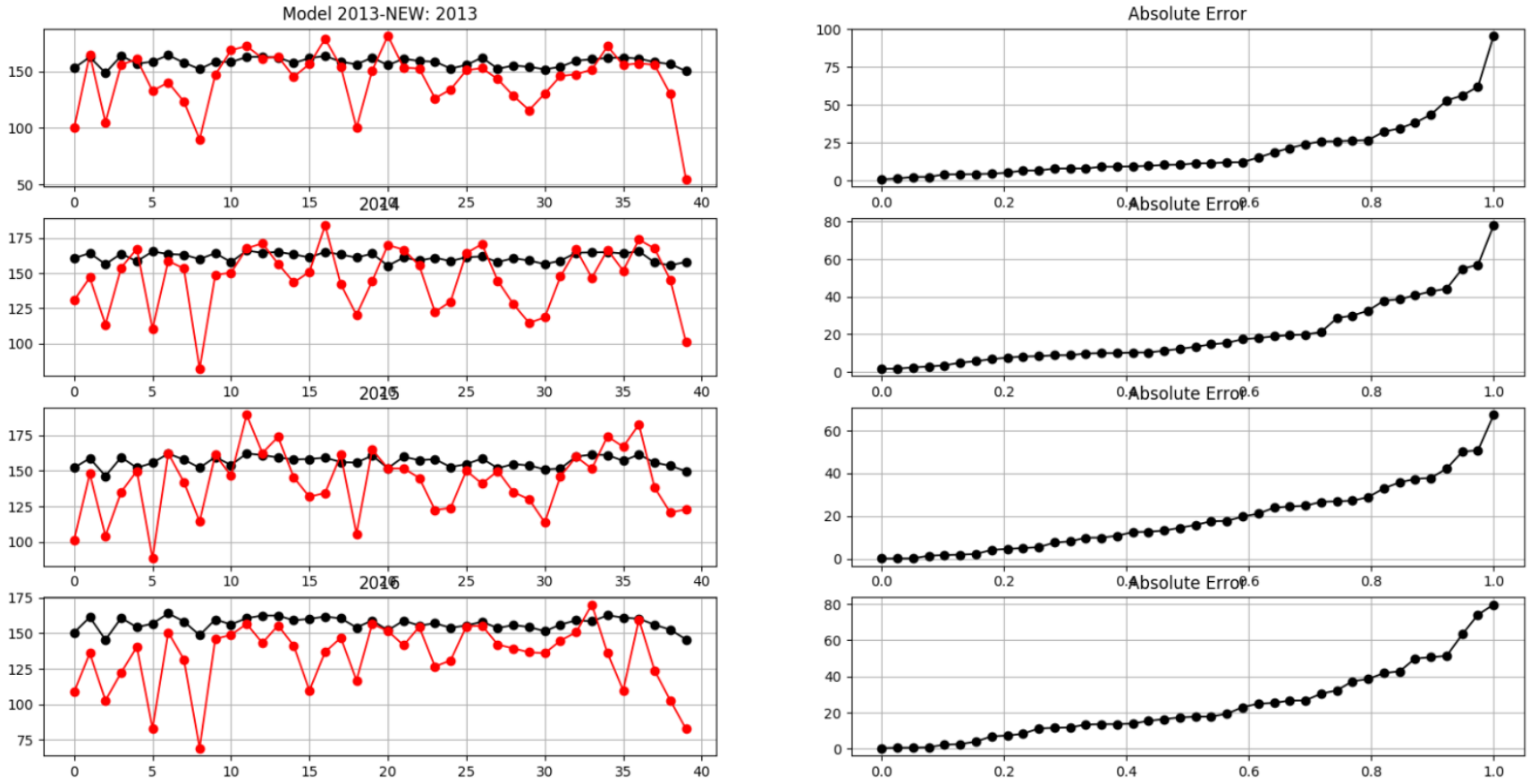
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.16: Prediction results for the “best” Indiana LSTM model with constant adjustment



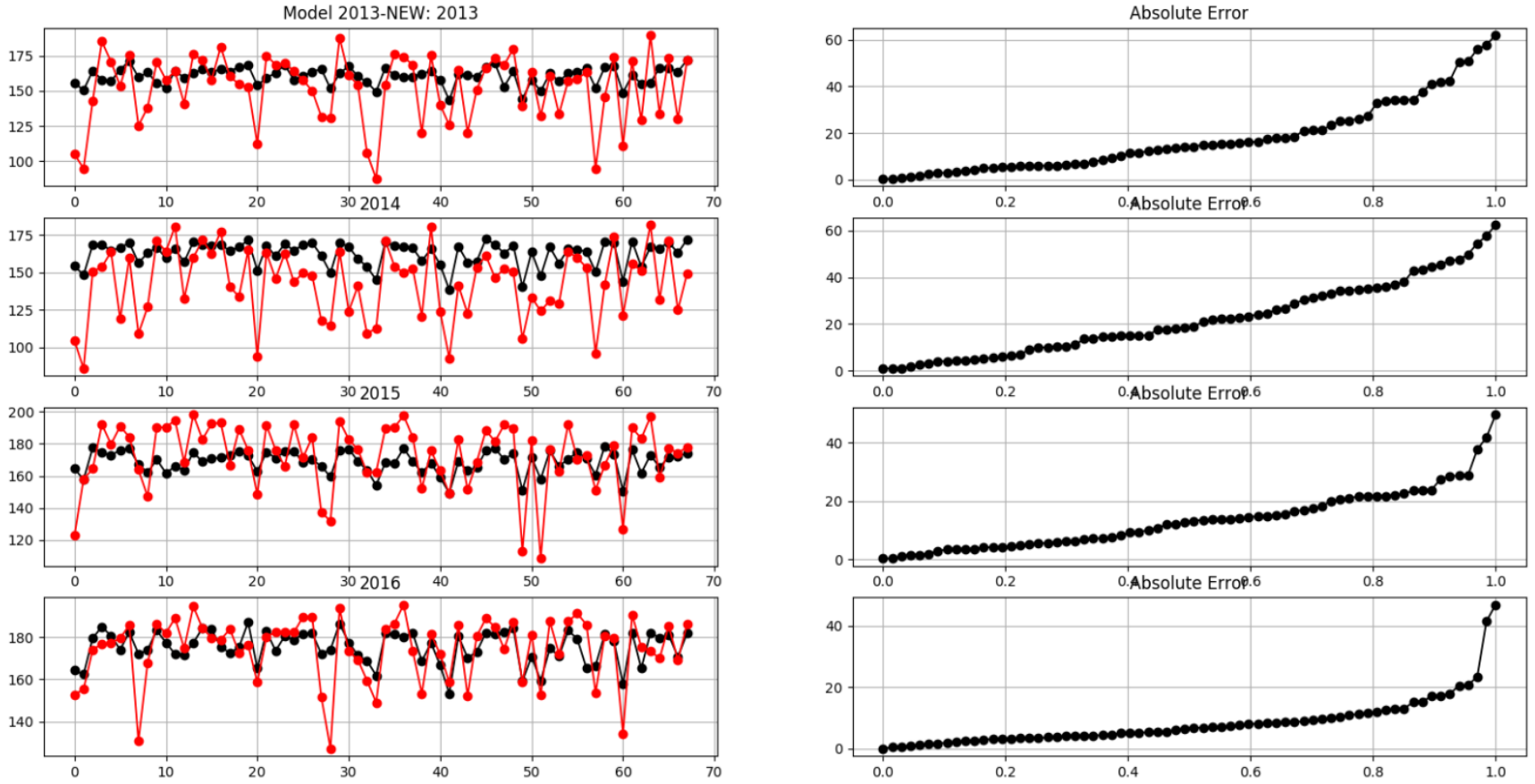
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.17: Prediction results for the “best” Kansas LSTM model with constant adjustment



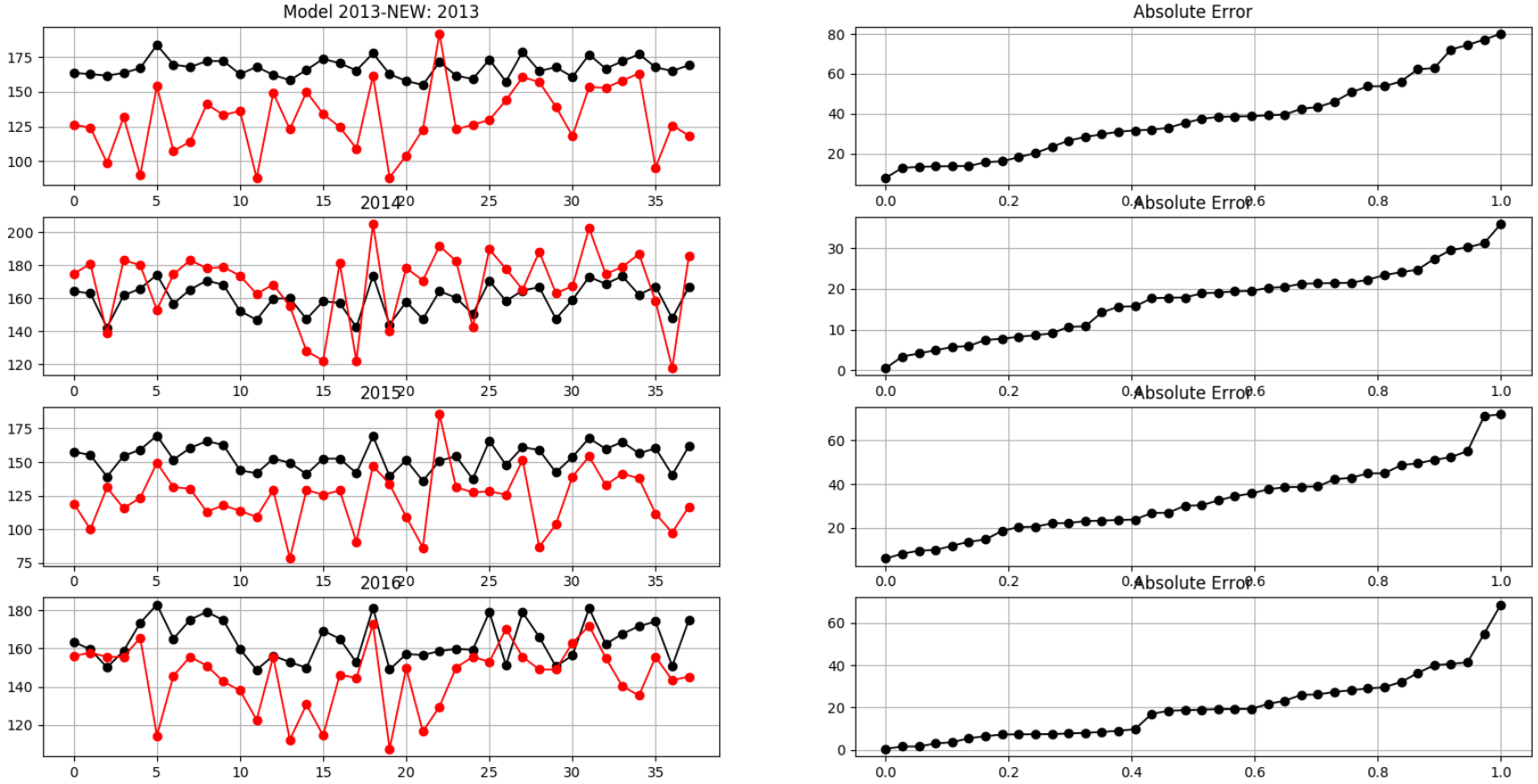
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.18: Prediction results for the “best” Michigan LSTM model with constant adjustment



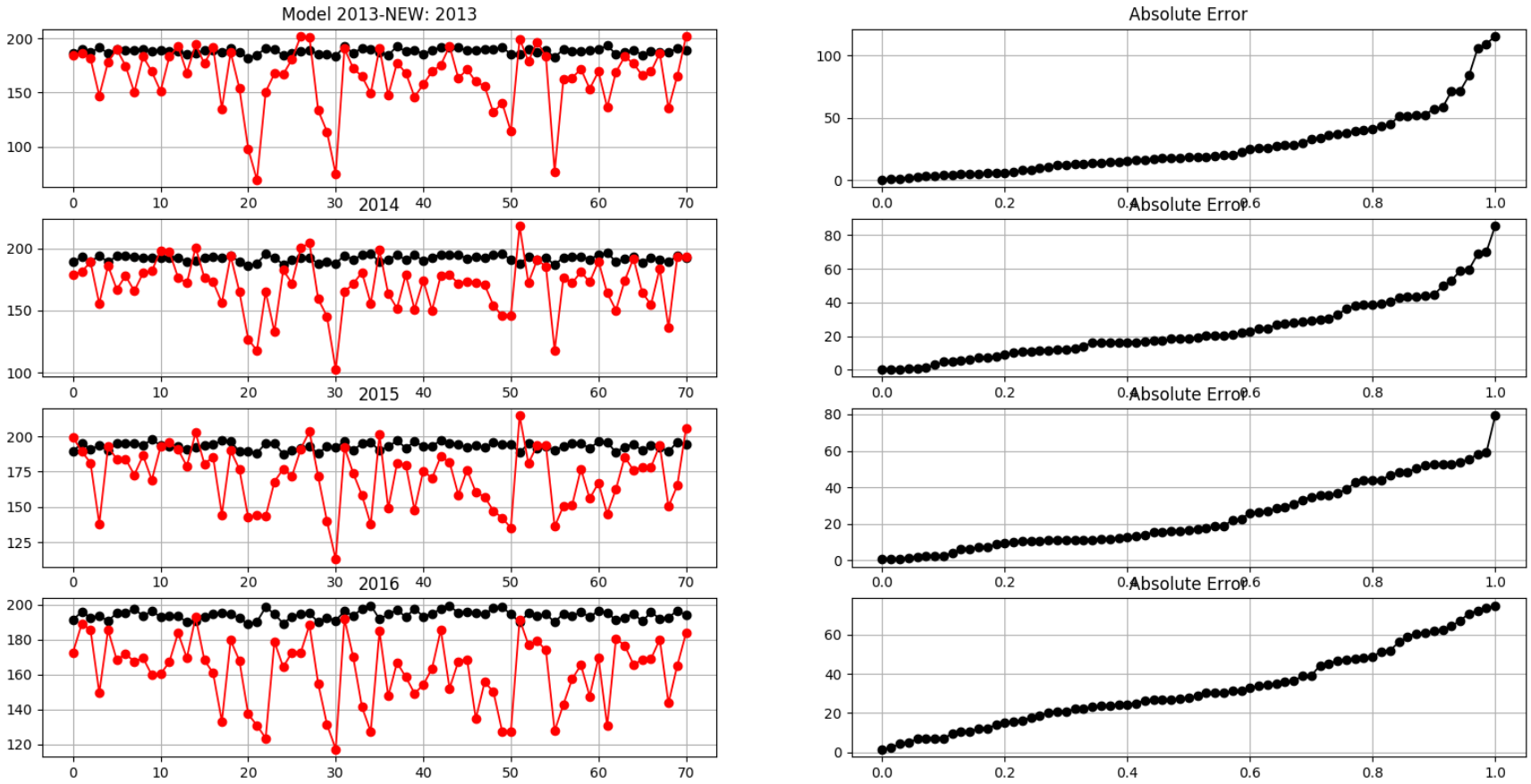
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.19: Prediction results for the “best” Minnesota LSTM model with constant adjustment



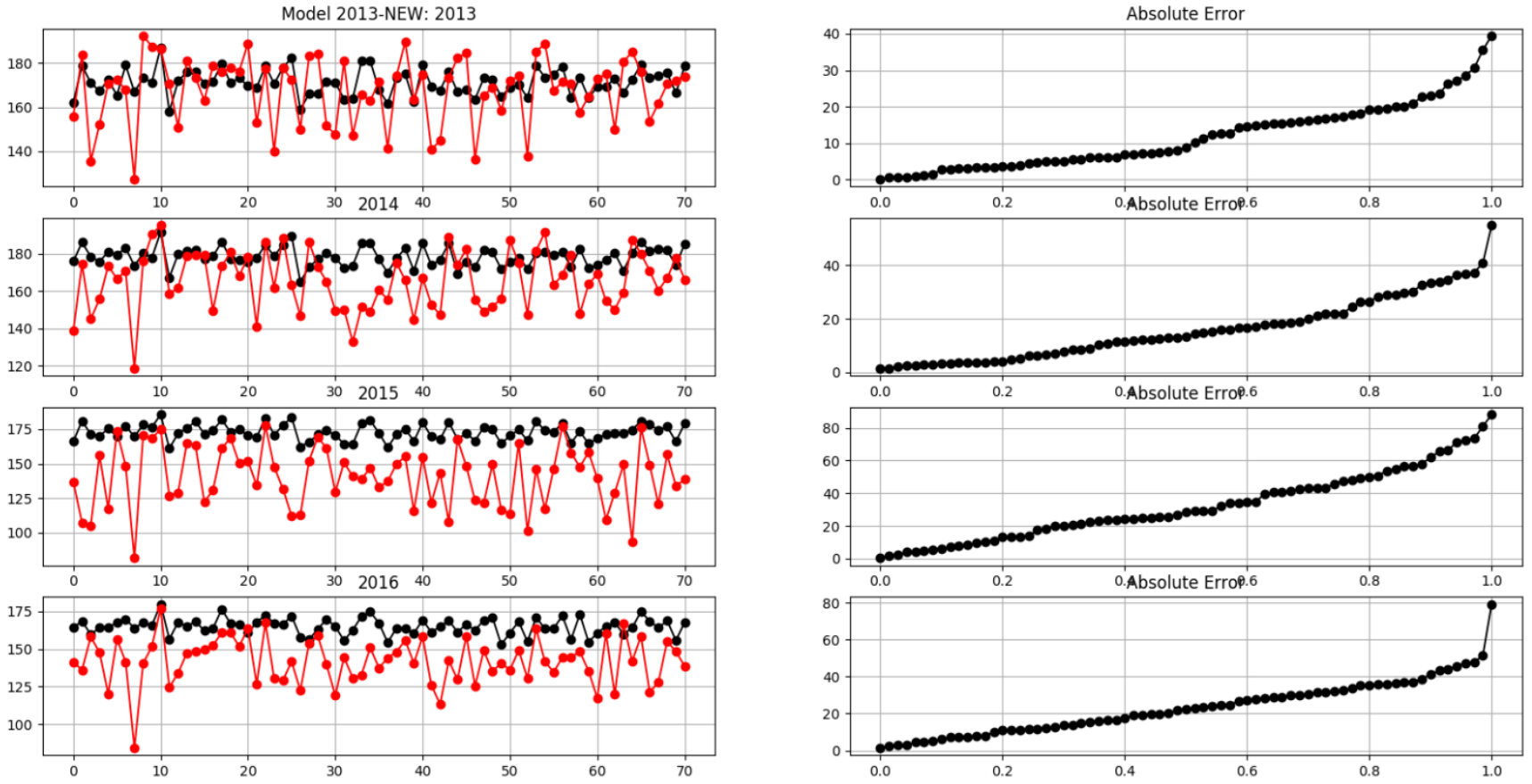
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.20: Prediction results for the “best” Missouri LSTM model with constant adjustment



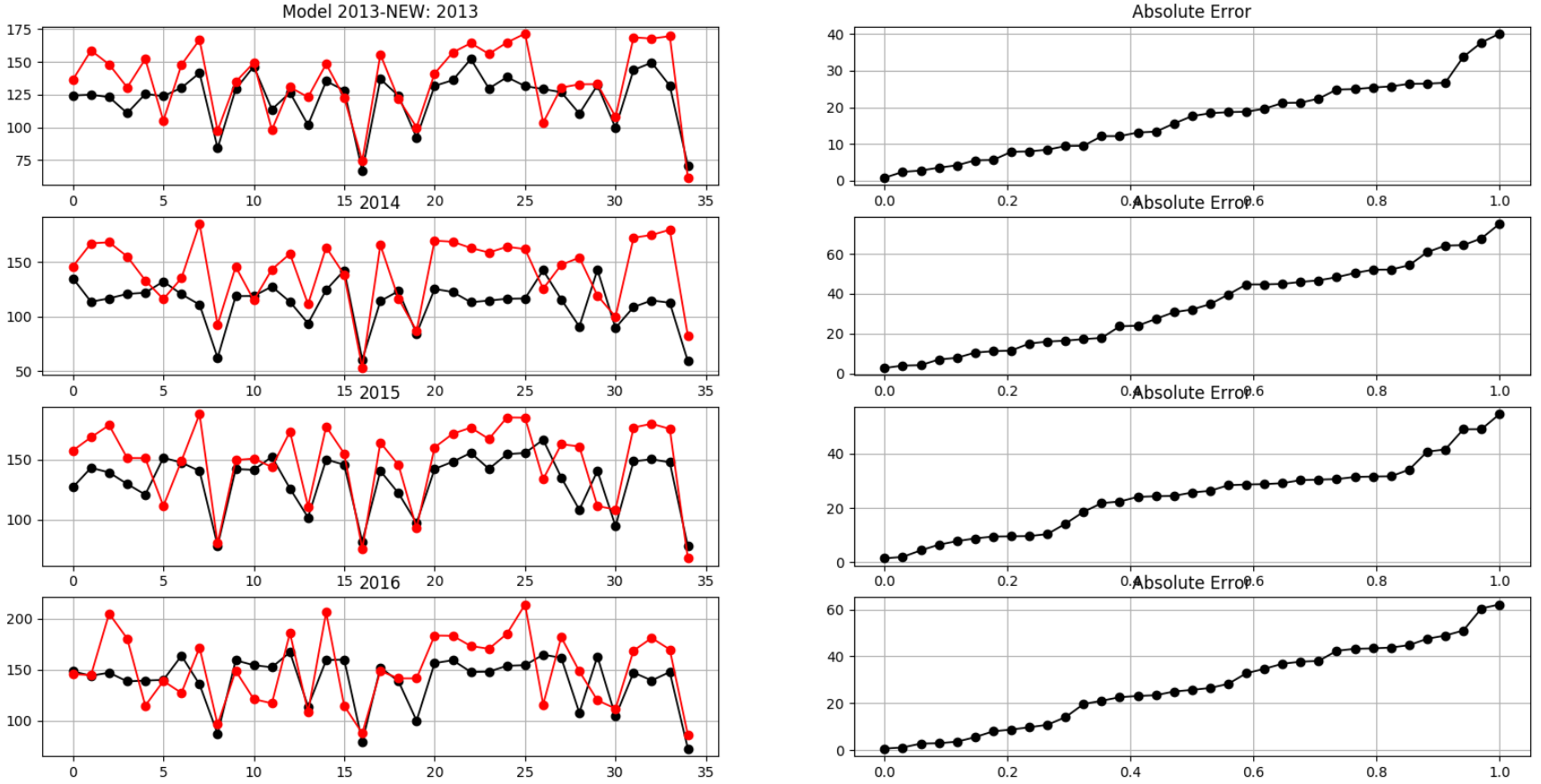
Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.21: Prediction results for the “best” Nebraska LSTM model with constant adjustment



Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.22: Prediction results for the “best” Ohio LSTM model with constant adjustment



Note: Red line is the actual yield, black line is the LSTM prediction

Figure A.23: Prediction results for the “best” South Dakota LSTM model with percentage adjustment