Yield and Price Forecasting for Stochastic Crop Decision Planning

Nantachai Kantanantha, Nicoleta Serban and Paul Griffin [1]

Crop decision planning is an important part of effective farm management. Because of the many uncertain factors such as weather variations, technology advances, and crop yields and prices, all of which prove to change considerably, decision planning can be very complex. The focus of this paper is to develop accurate yield and price forecasting models to aid in decision planning. For yield forecasting, we establish a crop-weather model using regression methods. We propose a semiparametric regression model, which accounts for both within- and between-year relationships in the data. For price forecasting, we develop a futures-based model, which predicts the cash price from futures price and basis. We estimate the yearly basis pattern using a functional model-based approach and adjust the basis estimates with the futures prices to forecast the crop cash price. In both forecasting models, we estimate confidence bands for the predictions that can be further integrated in the decision planning model.

Key words and phrases: functional principal component analysis, functional linear regression, p-splines, functional model-based clustering, crop decision planning, yield forecasting, price forecasting.

1 Introduction

Agriculture and related businesses are large industries in the U.S. In 2004, they had a value added of approximately \$141.6B, a 44% increase from \$98B in 2000 (Bureau of Economic Analysis) and employed about 2.1 million workers (Bureau of Labor Statistics). In addition, the agribusiness has become very complex in recent years, and hence the importance of agricultural planning has increased.

Crop producers often suffer from a lack of accurate information on which to base decisions for crop production and evaluation. Two examples are yield and price

[1]Nantachai Kantanantha is PhD candidate, Nicoleta Serban is Assistant Professor and Paul Griffin is Professor in the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology.

forecasts. The objective of this paper is to develop accurate yield and price forecasting models along with prediction confidence bands. This will allow us to explore the effects of yield and price behavior in the context of a decision planning model.

Timely and accurate crop yield forecasting is essential for crop production, marketing, storage, and transportation decisions and also helps managing the risk associated with these activities (Bannayan and Crout 1999, Lee 1999, Potgieter et al. 2005). The most well-known and widely used crop information source is the monthly USDA Crop Production report (Krog 1988). These reports, prepared by the National Agricultural Statistics Service (NASS), provide statistics and related information on crop production in the US. Even though these reports supply crop yield forecasts that are broadly utilized, they generate only the mean estimate for each state (Lee 1999). This numerical estimate may not reflect the true yield in any particular area in the state. Moreover, the projected grain yields, which are released monthly from August through November, may not be updated in a timely way that allows farmers to make decisions at the beginning of the planting season.

Crop growth and development are largely affected by environmental conditions. This can result in significant variation in crop yields by year and location. Many previous studies have incorporated the effects from temperature (Peng et al. 2004, Wheeler et al. 2000, Batts et al. 1997, Mitchell et al. 1993), rainfall (Mkhabela et al. 2005, Seif and Pederson 1978), or several weather factors (Hoogenboom 2000, Kandiannan et al. 2002) on yield. In line with the current literature, we develop a crop-weather model using a regression approach where the weather factors are rainfall and temperature. Additionally, we account for the economic growth by including GDP as a regressor. Unlike the classical parametric regression model that considers only the relationship of the data within one particular year, the semiparametric regression model introduced in this paper takes into account the within- and between-year relationships of the data.

Estimating future selling prices of crops is another important piece of information for farmers since it helps to determine what crops they should plant and how much return they would receive. Since commodity prices are volatile, growers and agribusinesses have to rely on price forecasting. The U.S. Department of Agriculture (USDA) publishes reports on crop supply, demand, yields, prices, and other related information. Agricultural prices are provided by the World Agricultural Outlook Board (WAOB) in the World Agricultural Supply and Demand Estimates (WASDE) report. However, the WASDE report is distributed once a month and the prices are national averages. Consequently, a crop price forecasting model is needed to predict the upcoming prices in any specific location.

A number of models have been developed to forecast the cash prices. Many researchers study the role of futures contract prices in agricultural price forecasting (Dow 1940, Gardner 1976, Kenyon et al. 1993, Tomek and Gray 1970, Working 1942). Futures price is found to be a good measure of the cash price (Eales et al. 1990). In this research, the price forecasting model is developed under a futures-based framework where the cash price is forecasted from the futures price and basis.

We predict the cash price by obtaining a forecast of the basis distribution over one year. We derive the cash price forecast by adding the basis forecast and the expected future price. In order to estimate the one-year basis distribution, we use a functional model-based approach. As we obtain the distribution rather than expectation alone, we can also compute prediction confidence bands for the one-year basis forecast.

The layout of the paper is as follows. Data used for yield and price forecasting are discussed in Section 2. The methods explored for yield forecast are described in Section 3. We apply our model to corn and soybean yields and weather data in Illinois. The data acquired for evaluating yield forecast models is discussed in Section 2.1. Fit and prediction results of the yield forecasting model are presented in Section 4.4. In Section 2.2, we present the cash and futures price data evaluated in this paper. The methodology on basis and cash price forecasts is introduced in Sections 5.1 and 5.2. The proposed price forecasting model is applied to corn and soybean prices in Illinois. Our results on price forecasts are discussed in Section 6. Finally, in Section 7 we discuss how these results could be used in crop planning tools as well as future work.

2 Data Background

2.1 Data for Yield Forecasting

We base our yield forecast on historical corn and soybean data, weather data and nominal GDP. Historical corn and soybean yield data are acquired from Quick Stats, an agricultural statistics database, provided by the National Agricultural Statistics Service. Yield data are expressed as a number of bushels harvested per acre. Both corn and soybean yield data are from Hancock County in Illinois from 1927 to 2005. We chose Illinois as our primary state in our study since based on the Agricultural Statistics report (National Agricultural Statistics Service 2005), Illinois is the largest soybean producer, and the second largest corn producer in the US. Hancock county is chosen as a representative county in Illinois because its corn and soybean yields in 2005 are about the same level as the state average. However, our methodology applies to any yield producer across the country.

In our weather based model, we use the weather data from National Climatic Data Center (NCDC). These data are collected from La Harpe station in Hancock County, Illinois, from 1927 to 2005. The rainfall variable is the total monthly rainfall in inches and the temperature variable is the average monthly temperature in degree Fahrenheit during the cropping season. Based on the Usual Planting and Harvesting Dates for U.S. Field Crops report (National Agricultural Statistics Service 1997), in Illinois, corn is usually planted around the end of April and harvested in late September. Similarly, soybean is planted in the beginning of May and harvested in late September. Therefore, we include only the temperature and rainfall data from May to September in our study.

Another variable used in yield forecasting is annual GDP from 1926 to 2004. We use the nominal GDP acquired from Economic History Services. We utilize the one-year lag nominal GDP since the current yield is the reflection of past year's economic growth. We evaluate our forecast by dividing the data into training data including years 1927 to 1995, and testing data including years 1996 to 2005.

2.2 Data for Price Forecasting

We base our yield price forecast on both futures and cash price data. They are used to calculate the basis by subtracting futures price from the corresponding cash price. The basis reflects the local market conditions which are influenced by several factors, such as local supply and demand conditions, interest, storage costs, transportation costs, handling costs, and profit margins. We focus on forecasting the basis because it typically does not vary as much as cash price and can generally be predicted from historical basis patterns (Chicago Board of Trade 2000).

The futures price data are acquired from an agricultural package provided by the Chicago Board of Trade. This package includes Open, Close, High, Low, and Settlement futures prices of corn and soybean for every trading day. Investors can trade several years in advance. A futures contract is classified by its delivery month or contract month. However, there are specific delivery months for each crop. Corn futures contracts are delivered only in March, May, July, September, and December. On the other hand, soybean futures contracts can be delivered in January, March, May, August, September, and November. In this research, we select the nearby settlement price to represent the futures price. A nearby contract is the futures contract that is closest to expiration. For example, December corn futures is the nearby futures for corn in October. The settlement price is determined by averaging a range of closing prices.

We acquire the cash price data from the USDA Springfield regional office. This office provides the average cash prices of corn and soybean traded in central Illinois. Both futures and cash prices are collected every business day from 1991 to the first quarter of 2006. We compute the basis from nearby futures and cash prices. Basis values from years 1991 to 2003 form the training data and basis values from years 2004 and 2005 are used to evaluate our forecasting model (testing data).

3 Yield Forecasting: Method

3.1 Model Formulation

In this section, a semiparametric crop-weather regression model is developed to predict the crops' yield. The weather factors used in the model are rainfall and temperature during the growing season. In addition to these factors, we also incorporate the GDP macroeconomic indicator to account for the economic growth, which indirectly reflects in the yield change over time. Since we expect a considerable advance in agricultural technology (i.e. new equipment, better seed, better fertilizer, etc.) over the past few decades, we allow for the technology change through the mean function, which may vary with time only. A functional linear regression analysis is applied to find the relationship between the response variable, which is yield, and the predictor variables, which are temperature, rainfall, and GDP. The initial model incorporates both the whether variables and GDP as follows

$$\begin{aligned} Y_i = \mu(t_i) + \alpha_1^{(T)}(t_i)T(s_1, t_i) + \ldots + \alpha_5^{(T)}(t_i)T(s_5, t_i) + \alpha_1^{(R)}(t_i)R(s_1, t_i) + \ldots + \\ \alpha_5^{(R)}(t_i)R(s_5, t_i) + \alpha^{(GDP)}(t_i)GDP(t_{i-1}) + \epsilon_i, \quad i = 1, \ldots, N, \end{aligned} \tag{1}$$

We assume errors are identically normally distributed with $\mathbb{E}(\epsilon_i) = 0$. In this model, Y_i is the yield observation of the i^{th} year and N is the number of years. The set of temperature variables consists of $T(s_1, t_i)$, the May temperature in i^{th} year, $T(s_2, t_i)$, the June temperature in i^{th} year, and so on. The set of rainfall variables consists of $R(s_1, t_i)$, the May rainfall in i^{th} year, $R(s_2, t_i)$, the June rainfall in i^{th} year, and so on. We also use the one-year lag nominal GDP - $GDP(t_{i-1})$. We estimate the relationship between the set of predictors above and the yield response by allowing for time (year) dependence in the regression coefficients.

There are several difficulties associated with the yield forecast using the weather-based regression model in 1, however. First, it includes a large number of predictors (five for temperature, five for rainfall and one for GDP). Second, there is a within-year dependence among the predictors; temperature and rainfall are observed over a five-month period within each year and used to predict yearly yields. Consequently, the predictors are not uncorrelated as commonly assumed in regression analysis. Predictor dependence or multi-collinearity may not affect the goodness of fit, but the estimated regression coefficients may be unstable due to their joint effect (non-identifiability).

One way to overcome these two model limitations is to use principal component analysis (PCA) to transform possibly correlated variables into a smaller set of uncorrelated variables without a significant loss of information. However, the explanatory variables, temperature and rainfall, depend continuously over time and are therefore naturally described as functionals. In order to transform the temperature and rainfall variables into a set of uncorrelated variables and to allow for within-year dependence, we use the functional version of PCA (FPCA). The key reference for FPCA is Ramsay and Silverman (2002, 2005), but recently, other methods for estimating FPC's have been introduced. For example, Yao, Müller and Wang (2005) developed a method that allows for a sparse design as seen in our application. We apply FPCA to temperature and rainfall data separately.

Denote $\omega_j(s,t)$ the weight functions or principal components for temperature data where s is the month index ($s = 1, \ldots, 5$) and t is the year index ($t = 1, \ldots, N$) for $j = 1, \ldots, 5$. They are functional eigenvectors of the covariance matrix of the temperature data and they form an orthonormal basis in the sense that

$$\int_s \omega_j^2(s,t)\, ds = 1 \text{ and } \int_s \omega_j(s,t)\omega_k(s,t)\, ds = 0. \tag{2}$$

The temperature scores are defined by

$$P_j(t) = \int_s T(s,t)\omega_j(s,t)\, ds,$$

and have mean zero and $\mathbb{E}(P_j(t)^2) = \lambda_j$ where λ_j is the j^{th} eigenvalue of the temperature covariance matrix. Similarly, for rainfall data, we denote $\rho_j(s,t)$ be the principal components and denote $S_i(t)$ the scores defined by

$$S_j(t) = \int_s R(s,t)\rho_j(s,t)\,ds,$$

where $R(s,t)$ is the rainfall in year t varying with s over the year. The rainfall scores also have mean zero and $\mathbb{E}(S_j(t)^2) = \nu_j$ where ν_j is the j^{th} eigenvalue of the rainfall covariance matrix.

The constraints in equation (2) guarantee that the principal components or the weight functions are mutually orthogonal and hence the scores become uncorrelated. Because the first few principal components generally explain most of the variability in the observations, we can reduce the number of variables by discarding the principal components of lesser significance or variability. Let I_T and I_R denote the number of temperature and rainfall principal components selected according to their variability. Since the temperature and rainfall variables are correlated with one another in the additive model (1), we replace the temperature and rainfall predictors with their scores corresponding to the first I_T principal components for temperature data and the first I_R principal components for rainfall data. The model becomes

$$\begin{aligned} Y_i = \mu(t_i) + \alpha_1^{(T)}(t_i)P_1(t_i) + \ldots + \alpha_{I_T}^{(T)}(t_i)P_{I_T}(t_i) + \alpha_1^{(R)}(t_i)S_1(t_i) + \ldots + \\ \alpha_{I_R}^{(R)}(t_i)S_{I_R}(t_i) + \alpha^{(GDP)}(t_i)GDP(t_{i-1}) + \epsilon_i, \quad i = 1, \ldots, N. \end{aligned} \tag{3}$$

This model differs from the classical parametric regression model in the sense that the latter method considers only the relationship of the data within one particular year disregarding between-year relationships. We allow for between-year relationships through the regression coefficients, which are functions of time (year).

3.2 Model Estimation

We estimate the regression coefficient functions μ, $\alpha_j^{(T)}$ for $j = 1, \ldots, I_T$, $\alpha_j^{(R)}$ for $j = 1, \ldots, I_R$, and $\alpha^{(GDP)}$ using p-splines. We closely follow the estimation procedure of a penalized spline using Best Linear Unbiased Prediction (BLUP) in a mixed model outlined by Ruppert et al. (2003).

In 3, we assume the following decomposition of the coefficient functions

$$\begin{aligned}\mu(t) &= \beta_0 \\ \alpha_j(t) &= \beta_j t + \sum_{k=1}^{K} u_{jk} \left| t - \kappa_k \right|^3 .\end{aligned}$$

The decomposition in $\alpha_j(t)$ is based on the radial basis spline functions where κ_k are the fixed knots and K is the number of knots. In the decomposition defined above, β_i are fixed effects and u_{jk} are random effects. We assume that the random effects, u_{jk}, have a normal distribution:

$$\Omega^{1/2} u_{jk} \sim N\left(0, \sigma_u^2 I_K\right), \text{ where } \Omega = \left[|\kappa_k - \kappa_{k'}|^3 \right].$$

Other assumptions in the model are

$$\begin{aligned}\mathbb{E}\left[P_j(t)\right] = 0 \quad &\text{and} \quad \mathbb{V}\left[P_j(t)\right] = \sigma_j^2 \quad \forall j = 1, \dots, I_T, \\ \mathbb{E}\left[S_j(t)\right] = 0 \quad &\text{and} \quad \mathbb{V}\left[S_j(t)\right] = \sigma_j^2 \quad \forall j = 1, \dots, I_R, \\ \mathbb{E}\left[GDP(t)\right] = 0 \quad &\text{and} \quad \mathbb{V}\left[GDP(t)\right] = \sigma^2.\end{aligned}$$

In addition, $P_j(t)$ and $S_j(t)$ are uncorrelated. These assumption hold since $P_j(t)$ and $S_j(t)$ are the scores of functional principal components for temperature and rainfall. GDP used in this model is standardized.

Let N be the number of years and define the X and Z matrices as

$$\begin{aligned}X &= \left[1 \;\; t_i P_1(t_i) \;\; \dots \;\; t_i P_{I_T}(t_i) \;\; t_i S_1(t_i) \;\; \dots \;\; t_i S_{I_R}(t_i) \;\; t_i GDP(t_{i-1})\right]_{i=1,\dots,N}, \\ Z_{T_j} &= \left\{\left[|t_i - \kappa_1|^3 \;\; \dots \;\; |t_i - \kappa_K|^3\right] T_j(t_i)\right\}_{i=1,\dots,N}, \quad j = 1, \dots, I_T, \\ Z_{R_j} &= \left\{\left[|t_i - \kappa_1|^3 \;\; \dots \;\; |t_i - \kappa_K|^3\right] R_j(t_i)\right\}_{i=1,\dots,N}, \quad j = 1, \dots, I_R, \\ Z_{GDP} &= \left\{\left[|t_i - \kappa_1|^3 \;\; \dots \;\; |t_i - \kappa_K|^3\right] GDP(t_i)\right\}_{i=1,\dots,N}.\end{aligned}$$

In order to allow for uncorrelated random effects, we scale Z matrices by $\Omega^{-1/2}$:

$$\begin{aligned}
\widetilde{Z}_{T_j} &= \Omega^{-1/2} Z_{T_j}, \quad j = 1, \ldots, I_T, \\
\widetilde{Z}_{R_j} &= \Omega^{-1/2} Z_{R_j}, \quad j = 1, \ldots, I_R, \\
\widetilde{Z}_{GDP} &= \Omega^{-1/2} Z_{GDP}.
\end{aligned}$$

Finally, define $\widetilde{Z}$ matrix as

$$\widetilde{Z} = \left[\widetilde{Z}_{T_1} \quad \ldots \quad \widetilde{Z}_{T_{I_T}} \quad \widetilde{Z}_{R_1} \quad \ldots \quad \widetilde{Z}_{R_{I_R}} \quad \widetilde{Z}_{GDP} \right].$$

With the above formulation, (3) can be written in the form of a linear mixed model as

$$Y_i = \beta X + u\widetilde{Z} + \varepsilon_i, \quad i = 1, \ldots, N, \tag{4}$$

where β and u are vectors of coefficients.

The approximate 100(1-α)% pointwise confidence band with bias allowance, as proposed by Ruppert et al. (2003), is defined as

$$\widehat{Y}(t^*) \pm z_{\left(1-\frac{\alpha}{2}\right)} \widehat{\sigma}_\varepsilon \sqrt{C_{t^*} \left(C^T C + \frac{\sigma_\varepsilon^2}{\sigma_u^2} D \right) C_{t^*}^T}, \tag{5}$$

where t^* is the predicted year, $C = \left[1 \; X \; \widetilde{Z} \right]$, C_{t^*} is the row of C where $t = t^*$, and $D = diag(0, \ldots, 0, 1, \ldots, 1)$. The number of zeros in D is equal to the number of columns in X plus one (for the intercept) and the number of ones in D is equal to the number of columns in $\widetilde{Z}$. Here, we use a pointwise confidence band rather than a uniform band because we are only interested in the confidence interval of the predicted year, not the past years.

4 Yield Forecasting: Results and Discussion

4.1 Functional Principal Component Analysis

We apply FPCA described in Section 3.1 to temperature and rainfall data from 1927 to 2004. Data in 2005 will be used as testing data to evaluate the performance of the yield forecasting model.

The smoothed mean functions of the weather data are shown in Figure 1. Figure 1(a) shows the mean function of the temperature data. Notice the average temperature gradually increases from May and reaches the highest average temperature at 76 F in July, indicated by the dashed line, and then steadily decreases until the end of the harvesting period in September. Figure 1(b) displays the mean function of the rain data. The total rainfall is highest in the beginning of June, indicated by the left dashed line, and decreases rapidly until it reaches its minimum in August, indicated by the right dashed line.

The principal component weight functions are estimated as discussed in Section 3.1. Figure 2 depicts the first four principal component curves of the temperature data. Each panel shows the weight function for the temperature data, which are deviations from the overall mean in Figure 1(a) and the variability explained by each principal component. The first weight function, displayed in the upper left panel, is negative throughout the year. The highest absolute weight is in July, which is the month that has the highest temperature. The lowest absolute weight is in September, which is about a half of the highest absolute weight. This implies that the greatest variability between years can be found by heavily weighting May to August and lightly weighting September. The second weight function is displayed in the upper right panel. This function has a sinusoidal shape. The weight gradually increases from May until it reaches its peak in June. This component consists of a positive contribution for temperature before July and a negative contribution for temperature after July.

The first four principal component functions for the rainfall data are displayed in Figure 3. The first weight function is shown in the upper left panel. The highest absolute weight is placed on July and moderate absolute weight on June. Small positive weights are assigned to August and September and a small negative weight is assigned to May. In contrast, the second weight function places positive weights on July, August, and September and negative weights on May and June.

4.2 Additive Regression Model

The scores of temperature and rainfall principal components are used as predictors in the semiparametric regression model as described in (3). First, we incorporate the scores of all principal components along with the standardized lag nominal GDP. The full model is defined in (3) for $I_T = 5$ and $I_R = 5$, which we refer to as Model 1. The prediction results for 1996 to 2005 are illustrated in Figure 4. This model provides good predictions for only a few years with large prediction errors for the beginning and the end of the prediction period.

According to the amount of variation explained by each principal component, the first two temperature principal components and the first three rainfall principal components explain most of the variation in the data. This suggests using only these five principal components and the standardized lag nominal GDP to predict the yearly yield. The model is described in equation (3) for $I_T = 2$ and $I_R = 3$, which we refer to Model 2. Figure 5 depicts the forecasting results for 1996 to 2005. This model delivers a better forecast than the previous model even though it has large prediction errors in years 1999 and 2005. This implies that discarding high order principal components improves the performance of the model.

The output of the semiparametric regression model (not shown here) indicates that some coefficient functions of weather principal components are approximately linear. This suggests using linear functions to estimate some of the regression coefficients. Using linear rather than nonlinear coefficient functions entails a more parsimonious model, which will be easier to predict and interpret. Next, we search exhaustively starting with the full model (Model 1) to identify a set of predictors and the shape of their coefficient functions (linear vs. nonlinear) that will provide the best overall prediction and fitting. This gives the final model (Model 3). This model consists of two nonparametric nonlinear components, standardized lag nominal GDP and first temperature principal component, three linear components, second and third temperature principal components, and first rainfall principal component. The final model is given by

$$Y_i = \mu(t_i) + \alpha_1^{(T)}(t_i)P_1(t_i) + \alpha_2^{(T)}P_2(t_i) + \alpha_3^{(T)}P_3(t_i)+$$

$$\alpha_1^{(R)}S_1(t_i) + \alpha^{(GDP)}(t_i)GDP(t_{i-1}) + \epsilon_i, \quad i = 1, \ldots, N. \tag{6}$$

The predicted yields from 1996 to 2005 are shown in Figure 6. The predicted yields are very close to the observed yields. There is a moderate error in year 2005.

4.3 Linear Regression Analysis

The common approach to predicting yield from weather data is linear regression (Kandiannan et al. 2002, Sheehy et al. 2006, Seif and Pederson 1978) as provided by

$$Y_i = \mu + \alpha_1^{(T)}T(s_1, t_i) + \ldots + \alpha_5^{(T)}T(s_5, t_i) + \alpha_1^{(R)}R(s_1, t_i) + \ldots +$$

$$\alpha_5^{(R)}R(s_5, t_i) + \alpha^{(GDP)}GDP(t_{i-1}) + \epsilon_i, \quad i = 1, \ldots, N. \tag{7}$$

which we refer to as Model 4. The difference between (1) and Model 4 is that the coefficients are fixed over time. We can also replace the monthly temperature and rain predictors with the scores corresponding to their functional principal components and obtain

$$Y_i = \mu + \alpha_1^{(T)}P_1(t_i) + \ldots + \alpha_5^{(T)}P_5(t_i) + \alpha_1^{(R)}S_1(t_i) + \ldots +$$

$$\alpha_5^{(R)}S_5(t_i) + \alpha^{(GDP)}GDP(t_{i-1}) + \epsilon_i, \quad i = 1, \ldots, N. \tag{8}$$

This is referred to as Model 5. The prediction results of Models 4 and 5 are provided in Figures 7(a) and 7(b), respectively. These two linear regression models provide a good prediction with moderate error in years 1999 and 2005 and small errors in the first three years. It is not surprising that both models give similar prediction results since principal components are linear combinations of the original data.

4.4 Model Evaluation

We use the mean square error (MSE) criterion to evaluate the performance of the forecasting models. Table 1 summarizes the performance of these models for corn yield prediction as provided by MSE. Both parametric regression models, Models 4 and 5, have similar MSE. The semiparametric regression Models 1 and 2, on the other hand, have much higher MSE than the linear regression Models 4 and 5. The selected model, Model 3, has the smallest MSE. It is about one-half of MSE for Models 4 and 5, and one-fourth of MSE for semiparametric Models 1 and 2. Therefore, we can conclude that allowing for time-dependent relationships reflected in the nonlinear coefficient functions together with the functional principal component analysis improve the performance of the forecasting model. Note that the observed corn yield in year 2005 is lower than the prediction in all five models. This is because there were extreme drought conditions during the 2005 growing season (Zhang et al. 2006). The proposed models can capture part of this drought effect. This can be seen from lower forecasted yield in 2005 than one in 2004.

We obtain similar results when we apply these regression models to soybean yield. The selected model for soybean yield data provides the smallest MSE. This model includes the lag nominal GDP and the second rainfall principal component with coefficients estimated as functions of time. The linear components are the first, second, and fifth temperature principal components and the first rainfall principal component. Also, the parametric models have lower adjusted coefficient of determination than all semiparametric models. The soybean yield forecasting results and the performance of the models are summarized in Table 2.

4.5 Confidence Band

Since Model 3 provides the best prediction, we employ it as the base model to determine the pointwise confidence band defined in Section 3.2. Predicted corn yield for year 2005, as shown in Table 1, is 170.28 bushels per acre. The 95% pointwise confidence band for year 2005 in bushels per acre is (143.80, 196.75). The observed corn yield in year 2005 is 142 bushels per acre. It is slightly less than the lower bound of the estimated confidence interval due to the severe drought conditions mentioned in Section 4.4. The pointwise confidence band of Model 3 is illustrated in Figure 8 for both corn and soybean. The confidence band will give a range of possible outcomes, and will allow us to cope with different scenarios that may occur over the planning horizon for crop decisions.

5 Price Forecasting: Method

5.1 Model Formulation

In this section, we develop a cash price forecasting model under the futures-based framework where cash price is forecasted from futures price and basis. We forecast the cash price by first obtaining a forecast of the basis over one year and then adding the expected futures price to it. We divide the data into 16 different functional observations, each observation consisting of basis values observed over a one-year period. Therefore, our data are both functional and longitudinal

$$Y_j(t_i), i = 1, \ldots, n_j, \quad j = 1, \ldots, n = 16.$$

Our goal is to identify common patterns among the 16 years and use them to predict the basis of the upcoming year. We adopt a model-based approach to estimate the density function of the basis distribution. Model-based clustering, introduced by Banfield and Raftery (1993) relies on estimation of a mixture density function, each component in the mixture corresponding to one cluster. Within this method, the main assumption is that the observations $y_1, \ldots, y_n$ are random variables from a mixture distribution with K components. However, in Banfield and Raftery (1993), Celeux and Govaert (1995), and Dasgupta and Raftery (1998), the model framework does not allow for functional relationships of the data. Since our basis data are functional, we instead exploit a functional data model-based clustering framework. Consequently, we follow the approach for clustering functional data proposed by James and Sugar (2003).

We assume that the predicted basis curve belongs to a cluster or model component with some probability as estimated using the mixture likelihood approach. Under this approach, the cluster memberships Z_j's are treated as missing data assuming that Z_j for $j = 1, \ldots, n$ are multinomial with parameters $(\pi_1, \ldots, \pi_K)$ and π_k is the probability that a basis curve belongs to the k^{th} cluster. Let $f_k(y_j|\theta_k)$ be the density function corresponding to the k^{th} cluster, parameterized by θ_k. The parameters are estimated by maximizing

$$L(\theta_1, \ldots, \theta_K; \pi_1, \ldots, \pi_K | y_1, \ldots, y_n) = \prod_{j=1}^{n} \sum_{k=1}^{K} \pi_k f_k(y_j|\theta_k). \tag{9}$$

Let b_{ij}, β_{ij}, and ε_{ij} be, respectively, the observed basis value, true basis value, and measurement error in year j and time t_i, i.e. $b_{ij} = b_j(t_i)$ and $\beta_{ij} = \beta_j(t_i)$. The basis model can be formulated as

$$b_{ij} = \beta_{ij} + \varepsilon_{ij}, \quad t_i = t_1, \ldots, t_{n_j}, \quad j = 1, \ldots, n,$$

where n is the number of years and n_j is the number of time points in year j. This model assumes β_{ij} follows a Gaussian process, and the measurement errors have mean zero and are uncorrelated with each other and true basis values. For each year j, we expand the true function $\beta_j(t)$ using a set of spline basis functions, and for each group k, we compute the mean, $\mu_k(t)$, along with the cluster proportion parameter π_k as extensively discussed in James and Sugar (2003) and summarized in the next section.

Under the mixture likelihood framework

$$y(t) \sim \sum_{k=1}^{K} \pi_k f(\mu_k(t), \Sigma_k(t)), \tag{10}$$

where K is the number of clusters, f is the density function of the cluster, and Σ_k is the covariance matrix of the k^{th} cluster.

We apply the functional model-based clustering to the standardized basis data to forecast the basis distribution on a common scale for all years. We calibrate the scaled basis distribution to forecast yearly basis by transforming back to its original scale. Finally, the predicted cash price over one year is calculated by adding the estimated basis with deferred futures price. Deferred futures price is the current price of the futures contract that expires in the distant future, beyond the expiration of the nearby contract and it is closest to, but not before, the time of the forecasted cash price.

5.2 Model Estimation

We expand the true basis value by a set of spline basis functions $\beta_j(t) = s(t)^T \varphi_j$, where $s(t)$ is a vector of spline basis and φ is a spline coefficient vector. The spline coefficients are modeled by assuming a Gaussian distribution:

$$\varphi_j = \mu_{z_j} + \gamma_j, \quad \gamma_j \sim N(0, \Gamma),$$

where μ_{z_j} is the cluster mean or the cluster fixed effect, z_j is the unknown cluster membership of year j, and γ_j is a random effect of year j. Define $S_j = \big(s(t_1), \ldots, s(t_{n_j})\big)^T$ to be the spline basis matrix corresponding to the j^{th} year, b_j to be the vector of the observed values, and ε_j to be the vector of measurement errors. The functional clustering model (FCM) can be written as

$$b_j = S_j(\mu_{z_j} + \gamma_j) + \varepsilon_j, \quad j = 1, \ldots, n, \tag{11}$$

$$\varepsilon_j \sim N(0, \sigma^2 I), \quad \gamma_j \sim N(0, \Gamma).$$

In this model, we assume the covariances of the ε_j's and γ_j's to be, respectively, $\sigma^2 I$ and Γ for all clusters. Therefore, under this formulation, the distribution of β_j is

$$b_j \sim N(S_j \mu_{z_j}, \Sigma_j), \text{ where } \Sigma_j = \sigma^2 I + S_j \Gamma S_j^T. \tag{12}$$

As in FCM, we estimate the parameters by maximizing the mixture likelihood function (9). The estimation procedure is fully described in James and Sugar (2003).

Under the Gaussian process assumption, we assume that our predicted curve follows a mixture of normals whose k^{th} component has mean μ_k and variance Σ_k. We can further derive the $100(1-\alpha)\%$ pointwise confidence band of the functional mixture model by

$$\widehat{b}(\tau, t) = \sum_{k=1}^{K} \widehat{\pi}_k \left(\widehat{\mu}_k(t) \pm \Phi^{-1}(1 - \frac{\alpha}{2}) \widehat{\upsilon}_k(t) \right), \tag{13}$$

where Φ^{-1} is the inverse cumulative density function and υ_k is the diagonal of the covariance matrix of the k^{th} cluster.

6 Price Forecasting: Results and Discussion

6.1 Basis Information

We constructed the basis history from futures and cash price data by subtracting nearby futures price from the corresponding cash price. The corn basis plots from years 1991 to 2006 are shown in Figure 9. Note that the basis plot in 2006 is incomplete because we only have the data up to February 22. Overall, the basis fluctuates within a year with five common local maxima and one local minimum. In these plots, the local maxima are indicated by dashed lines and the local minimum is marked by a straight line. However, the 1996 corn basis has an outlying pattern since it behaves differently from other years. This outlier may come from the passage of the 1996 Farm Act, which increased the planting flexibility and resulted in a large amount of corn released to the market. Similar patterns occur in the soybean basis data (not shown here) but they are not as consistent as for the corn basis.

6.2 Functional Clustering Analysis

The functional model-based clustering technique outlined in Section 5.1 is applied to the basis data. Similarly to yield forecasting, the basis is forecasted yearly. We show the corn basis forecasts for year 2005, using data from years 1991 to 2004. As mentioned in Section 6.1, basis data in year 1996 is excluded from this analysis since it is an outlier for our 16-year period.

We first perform the clustering analysis on the standardized basis data. For the method briefly discussed in Section 5.2 and developed by James and Sugar, we need to specify the dimension of the spline basis, p, and the number of clusters, K. The clustering membership using data from years 1991 to 2004 (except 1996) is shown in Table 3 for $K = 2, 3, 4, 5$. Under the mixture likelihood framework (10), the predicted basis curves for each K are almost identical as depicted in Figure 10. Since our primary goal is pattern prediction and not the clustering membership, we can choose any of these value for K. There are techniques that can be used to determine the number of clusters, for example, Bayes factors (Kass and Raftery 1995), gap statistic (Tibshirani et al. 2001), and jump method (Sugar and James 2003) but here we choose a low value for K.

The tuning parameter that has a significant effect on the prediction is the dimension of the spline basis or the number of knots, p. Figure 11 shows the prediction outcome when p equals 10, 15, 20, and 25. In our further analysis, we choose $p = 20$ because we believe that at this smoothing level, we best balance the local and global variations.

6.3 Confidence Band

The 95% confidence band of the forecasted basis, as defined in Section 5.2, is computed for $\alpha = 0.05$. The result is illustrated in Figure 12(a). Even though we have daily values available for basis historical data, it is more convenient to use weekly average values rather than daily values. This is because of the lack of synchronization between the dates across years. For example, January 2 was Friday in 2004 but Sunday in 2005. Moreover, the futures markets close on Saturday and Sunday so we do not have available data for weekend days. Consequently, the forecasted daily basis is averaged to a weekly value using forecasted data only from Monday to Friday. The confidence band in average weekly basis is shown in Figure 12(b). We can see that there is not much loss of information from aggregating daily basis values to weekly ones.

6.4 Calibration

Since the basis data is first standardized, the forecasted result will be on the standardized scale. The basis on this scale provides information about the predicted pattern but not about the predicted values. In this section, we propose a simple calibration method.

As provided in Figure 9, the basis patterns are on similar scales as the scale of their adjacent year. Therefore, we use the basis information from 2004 to calibrate the 2005 forecasted basis. In addition, since we want to forecast the basis price starting at the time of planting, which is May for corn and soybean in Illinois. We calibrate using the first four months of the year and the last eight months of the previous year to have a full year of reference basis data. In our calibration method, we first adjust for the mean difference between the current and previous year. Therefore,we determine the difference between the means of the first four months of the observed basis curves in 2004 and 2005. Then we subtract the difference from the basis data in 2004 to shift the 2004 basis curve to the same level as one in 2005. Second, we re-scale the predicted pattern to the scale corresponding to the predicted year. We call this calibration with difference adjusted. The result is illustrated in Figure 13. Most of the observed basis values are captured by the band. However, there is still a large difference during weeks 35 to 45. This may come from the severe drought condition in 2005, which affected the corn production and hence the price.

The proposed calibration technique works well in the years with no severe conditions. Figure 14 shows the calibrated band for corn basis in 2004. Most of the observed basis are contained within the band and there are only a few points that lie far from the band. We also apply this calibration method to soybean. The calibrated confidence band capture most of the observed soybean basis, as depicted in Figure 15.

6.5 Forecasted Cash Price

In crop decision planning, we need to incorporate cash price but not the basis. As mentioned in Section 5.1, cash price is calculated by adding the forecasted basis and the expected futures price. For example, suppose we want to sell corn right after harvest (late September) and the December futures contract (a deferred contract for September) is traded at $2.04 per bushel. From the 2005 basis forecast, the corn basis in last week of September (week 39) is -$0.264 per bushel. Then the expected cash price is $2.04 + (-$0.264) = $1.776 and the corresponding confidence interval is ($1.7314, $1.8207). It is important to remember that the accuracy of the forecasted cash price depends not only on the basis forecast but also on the futures price. The futures contract price changes continuously; even for the same delivery month the prices may not be the same if observed at different times. In the previous example, suppose that the day before our calculation the December futures contract was traded at $2.0325 per bushel. The expected cash price would change to $2.0325 + (-$0.264) = $1.7685.

7 Discussion

We first provide discussion of our methods and results and then conclude with a section on the motivation and background of how the forecasts would be used in a decision planning model.

7.1 Yield Forecasting

To the best of our knowledge, parametric linear regression has been used as the forecasting tool in almost all studies on yield forecasting that use a regression approach. Linear regression has been mainly considered since it is easy to use and interpret. However, as pointed out in Section 3, it does not incorporate the between-year relationships in the weather data. Disregarding these relationships may result in inaccurate prediction as already discussed in Section 4.4. Subsequently, we proposed a semiparametric regression model, which allows for these relationships. One important contribution is estimating confidence bands for the yield forecast, which can be incorporated in our decision planning model.

There are several observations arising from this forecasting analysis. Temperature gives more information than rainfall, especially in the corn yield prediction. In the final corn yield prediction model, we use three temperature principal components and only one rainfall principal component. In soybean yield forecasting, we use three temperature principal components and only two rainfall principal components.

7.2 Price Forecasting

This study explores the use of functional model-based analysis in the crop's price forecasting. Under the futures-based model where expected cash price is equal to futures price plus expected basis, we focus on basis estimation. A multiple-year average technique is often used as a tool to compute the expected basis. It is simple and provides relatively insightful results. However, it returns only the expectation of the basis. Using only the expectation in decision making may lead to inaccurate decisions. On the other hand, our model formulation allows us to estimate (pointwise) confidence bands since it provides the density function of the basis distribution.

For price forecasting, we have only few years of data. If we were to have available data for a longer period of time, we might be able to cluster years with severe conditions like flood and drought. This would allow us to accurately identify 2005 as a drought year and correctly account for this severe condition.

7.3 Crop Decision Planning

Decision planning plays an important role in agriculture just as it does in other industries. It is a key factor that determines in part the success and failure of business. The decisions that farmers have to make include: i) which crops to produce, ii) how much land to allocate to each crop, and iii) when to grow, harvest, and sell. Uncertain factors such as weather and demand, along with the limited resources used to cultivate, store, and supply crops make crop decisions difficult for farmers and can therefore significantly affect returns.

In order to make good decisions under these uncertainties, forecasting of important factors is a crucial step. The models described in previous sections can help establish the expected values of yield and price along with their confidence bands. These results enable analysis to be performed on different scenarios. For example, before each cropping period begins, farmers have to consider which crops they will grow. This can be a difficult decision since farmers do not have information for the outcome of their crops before cropping begins. The expectations for yield and price would help producers to estimate the expected return for each crop. The confidence bands would provide possible outcomes beside the expected values.

A second example is the decision at the harvest time, when the yields and cash prices at are realized. If crops are storable, farmers may keep them for sale at higher prices after the harvest time. By doing so, producers may obtain higher return. However, this decision will depend on whether the increased price is high enough to compensate for the storage costs. The price confidence band would help the growers to estimate the maximum loss or maximum gain. With these estimations along with the realized harvest cash price, farmers can make a better decision about whether they should keep their crops for later sale or not.

We are currently working on an optimization-based decision planning model for these decision problems. Since yield and price are stochastic and affect both revenues and costs, these stochastic variables will be incorporated in the model to derive probabilities for different planning scenarios. Our model will take into account the resource limitations as its major constraints as well as the limitation on the planting and harvesting periods of each crop.

References

[1] Banfield, J.D., Raftery, A.E. (1993), "Model-based Gaussian and non-gaussian clustering", *Biometrics*, 49, 803-821.

[2] Bannayan M., Crout, N.M.J (1999), "A stochastic modelling approach for real-time forecasting of winter wheat yield", *Field Crops Research*, 62, 85-95.

[3] Batts, G.R., Morison, J.I.L., Ellis, R.H., Hadley, P., Wheeler, T.R. (1997), "Effects of CO_2 and temperature on growth and yield of crops of winter wheat over four seasons", *European Journal of Agronomy*, 7, 43-52.

[4] Bureau of Economic Analysis, United State Department of Commerce.

[5] Bureau of Labor Statistics, United State Department of Labor (2005), "Career guide to industries 2006-2007 editon".

[6] Celeux, G., Govaert, G. (1995), "Gaussian parsimonious clustering models", *The Journal of the Pattern Recognition Society*, 28, 781-793.

[7] Chicago Board of Trade (2000), "Understanding basis".

[8] Dasgupta, A., Raftery, A.E. (1998), "Detecting features in spatial point porcesses with clutter via model-based clustering", *Journal of the American Statistical Association*, 93, 294-302.

[9] Dow, J.C.R. (1940), "A theoretical account of futures markets", *The Review of Economic Studies*, 7, 185-195.

[10] Eales, J.S., Engel, B.K., Hauser, R.J., Thompson, S.R. (1990), "Grain price expectations of Illinois farmers and grain merchandisers", *American Journal of Agricultural Economics*, 72, 701-708.

[11] Gardner, B.L. (1976), "Futures prices in supply analysis", *American Journal of Agricultural Economics*, 58, 81-84.

[12] Hoogenboom, G. (2000), "Contribution of agrometeorology to the simulation of crop production and its applications", *Agricultural and Forest Meteorology*, 103, 137-157.

[13] James, G.M., Sugar, C.A. (2003), "Clustering for sparsely sampled functional data", *Journal of the American Statistical Association*, 98, 397-408.

[14] Kandiannan, K., Chandaragiri, K.K., Sankaran, N., Balasubramanian, T.N., Kailasam, C. (2002), "Crop-weather model for turmeric yield forecasting for Coimbatore District, Tamil Nadu, India", *Agricultural and Forest Meteorology*, 112, 133-137.

[15] Kass, R.E., Raftery, A.E. (1995), "Bayes factors", *Journal of the American Statistical Association*, 90, 773-795.

[16] Kenyon, D., Jones, E., McGuirk, A. (1993), "Forecasting performance of corn and soybean harvest futures contracts", *American Journal of Agricultural Economics*, 75, 399-407.

[17] Krog, D.R. (1988), Plant-process model corn yield forecasts for Iowa, Ph.D. Dissertation, Iowa State University, Iowa.

[18] Lee, R. (1999), Modeling corn yields in Iowa using time series analysis of AVHRR data and vegetation phenological metrics, Ph.D. Dissertation, University of Kansas, Kansas.

[19] Mitchell, R.A.C., Mitchell, V.J., Driscoll, S.P., Franklin, J., Lawlor, D.W. (1993), "Effects of increased CO_2 concentration and temperature on growth and yield of winter wheat at two levels of nitrogen application", *Plant, Cell and Environment*, 16, 521-529.

[20] Mkhabela, M.S., Mkhabela, M.S., Mashinini, N.N. (2005), "Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR", *Agricultural and Forest Meteorology*, 129, 1-9.

[21] National Agricultural Statistics Service, United State Department of Agriculture (2005), "Agricultural statistics 2005".

[22] National Agricultural Statistics Service, United State Department of Agriculture (1997), "Usual planting and harvesting dates for U.S. field crops".

[23] National Climatic Data Center, United State Department of Commerce.

[24] Peng, S., Huang, J., Sheehy, J.E., Laza, R.C., Visperas, R.M., Zhong, X., Centeno, G.S., Khush, G.S., Cassman, K.G. (2004), "Rice yields decline with higher night temperature from global warming", *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9971-9975.

[25] Potgieter, A.B., Hammer, G.L., Doherty, A., de Voil, P. (2005), "A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia", *Agricultural and Forest Meteorology*, 132, 143-153.

[26] Ramsay, J.O., Silverman, B.W. (2002), *Applied Functional Data Analysis*, Springer, New York.

[27] Ramsay, J.O., Silverman, B.W. (2005), *Functional Data Analysis*, Springer, New York.

[28] Ruppert, D., Wand, M.P., Carroll, R.J. (2003), *Semiparametric Regression*, Cambridge, New York.

[29] Seif, E., Pederson, D.G. (1978), "Effect of rainfall on the grain yield of spring wheat, with an application to the analysis of adaptation", *Australian Journal of Agricultural Research*, 29, 1107-1115.

[30] Sheehy, J.E., Mitchell, P.L., Ferrer, A.B. (2006), "Decline in rice grain yields with temperature: Models and correlations can give different estimates", *Field Crops Research*, 98, 151-156.

[31] Sugar, C.A., James, G.M. (2003), "Finding the number of clusters in a data set: An information theoretic approach", *Journal of the American Statistical Association*, 98, 750-763.

[32] Tibshirani, R., Walther, G., Hastie, T. (2001), "Estimating the number of clusters in a data set via the gap statistic", *Journal of the Royal Statistical Society, B*, 63, 411-423.

[33] Tomek, W.G., Gray, R.W. (1970), "Temporal relationships among prices on commodity futures markets: Their allocative and stabilizing roles", *American Journal of Agricultural Economics*, 52, 372-380.

[34] Wheeler, T.R., Craufurd, P.Q., Ellis, R.H., Porter, J.R., Prasad P.V.V. (2000), "Temperature variability and the yield of annual crops", *Agriculture, Ecosystems & Environment*, 82, 159-167.

[35] World Agricultural Outlook Board, United State Department of Agriculture, "World agricultural supply and demand estimates".

[36] Working, H. (1942), "Quotations on commodity futures as price forecasts", *Econometrica*, 10, 39-52.

[37] Yao, F., Müller, H.G., Wang, J.L. (2005), "Functional data analysis for sparse longitudinal data", *Journal of the American Statistical Association*, 100, 577-590.

[38] Zhang, P., Anderson, B.T., Myneni, R. (2006), "Monitoring 2005 corn belt yields from space", *EOS, Transactions American Geophysical Union*, 87, 150.

Predicted Year	Observed Corn Yield	Model 1	Model 2	**Model 3**	Model 4	Model 5
1996	141	158.13	145.27	142.91	153.38	158.27
1997	140	136.42	131.16	137.50	148.09	148.16
1998	133	154.75	128.98	131.86	146.92	140.63
1999	133	145.21	173.43	149.40	162.27	162.35
2000	158	144.87	140.84	150.86	160.45	161.24
2001	155	158.64	149.07	152.89	158.35	158.45
2002	155	170.39	161.90	159.85	152.86	153.85
2003	171	154.74	166.31	167.47	168.98	169.35
2004	193	233.66	195.96	192.99	197.17	196.58
2005	142	185.92	187.68	170.49	168.26	168.45
MSE		519.75	424.15	**118.34**	200.72	202.33
R *		90.96%	89.52%	89.21%	79.50%	79.50%
R-adj *		89.45%	**88.63%**	88.47%	76.10%	76.10%

Table 1: Corn yield prediction results for Model 1 to Model 5. (* Using Data 1927-2004)

Predicted Year	Observed Soybean Yield	Model 1	Model 2	**Model 3**	Model 4	Model 5
1996	43.5	45.56	43.21	44.24	43.61	44.16
1997	47.5	47.07	44.41	46.21	46.77	46.80
1998	44	41.16	39.36	40.81	44.64	44.05
1999	39	49.56	51.35	50.65	50.92	50.92
2000	46	44.73	42.12	43.07	48.51	48.60
2001	46	47.07	47.89	48.33	50.11	50.07
2002	47	52.14	50.56	48.59	51.54	51.70
2003	43	44.57	44.75	46.57	50.17	50.26
2004	53	63.03	51.57	52.69	56.64	56.50
2005	47	42.84	42.53	46.68	49.29	49.40
MSE		27.35	24.00	**17.75**	25.66	25.92
R *		87.81%	86.06%	86.87%	78.60%	78.60%
R-adj *		85.77%	84.88%	**85.96%**	75.00%	75.00%

Table 2: Soybean yield prediction results for Model 1 to Model 5. (* Using Data 1927-2004)

Number of Clusters	1991	1992	1993	1994	1995	1997	1998	1999	2000	2001	2002	2003	2004
K=2	2	2	1	2	1	2	1	1	1	1	1	2	2
K=3	2	2	1	2	1	2	1	1	1	3	3	2	2
K=4	2	4	2	2	1	4	1	1	1	3	3	2	2
K=5	5	4	2	2	5	4	1	1	1	3	3	2	2

Table 3: Cluster membership for corn basis data from 1991 to 2004 (except 1996).

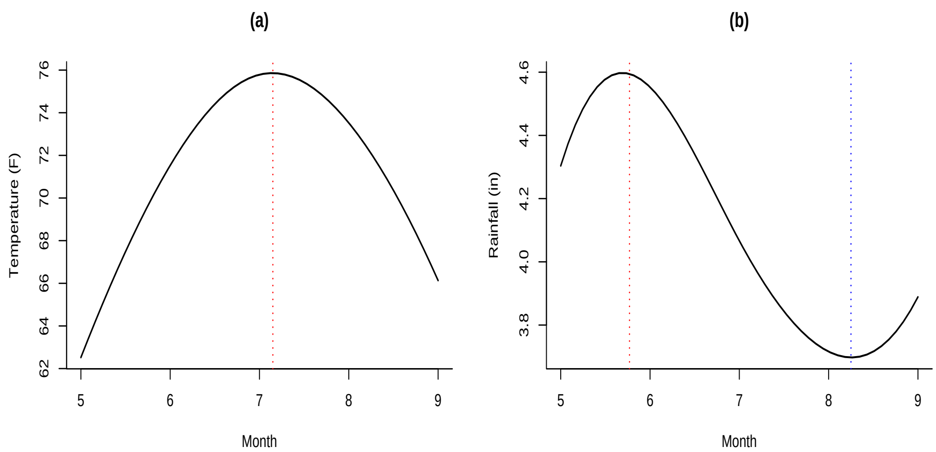

Figure 1: Smooth mean functions of (a) temperature data and (b) rainfall data from 1927 to 2004.

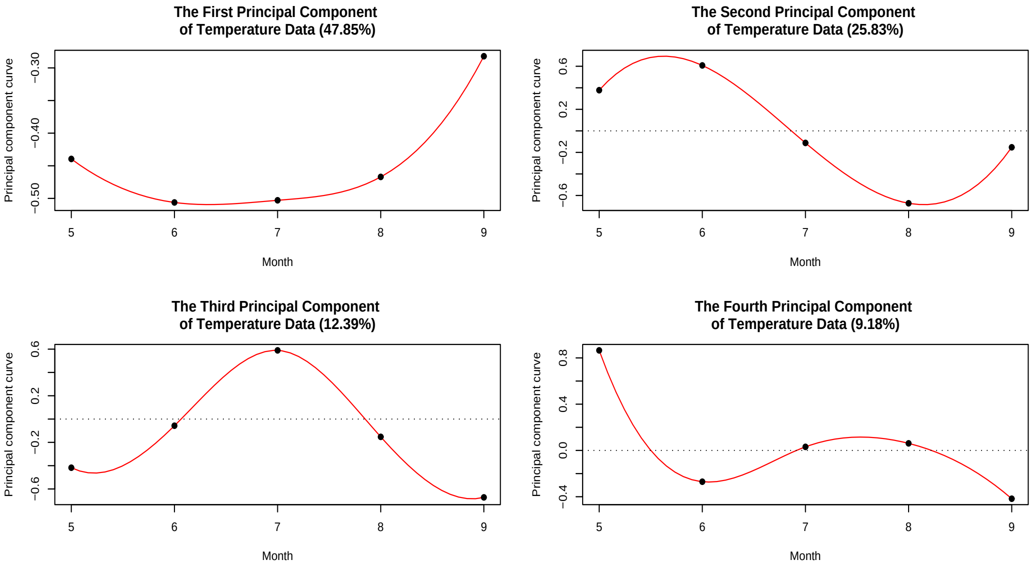

Figure 2: The first four principal component curves of temperature data from 1927 to 2004.

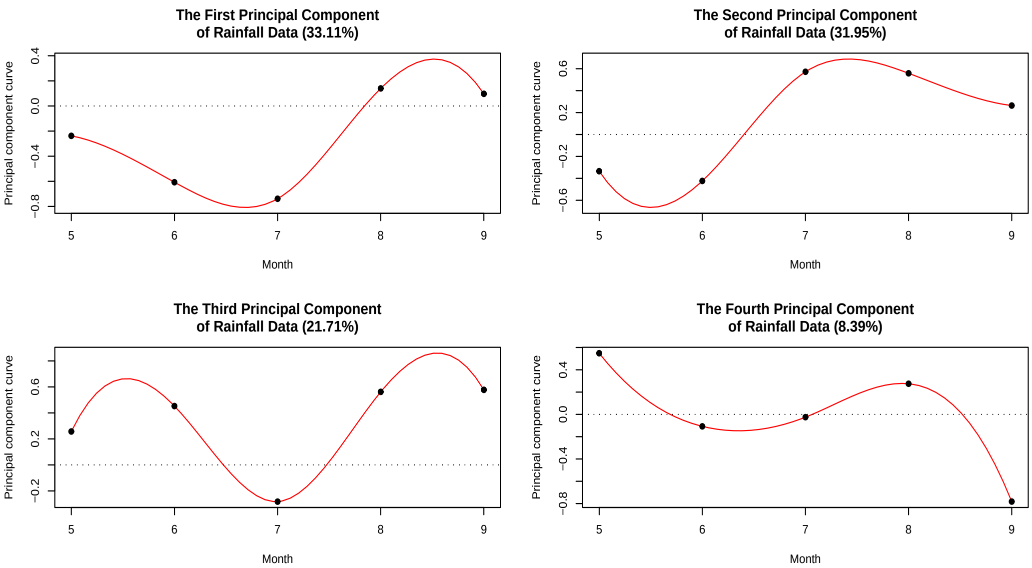

Figure 3: The first four principal component curves of rainfall data from 1927 to 2004.

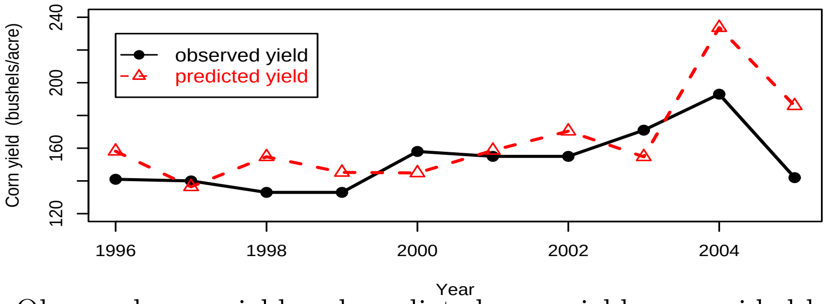

Figure 4: Observed corn yield and predicted corn yield as provided by Model 1.

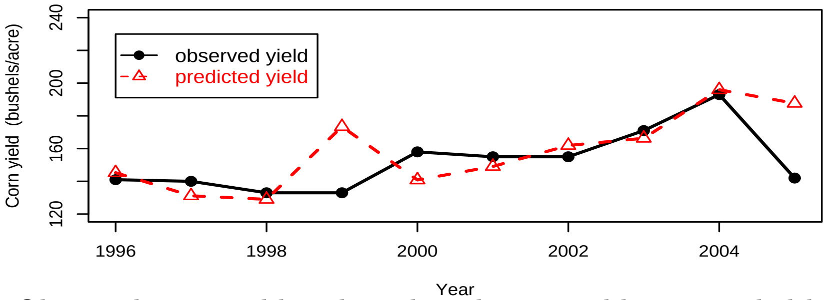

Figure 5: Observed corn yield and predicted corn yield as provided by Model 2.

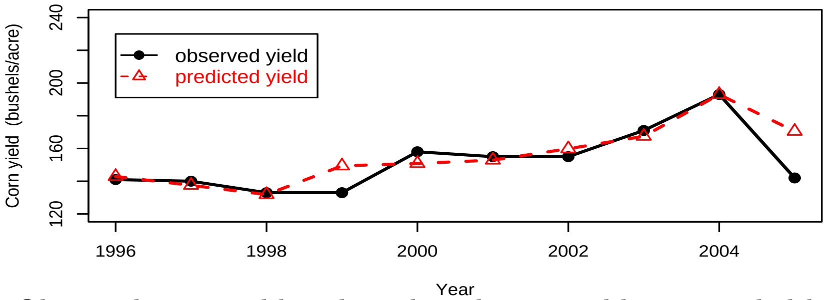

Figure 6: Observed corn yield and predicted corn yield as provided by Model 3.

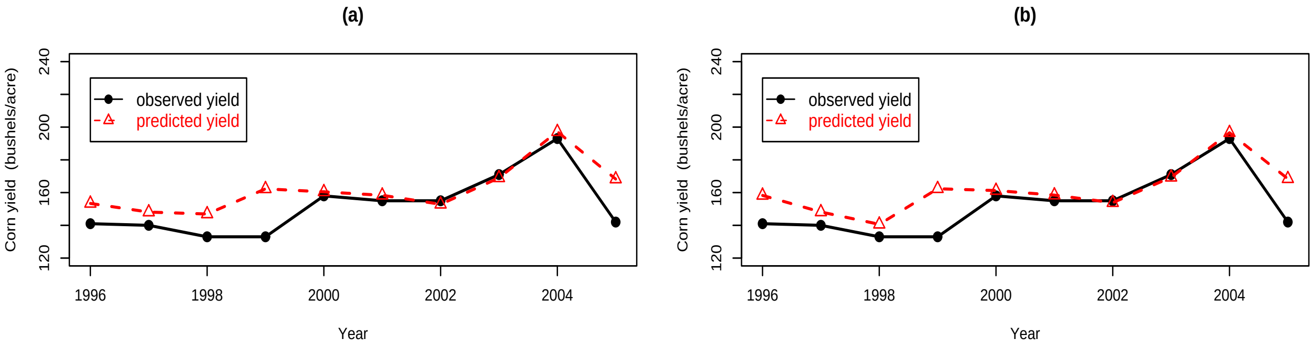

Figure 7: Observed corn yield and predicted corn yield as provided by Model 4 in (a) and Model 5 in (b).

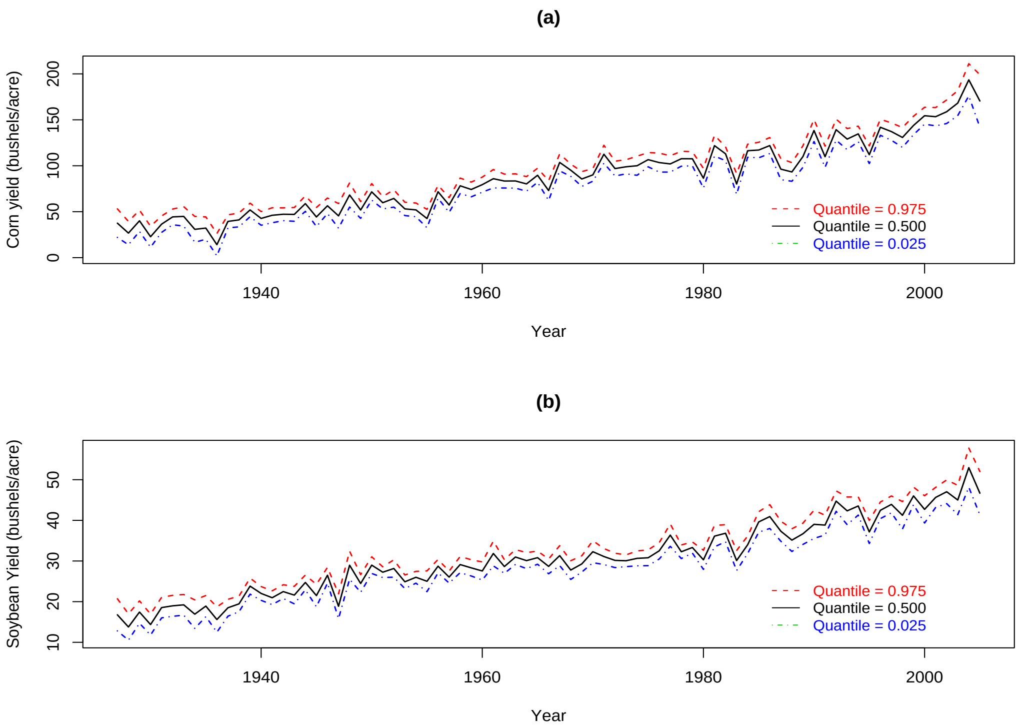

Figure 8: Plot of 95% pointwise confidence band of the fitted corn (a) and soybean (b) yields from 1927 to 2005.

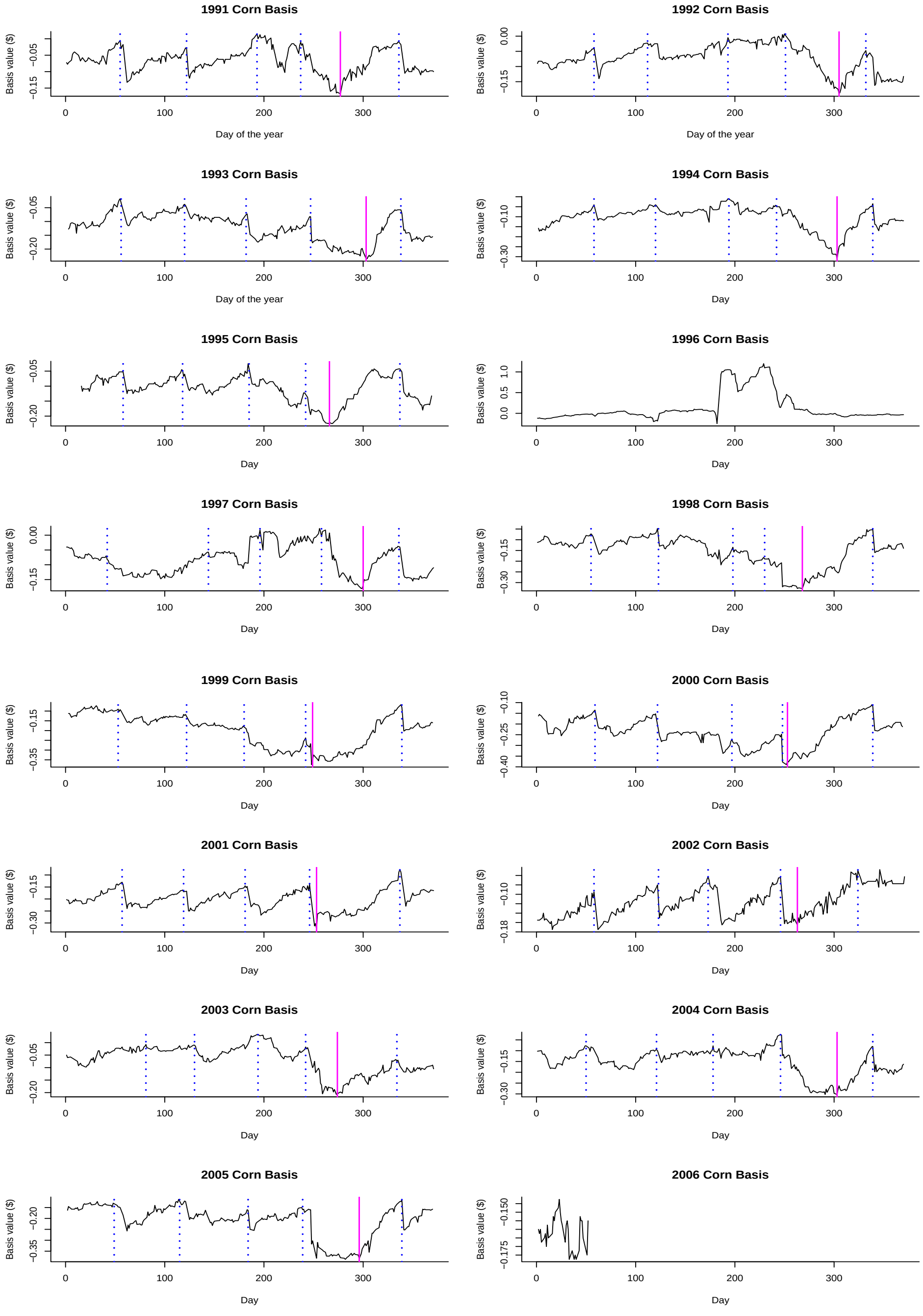

Figure 9: Corn basis plots from 1991 to 2006.

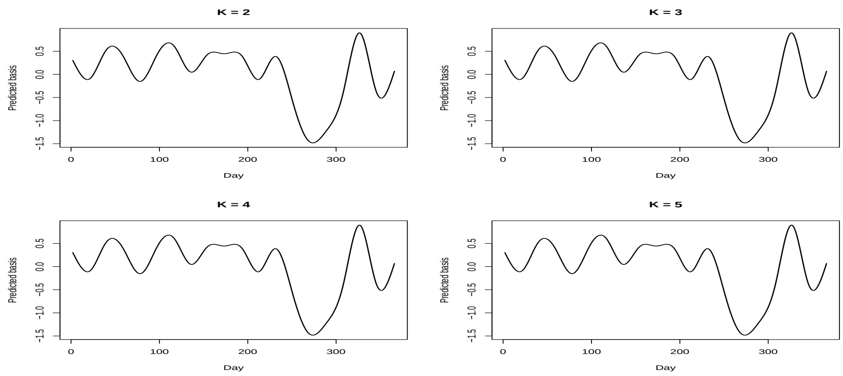

Figure 10: Comparison of 2005 corn basis forecast when K = (2,3,4,5).

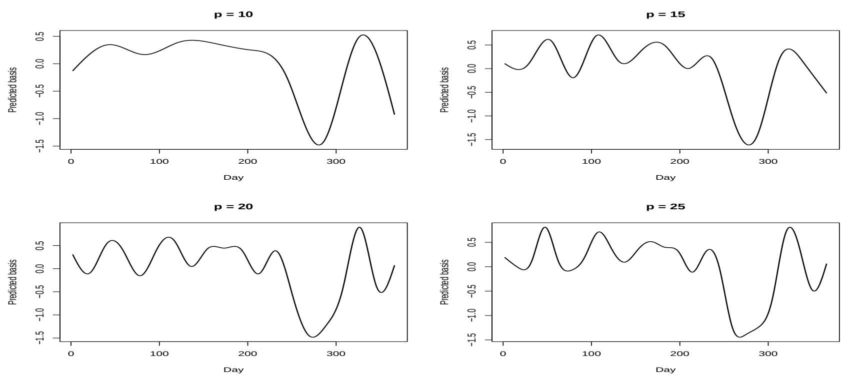

Figure 11: Comparison of 2005 corn basis forecast when p = (10,15,20,25).

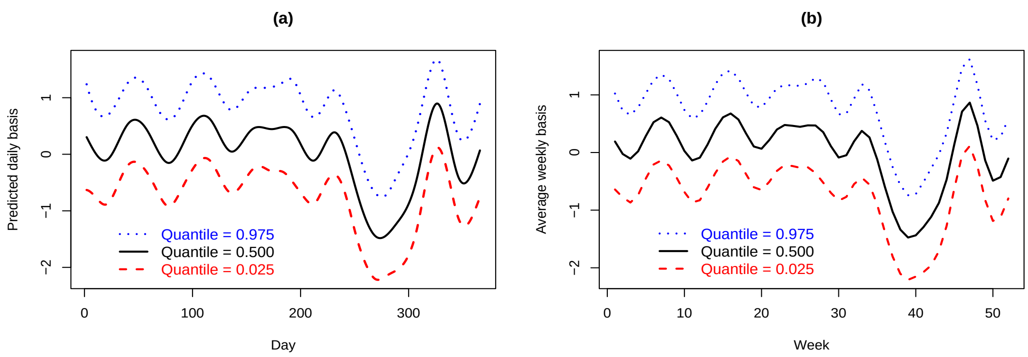

Figure 12: Plot of 95% pointwise confidence band of the predicted corn basis in daily (a) and average weekly (b).

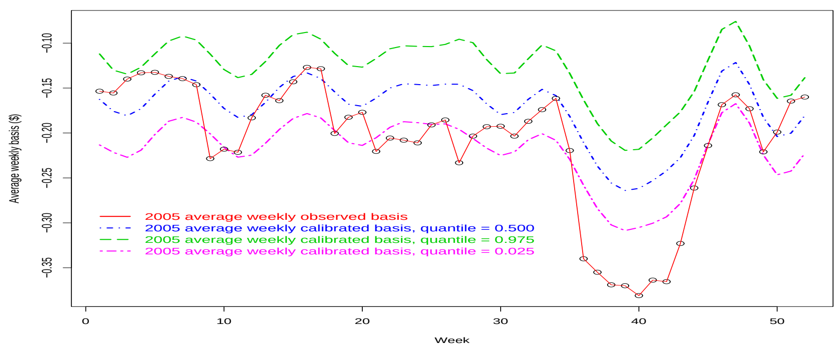

Figure 13: Plot of the 2005 average weekly observed corn basis and the calibrated confidence band with difference adjusted.

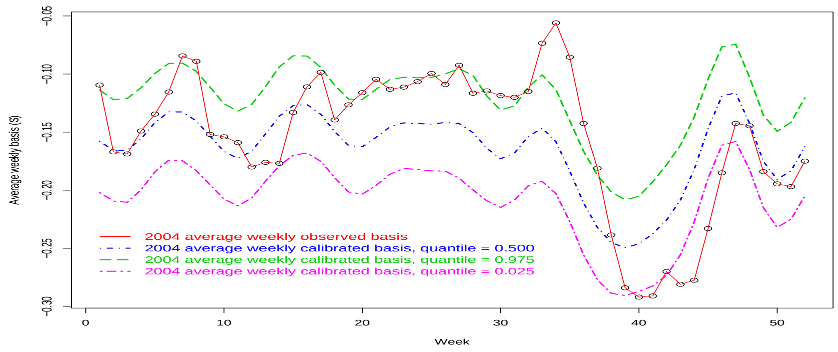

Figure 14: Plot of the 2004 average weekly observed corn basis and the calibrated confidence band with difference adjusted.

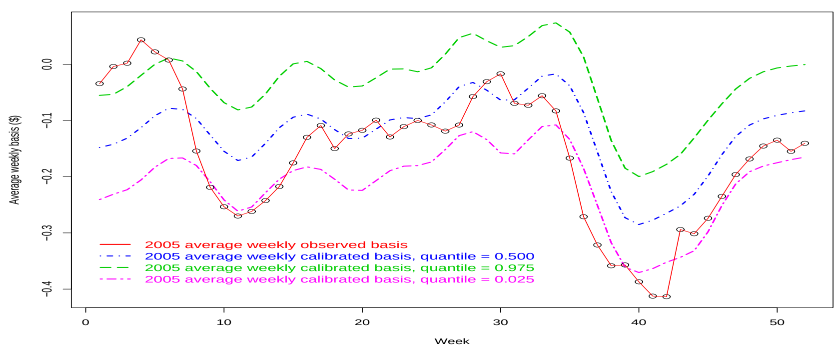

Figure 15: Plot of the 2005 average weekly observed soybean basis and the calibrated confidence band with difference adjusted.