Development of a Methodology for Projecting Domestic Percent Crop Treated

Economic Analysis Branch
Biological and Economic Analysis Division
Office of Pesticide Programs
U.S. Environmental Protection Agency
August 2002

Executive Summary

One of the responsibilities of the Environmental Protection Agency's Office of Pesticide Programs is to estimate the typical and maximum percent of a crop treated (PCT) with a particular pesticide. These estimates, referred to as "likely average PCT" and "likely maximum PCT", could be reflective of expected pesticide use in the short-term future (three to five years). The Office of Pesticide Programs (OPP) may estimate forecasts (projections) of PCT values. OPP worked for some time to develop a methodology to calculate PCT. This paper details recent progress towards refining these PCT projections.

To improve estimates of PCT, OPP considered various methods for estimating typical and maximum PCT along with criteria for selecting an appropriate model. This paper presents a brief account of the advantages and disadvantages of these methods and criteria. The finalized version of the forecasting methodology outlined by OPP includes a forecasting method, various "models" within the forecasting method to project typical PCT, a model selection criterion that identifies the most appropriate model and a model specific upper prediction interval (upper bound) to project maximum PCT. Additionally, OPP performed an evaluation of the forecasting methodology's accuracy. An objective measurement commonly used in forecasting "competitions" to quantify accuracy was employed to compare the proposed forecasting methodology to some benchmark methods, including the method currently in use.

Table of Contents

1 2 Introduction

1.1 Regulatory Overview

Pesticides are regulated in the U.S. under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Federal Food, Drug and Cosmetics Act (FFDCA). In 1996, Congress passed the Food Quality Protection Act (FQPA), which amended both FIFRA and FFDCA by requiring that aggregate and cumulative risks be considered by the Environmental Protection Agency in granting pesticide tolerance petitions and in assessing whether pesticides can be reregistered for use. Through these statutes, EPA evaluates risks posed by the use of each pesticide to make a determination of safety. Only if the Agency determines that such residues would be “safe”, may it authorize a tolerance to allow a pesticide residue in food.

One of the responsibilities of EPA’s Office of Pesticide Programs (OPP) is to assess the potential risks from pesticide residues for food consumption. The size of the potential risks depends on a variety of factors including the toxicity of the pesticide (how much harm, if any, is caused by specific amounts of the pesticide) and the magnitude of the exposure to the pesticide. In turn, exposure to a pesticide in the food supply depends on two factors: the amount of the pesticide present in food and how much food a person eats.

To develop estimates of such exposure, the Agency must use available and reliable, representative data for such risk assessments. These data include pesticide use statistics such as the percent of a crop treated (PCT) with a particular pesticide.

The FQPA-amended FIFRA also requires that OPP re-evaluate risks on a continuing basis. Specifically, the act permits the Agency to consider the percent of a crop that is treated with a pesticide (PCT), but requires that this information be re-evaluated (and, if necessary, the risk assessment be adjusted) after five years. Thus, estimates of PCT should be reflective of *future* pesticide use based on information from OPP data sources.

OPP is attempting to develop standard procedures that can be routinely used by a broad audience as a “first step” in projecting PCT. It is important than any such “first step” be well documented, clear and transparent, and reasonably simple to perform. Therefore OPP has compiled the current document to detail the development, realization, and evaluation of the methodology proposed to forecast “likely average” and “likely maximum” PCT. OPP recognizes that the function of any forecasting tool is not to rigidly dictate a forecast projection but rather to serve as a systematic means of illuminating and highlighting patterns and trends in data. Specialized professional expertise and experience, including specific knowledge of and judgment regarding agricultural practices and structural changes in the pesticide markets, can override forecasts based predominantly on standardized forecasting procedures. OPP believes that the methods described in this document will substantially improve our ability to realistically evaluate the potential exposure of individuals and the population to pesticides and contribute to the goal of protection of public health.

1.2 Forecasting Methodology and Policy Issues

It is important to note that implementing methods for forecasting pesticide use will necessarily involve and draw from a variety of "science policies." That is, implicit to any decision that involves prediction or forecasting are questions related to "How sure?", "How often?", "With what confidence?", "Over what time period?" and "How likely?." Each of these is an issue that can be informed by the science of statistical forecasting, but for which that discipline can offer no firm, uncontested, or incontrovertible answers. Any so-called "answers" to such questions are inherently judgmental in nature. This guidance does not investigate, nor even attempt to explore, the intricate nature of these decisions. Instead, it will simply recognize that consideration of these policy issues is on going and that further discussion in this area is needed.

Since the approaches discussed in the document are intended to apply only to the *methodological* aspects of the forecasting process, it is important to note that the approaches discussed herein do not support or prescribe the use of any one particular confidence level, percentile, percentage, or forecasting period associated with the process of regulatory decision-making. Thus, although the document may discuss a "95th percentile upper prediction interval" or a "five-year time horizon", these decisions have not been made and should not be inferred. Instead, they should be accepted solely as a simplification designed to make the technical discussion more concrete and the science policy "decision points" more apparent. Although this paper makes no attempt to directly address these issues, there are no intrinsic limitations in the methodology that would prevent such forthcoming decisions from being made or the described methodology from being adapted to include these decisions.

1.3 Scope and Organization of Document

Section 2 of this document details the development of the methodology proposed by OPP for forecasting PCT. An important component of detailing the development of the proposed methodology is a description of OPP's approach to forecasting PCT. Topics covered in this section include identifying candidate forecasting methods and model selection procedures. Documentation of this stage is motivated by EPA's practice of soliciting public participation and guidance for the development of its scientific methods. OPP believes an understanding of the decision process used to arrive at the proposed forecasting methodology will help to make this process transparent.

Section 3 describes the finalized version of the forecasting methodology, which is based on the exponential smoothing forecasting method. A desirable aspect of any standardized procedure employed by OPP for the purpose of forecasting PCT is that the process should be transparent, accessible, and reproducible. Therefore this section will provide a broad overview of the steps involved in producing PCT forecasts. These steps include parameter optimization, model selection, and calculation of PCT forecasts.
In order to gauge the accuracy of OPP's proposed methodology, section 4 includes an empirical

evaluation of the forecasts of various models. In an attempt to evaluate the ability of the methodology to select the "best" forecasting model, the forecasts of various models are compared to those of the methodically selected model. This "competition" is intended to evaluate the predictive accuracy of the methodically selected model.

2 Methodological Development

2.1 Candidate Forecasting Methods

Pesticide use is a dynamic process that is subject to unpredictable factors such as weather, pest population, and the pesticide market itself. These factors influence the pesticide applicators' decision-making process when seeking to answer questions such as: "Does a crop need to be treated this year?", "If so, how much of the crop should be treated?", "At what rate should the pesticide be applied?", and "Is the cost of pesticide application worth the increase in expected crop yield?." Modeling the complex relationships between these factors and the applicators' decision-making process, in order to forecast PCT, would require overwhelming amounts of information. As such, multivariate methods that attempt to model the relationship between percent crop treated and a wide variety of explanatory variables were ruled out as candidate methods. Rather OPP has focused on univariate methods where forecasts depend only on the past values of PCT.

The exclusive use of historic data for producing forecasts is the identifying characteristic of extrapolation methods. Methods that can be used to extrapolate time series data such as PCT include linear regression, Box-Jenkins methods, and exponential smoothing. The models provided by these extrapolation methods in addition to a simple mean/average model were considered as candidates for forecasting PCT. A brief description of the methods and/or model(s) and the reasons for eliminating/including them as part of OPP finalized methodology follow.

2.1.1 Mean/Average Model

The mean model is one of the simplest methods that could be used for forecasting time series data. To arrive at a forecast, all that is required is taking the arithmetic mean (i.e. average) of the past observations. By estimating percent crop treated as the mean of the past values, one assumes that the observations are independent samples from a common population and that any differences are due to some random error. In other words, any variation in the annual values is unexplained. Such a method would not account for any trend in the data. Initially, OPP considered using the mean model for time series that do not exhibit a trend. Using this method on a time series that is exhibiting a trend would expose the forecasts to serious criticism. For example, if the use of a pesticide has been increasing (or decreasing), one could argue that the average underestimates (or overestimates) the pesticide's use. Nonetheless, instances in which little data is available, such as a newly registered and/or reported use, the mean model could provide adequate forecasts. This method was eventually discarded in favor of other forecasting techniques, but still serves as "benchmark" method with which to compare forecasts.

2.1.2 OLS Regression

Although commonly employed as a multivariate method, linear regression can be used as a univariate method for time series data. The linear regression approach models the relationship between the data points and the time of their observation as a linear function. The linear relationship is specified by the slope and intercept parameters. Regression methods differ in the procedures used to estimate the values of these parameters. The most commonly used method is ordinary least squares (OLS), which estimates the parameters by minimizing the squared residuals. The residuals are the differences between the "predicted" values and actual values of the time series (here predicted refers to the value of a data point as estimated by the OLS model, not in the sense of forecasting).

The nature of the trend in a time series is related to the concept of "stationarity." Generally a time series is stationary if the mean and variance are constant over time and the value of the covariance between two time periods depends only on the distance between two time periods and not the actual time at which the covariance is computed. With linear regression, the mean of the time series is modeled to increase or decrease by the same amount for every time period; the change from one time period to the next is the slope parameter. Thus the time series is considered nonstationary. However with linear regression, one assumes the time series can be stationarized by accounting for the trend. In other words, if one were to subtract the trend from each observation, the time series would have a constant mean and variance. This type of trend is referred to as a deterministic trend. Generally a deterministic trend is constant throughout the time series; while a variable trend is referred to as being stochastic. OPP believes it is more realistic to assume that trends in PCT may change over time. Therefore linear regression methods for forecasting PCT were ruled out.

2.1.3 IRLS Robust Regression

In addition to OLS regression, OPP considered iteratively re-weighted least squares (IRLS) regression. IRLS regression is more specifically classified as a robust regression method. The term "robust" refers to the method's goal of obtaining robust parameter estimates by dampening "outlier" effects. An outlier is an observation with a relatively large residual. Sometimes the residual of an outlying observation is "balanced out" by the residuals of the other observations (more common for cross-sectional data than to time series data). Other times, outliers can greatly affect parameter estimation. The potential influence a data point has on parameter estimation is referred to as leverage. IRLS aims to diminish the leverage of these outliers by weighting the residuals via some weight function(s). Generally, observations with relatively large residuals are assigned smaller weights than those of observations with relatively small residuals; thus mitigating the leverage of outliers. As the name implies, IRLS repeats the process of weighting the residuals and calculating the parameter estimates until there is negligible difference between subsequent sets of weights (Hamilton, 1992).

In addition to modeling a deterministic trend, there is another disadvantage of using IRLS to forecast PCT. IRLS was examined due to its ability to "down-weight" outlying observations. OPP thought this would be helpful in situations where some of the earlier values for PCT were uncharacteristically high or low compared to more recent observations. Such a change in the "level" of PCT could be due to some shift in the market, such as the registration or cancellation of some competitive chemical; or a consistent increase or decrease in pest pressure. The hope was that IRLS would be able to discount these initial observations and start tracking the most recent level and trend of the time series. However, robust regression regards outlying observations the same whether they occur at the beginning or the end of the time series. If such change(s) were to take place and the most recent observations of PCT were reflective of such change(s), OPP certainly would not want to disregard such observations when calculating PCT forecasts.

2.1.4 Box-Jenkins Methods (ARIMA)

Box-Jenkins (BJ) methods were also considered for forecasting PCT. BJ methods model time series as autoregressive integrated moving average (ARIMA) processes. When modeling a time series as an ARIMA process, the first step is to stationarized the data. Differencing is a commonly used method for stationarizing time series data. The process of differencing a time series involves taking the difference between subsequent observations. The term "integrated" (I) in ARIMA refers to this differencing process. Once a stationarized, the data is modeled to be an autoregressive (AR) process and/or moving average (MA) process. Generally an AR process models an observed value to be depend upon previously observed value(s), a constant term (i.e. deterministic) and a stochastic term. An MA process models an observed value to be dependent upon a constant term and a linear combination (i.e. weighted average) of multiple stochastic terms. The above explanations provide a generally description of ARIMA processes; a determination of the ARIMA process which best fits a particular time series is a iterative procedure that involves analyzing the residuals of the ARIMA process.

The BJ methods were developed as a framework to recognize and exploit patterns of variability in time series data. Identifying characteristics of the time series are then used to select an appropriate ARIMA process to model. The fact that BJ methods incorporate procedures for identifying and modeling nonstationary time series (and variable trends) makes it an attractive univariate method. However, it is generally accepted that at least fifty observations is needed to employ such methods. Typically time series for PCT contain much fewer observations. As appealing as BJ methods are, OPP believes the majority of PCT time series would not meet the data requirements for applying BJ methods.

2.1.5 Exponential Smoothing

Exponential smoothing (ES) methods were considered by OPP for the purpose of forecasting PCT. ES methods model time series data in the manner similar to BJ methods. In fact many ES models have an equivalent ARIMA model. The ES models of interest to OPP, simple exponential smoothing (SES), linear exponential smoothing (LES), and damped-trend exponential smoothing (DES), all have ARIMA equivalencies. Although BJ and ES methods can model seasonality in time series data, seasonality is not a relevant characteristic of annual data such as PCT. Like BJ methods, the ES methods can be used on nonstationarity data. However unlike BJ methods, the ES model selection procedure is typically not based on examining the residuals to determine if a model effectively stationarizes the data.

Every ES model can be considered as having two components: level and trend. Both the level and trend have a corresponding smoothing parameter, α and β respectively. For the models of interest, the smoothing state is the arithmetic sum of these two components. The smoothing state is the estimated or fitted value of an observation for particular time period. ES models attempt to estimate the value of these components based on weighted averages of the observations (for the level) or differences in the observations (for the trend). The weights are specified such that the most recent observations have the greatest effect on a component's estimation. In fact, the name "exponential smoothing" is derived from the specification that the weights increase "exponentially" from the most distant to the most recent observation--thus providing "smoothed" estimates of the level and trend.

The smoothing parameters assume values from zero to one and determine the value of the weights. When a smoothing parameter is equal to one, all of the weight is given to the most recent observation. Thus the estimate of the corresponding component is completely determined by the previous observation. At the other extreme, when a smoothing parameter is equal to zero, the weights are all the same: zero. Thus the value of the component is never updated from its initial estimate. These extreme values can yield useful model equivalencies. For example, when α is equal to one, the SES model is equivalent to the "naïve" model where the forecast for a time series is simply equal to most recent observation. Typically however, values in between these extremities are used to specify ES models. The larger the value of a smoothing parameter, the more influence the observations from the recent past will have on the component estimations. Conversely when smoothing parameters assume smaller values, the influence of component estimation is more evenly distributed among observations from the distant and recent past.

Being the most basic of the ES models, the simple ES model can be considered as having a trend component equal to zero. In other words, the smoothing state is simply the level component. The SES model attempts to track the changing level of a time series. If the smoothing parameter is relatively small, then the time series is being model as having a level that does not change much from year to year. On the other hand with a relatively large smoothing parameter, the SES model updates its estimate of the level frequently. For the SES model, the forecasts are the same regardless of the number of years being forecasted.

The linear ES model has the added complexity of including a trend component (not equal to zero). The term "linear" refers to the additive nature of the trend component: the smoothing state is equal to level plus the trend. In conjunction with modeling the changing level of the observation, LES also models the shifting trend in the series. As with the simple model, the smoothing parameters govern how often the components are revised. Thus a time series can be modeled as having a fairly stable level (small α) and a trend component that changes frequently (large β). The LES forecasts increase by an amount equal to the last estimate of the trend component. To illustrate, if l_n is the last estimate of the level for a time series and b_n is the last estimate of the trend, then the forecast for next three years would be $l_n + b_n$, $l_n + 2*b_n$, and $l_n + 3*b_n$ respectively.

In addition to having a level and trend component, the damped-trend ES model includes a parameter not previously mentioned: the damping coefficient, f*. As the name implies, damped-trend model damps the trend component of the model. Like* a *and* b*, the damping coefficient can vary from zero to one. The closer* f *is to zero, the more rapidly the trend is damped. Conversely, the damping of the trend is more gradual when* f *is larger. In fact the linear model is a special case of the damped-trend model for which* f *is equal to one. The DES is a useful model when there is evidence to suggest that the current trend in the data is unlikely to continue. As an example the forecasts for the next three years would be* $l_n + (f)*b_n$, $l_n + (f + f^2)*b_n$ *and* $l_n + (f + f^2 + f^3)*b_n$ *respectively.*

The ES method has some very attractive characteristics: the models available with this method allow for a variety of trends, put more emphasis on the most recent observations and are not data intensive. Until recently the lack of a well-developed modeling framework presented some disadvantages to employing the ES method. However, in recent years Hyndman and others (Hyndman et al forthcoming) have done some innovative work to provide analytical formulae for the forecast variances of these models. These formulae allow for the calculation of prediction intervals for exponential smoothing forecasts. As such, ES models are the focus of OPP's proposed methodology for forecasting PCT.

2.2 Model Selection Criteria

This subsection details some procedures considered by OPP for selecting an appropriate exponential smoothing model to forecast a particular time series. The model selection criteria that are the basis of these procedures are the Gardner-McKenzie protocol and the Bayesian information criterion. The Gardner-McKenzie protocol selects a model by examining time series before fitting any of the models to the data. In contrast, the Bayesian information criterion selects the best model *after* the various exponential smoothing models have been fitted to the data. Ultimately use of the Gardner-McKenzie protocol lacked empirical support. While other model selection criteria similar to the Bayesian information criterion, such as Akaike's information criterion are favored by researchers (Hyndman et al 2002). Both methods are presented in the following section, although the Bayesian information criterion was ultimately chosen as the basis for model selection procedure.

2.2.1 Gardner-McKenzie Protocol

The Gardner-McKenzie protocol (GMP) is a simple procedure for *classifying* trends in time series data (Gardner 1988). Once the trend has been classified, Gardner suggests which exponential smoothing model to use. To determine the type of trend present in the time series, the variance of the observations is compared to the variance of first order differences of the observations and the variance of the second order differences of the observations. The first order differences are calculated by taking the arithmetic difference between subsequent observations. Similarly the second order differences are calculated by taking the arithmetic difference of the first order differences. The protocol is based on the proposition that if there is no trend in the time series then the variance of the observations is the minimum, if there is a moderate trend then the variance of the first order differences is the minimum and if there is a strong trend then the variance of the second order differences is the minimum. For a trend categorized as "none", "moderate" or "strong", Gardner suggests using simple, damped-trend and linear exponential smoothing respectively. An example demonstrating the use of the GMP is given in Table 1. The first order and second order differences are denoted as ΔPCT and Δ^2PCT respectively.

2.2.2 Bayesian Information Criterion

The Bayesian information criterion (BIC) is a statistic that quantifies the relative "goodness-of-fit" and "complexity" of a model. The component of the BIC that measure the goodness-of-fit of the model is the mean squared error (MSE). For the exponential smoothing models, the MSE is similar to the more familiar MSE associated with regression models. In general, the MSE is the average or "mean" of the "square" of the "errors", where the error is simply the difference between the *fitted* value of the model and the *observed* value of the time series data. Mathematically,

$$MSE = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n}$$

where n is the number of observations in the time series, Y_i and $\hat{Y}_i$ are the observed and fitted values of the observations respectively. A relatively small MSE is indicative of a model that fits the historical data well. At the same time, the number of parameters employed to calculate these fitted values can supplement the MSE measure of accuracy. These two concepts are central to the definition of the BIC, which can be formalized as:

$$BIC = n\ln(MSE) + k\ln(n)$$

where, as before, n is the number of observations in the time series, k is the number of parameters for the model, and ln(.) refers to the natural logarithm. As with the MSE, a minimal BIC indicates a "good" model. Holding the MSE constant, it can be seen that for a given time series as the number of parameters used to specify a model increases, the BIC increases. Thus the BIC "penalizes" models with more parameters. The purpose of using the BIC is to select the model that provides the "best" fit to the historical data with the minimal number of parameters. The rationale behind this model selection procedure is that the model that performs well at *fitting* will also perform well at *forecasting*.

2.3 Estimating Maximum PCT

When describing PCT forecasts, it has been understood that the pesticide use statistic referred to has been the "likely average PCT." "Likely average PCT" is the term OPP uses when denoting typical or average use for the pesticide crop combination of interest (usually at the national level). However, another pesticide use statistic utilized by OPP is the "likely maximum PCT", the maximum percent of a crop expected to be treated with a specific pesticide (again, usually at the national level). OPP considered several different techniques for estimating the likely maximum PCT. These include general techniques that could be used regardless of the specified model, and others that are model specific.

2.3.1 General Techniques

General techniques could be used for estimating maximum PCT that do not take into account the model being used to forecast the "likely average PCT" or its parameters. OPP considered several options, which can be characterized unsophisticated. Generally, the maximum of the historical observations might serve as an estimate of the "likely maximum PCT", and has been by at least one OPP stakeholder. OPP decided that such estimates were inadequate representation of maximum use. If historic maxima were used for a crop that has not experienced high pest pressure, they could greatly underestimate future maximum use. On the other hand, if a pesticide is exhibiting a decidedly decreasing trend in usage, the historic maximum may grossly overestimate the maximum PCT. Such general techniques do not utilize all relevant information when estimating likely maximum PCT, and OPP has focused on other options.

2.3.2 Model Specific Techniques

The model specific techniques employed by OPP involve calculating prediction intervals. To be more precise, the upper bound of a prediction interval or upper prediction interval (UPI) is of primary interest to OPP. Generally, a UPI is a point estimate plus some multiple of the estimated standard deviation the model error. Additionally, there is a specific "confidence level" associated with the UPI based on assumed distributional properties of the model errors. As an example, for the "mean model" described earlier, one could use a UPI for a normal distribution as detailed in Hahn (p. 63). Given *n* normally distributed observations with standard deviation *s*, there is a 100%X(1-α) probability that none of the next *m* future observations will exceed the upper bound,

$$Y_m = \bar{Y} + r_{(1-\alpha; m, n)} s$$

where *r* is a tabulated probability distribution specifically for calculating prediction intervals for observations from a normal distribution . In this case, the estimate standard deviation of the model error would be *s*, the square standard deviation of the observations. For α with a value of 0.05, the confidence level associated with this UPI would be 95% or 100%X(1 - 0.05). A conservative estimate (i.e. underestimate) for this upper bound that utilizes the more familiar and readily computable t statistic is,

$$Y_m = \bar{Y} + \left(1 + \frac{1}{n}\right)^{1/2} t_{(1-\alpha/m;\, n-1)} s$$

It is important to keep in mind that this UPI is not expected to be exceeded by *any* of the next *m* future observations. This is quite different than computing five separate prediction intervals for *each* of the next *m* future observations. Inspection of the formula for calculating this conservative estimate of the UPI for the mean model verifies some of its intuitive properties. The UPI becomes larger when the number of future observations being predicted (*m*) increases or when the variability of the data series (*s*) increases. However, the UPI diminishes as the number of observations (*n*) increases.

Similar UPIs can be calculated for other forecasting models considered in this paper. Analytic UPIs hinge on the calculation of an estimate of the error variance (square of the standard deviation of the model errors). For methods such as regression and Box-Jenkins, estimates of the error variance for the various models have been specified and are widely used. Until recently, this has not been the case with exponential smoothing methods. Some exponential smoothing models do have equivalent ARIMA models, from which "reasonable" estimates of the forecast error can be calculated (Armstrong 2001, p. 481). However, Hyndman and his colleagues laid the foundation for computing UPIs for several ES models (2002). For the ES models that have ARIMA counterparts, these estimates of the error variance are in agreement with one another.

Unlike the mean model, for these ES models there is no direct method for computing a UPI to contain multiple future observations-such UPIs are sometimes referred to as *simultaneous* UPIs. However, OPP is interested developing an estimate for the "likely maximum PCT" for more than one year, which would entail the use of simultaneous UPIs. Therefore, OPP is considering calculating such UPIs by specifying an upper bound such that the product of the confidence levels of the single year UPIs is equal to the desired overall confidence level. To illustrate, suppose for a specific PCT time series, OPP would like to calculate a UPI for which there would be a 95% probability that the UPI would not be exceeded for any of the next three years. The confidence levels for the *individual* UPIs would necessarily be greater than 95%. One possible combination of the individual confidence levels that would yield the desired overall confidence level of 95% would be 99.5%, 98.5% and 97% ($0.995 \times 0.985 \times 0.970 \approx 0.95$).

For any given ES model, the error of a point forecast is characterized by the MSE, the values of the estimated models parameters and the number of time periods beyond the last observation being forecast. Thus the estimated error variance and, hence, the UPI is determined by how well the model fit the data, how much weight is given to recent observations, how rapidly the trend is dissipating and how far in the future the estimate is being provided.

3 Proposed Methodology

Having described the development of the methodology, this paper will now focus on the methods and procedures that make up OPP's proposed methodology for forecasting PCT. Of the univariate techniques considered by OPP, exponential smoothing is simple, yet it provides an assortment of methods for modeling data. There are procedures for implementing ES methods such that the whole process of parameter selection, model selection, and forecasting can be automated. It is important to note that any automated forecasting procedure employed by OPP for the purpose of estimating PCT or any other pesticide use statistic will be reviewed by an analyst. Ultimately it is the analyst's responsibility to ensure that the estimate is reasonable and to adjust the estimate to be reflective of changes in the pesticide market not captured in existing usage data, such as the introduction or discontinuation of alternative pesticides.

3.1 Estimating Parameters of Forecasting Models

Among the various exponential smoothing models, OPP selected ones which appear to be appropriate for forecasting PCT: simple exponential smoothing (SES), linear exponential smoothing (LES), and damped-trend exponential smoothing (DES). Each model has one or more components, referred to as level and trend. The SES model has a single component and a smoothing parameter, α, for the level. Both the LES and DES models have two smoothing parameters, one for the level (α) and one for the trend (β). Additionally, the DES model has a damping coefficient (ϕ). The mean squared error (MSE) is used to estimate the parameters for each model. The values selected for α, β, and ϕ for each model are those that minimize the MSE. Recall that the MSE for these models is similar to that use in OLS regression in that it is based on the differences between the fitted values and the actual values of the time series. Thus the values of α, β, and ϕ are those that provide the "best fit" for the specific model.

3.2 Model Selection

Once model parameters have been estimated, forecasts of PCT can be calculated for each model, and the "best" of these three models can be identified. The model selection criterion employed in OPP's methodology is the Bayesian information criterion (BIC). For the BIC, the model with the best fit is that which minimizes the MSE with the fewest number of parameters. For example, if both the SES and the DES models provide comparable MSE's, one would select the SES model using the BIC because it has one parameter, while the DES model has three. The model with the smallest BIC is considered to be the "best" model.

3.3 Forecasting PCT with Optimal Model

Once the model that minimizes the BIC has been selected, one can forecast the values of "likely average PCT" and "likely maximum PCT" using that model. Depending on the number of years to be forecasted, the point forecast(s) of PCT from the "optimal" model is/are used for the likely average. For the simple ES model this forecast is simply the last estimated value of the model level. For the linear ES model, the point forecast is the last estimate of the level plus the last estimate of the trend. For the damped-trend ES model, the forecast is the last estimate of the level plus the damped-trend component, which is the last trend component adjusted by the damping coefficient. These estimates of likely average PCT are easily calculated. However, the estimates for likely maximum PCT are more complicated.

In order to obtain estimates of the likely maximum PCT, the upper bound of simultaneous prediction intervals (UPIs) is calculated. As mentioned earlier, a simultaneous UPI is an upper bound for more than one year of forecasts. For OPP's purposes, the number of years forecasted will generally be between three and five. The MSE serves as a basis for calculation of the forecast variance for the ES models. Other factors that affect the forecast variance are the parameter values of the model and the number years beyond the forecast horizon being forecasted. The forecast horizon is the last year (or other appropriate time unit) beyond which forecasts are obtained. As one would expect, since the reliability of a forecast decreases the further out one forecasts, the forecast variance increases as the years beyond the forecast horizon increases. Additionally, the more parameters there are, the larger the forecast variance.

Once the forecast variance of each point forecast has been calculated, the simultaneous UPI can be obtained. The UPI is chosen such that the product of the confidence levels is equal to 95% (approximately). In order to determine the confidence level for a point forecast, the probability of the point forecast *not to exceed* the UPI is calculated. Given its forecast variance, this probability is computed for each point forecast. In order to compute these probabilities, one must first standardize the UPI. This is done for each point forecast (Y_{h-i}) by subtracting it from the UPI (Y_{UPI}) and dividing by the square root of its forecast variance ($var(Y_{h-i})$),

$$z_i = \frac{Y_{UPI} - Y_{h+i}}{\sqrt{var(Y_{h+i})}}$$

Here the $h+i$ subscript denotes the i^{th} forecast beyond the forecast horizon h. These standardized values, the z_i's are assumed to be associated with a normal distribution. Each z_i has an associated probability denoted as $P_i(Z \leq z_i)$. That is, the probability of observing a value less than or equal to z_i. This probability represents the confidence level of Y_{UPI} associated with the Y_{h-i} forecast. Additionally the probability P_c is equal to the product of these individual probabilities,

$$P_c(Z \leq z_1 \mid \ldots \mid Z \leq z_m) = \prod_{i=1}^{m} P_i(Z \leq z_i)$$

Here m denotes the number of point forecasts being made. Thus P_c is the combined probability of observing values less than or equal to z_i for all m z_i's, which represents the confidence level of the simultaneous UPI for all m point forecasts. Since the forecast variance and the point forecasts are fixed values, the UPI can be found by varying its value until the combined probability is approximately equal 0.95.

4 Methodology Evaluation: An Empirical Example

The primary motive for updating the OPP methodology for forecasting PCT is to provide more reliable and consistent estimates of PCT. The model selection process is a concept of the proposed methodology that makes it more appealing than the current methodology, which employs only one forecasting model. Therefore a "competition" was performed to not only compare the PCT estimates produced by the various exponential smoothing models and those of the current model, but to examine the proposed methodology's ability to select the "best" model. Here the term "best" refers to the accuracy of forecasting model. A quantification of accuracy commonly used in empirical studies of forecasting methods is the mean absolute percentage error or MAPE. The term "error" refers to the forecasting error, the difference between the actual value (Y_t) and its forecasted value ($\hat{Y}_i$). The formula for computing the MAPE is

$$MAPE = \frac{100}{m} \sum_{i=1}^{m} \frac{\left| Y_i - \hat{Y}_i \right|}{Y_i}$$

In order to evaluate forecasts, a "hold out sample" must be specified. The hold out sample consists of observations that are not used to develop the model or the forecasts. The accuracy of forecasting model is evaluated by computing the MAPE for this hold out sample.

For OPP's purposes, a hold out sample consisted of five years of annual data from 1996 to 2000. Seventeen pesticide crop combinations (PCCs) were used in this evaluation. These time series were selected based on their "interesting" graphical properties and do not represent a random sample of PCT data. For nine of the PCCs the data are available from 1987; for the other eight data the earliest available year is 1990. In addition to the exponential smoothing models and current model, the mean model is used as a "benchmark" with which to compare the forecasts.

A side-by-side comparison of the various PCT forecasts is shown in Table 2. The highlighted forecasts are those of the forecast model selected by the proposed methodology and the dotted line represents the forecast horizon. The point forecasts, which represent the "likely average PCT" are in regular text and the interval forecasts, which represent the "likely maximum PCT" are italicized. Under each PCC, "flags" are displayed when an upper prediction interval is exceeded. These flags do not specify the year for which the UPI was exceeded and reflect decimal values not displayed in the table. Table 3 is a summary chart of the MAPE for the different forecast models. Again the highlighted values are those that represent the MAPE of the methodically selected model and the bold values represent the minimum MAPE for the particular PCC. In only three cases, the minimum MAPE was not generated by the exponential smoothing models. As can be seen by the average MAPE at the bottom of the column, on average, the ES models yielded considerably more accurate forecasts than those of the mean and current models. However, all three ES model performed comparably to one another. Nonetheless, the average MAPE of the forecasts using the BIC model selection method is smaller than that of any one forecasting model. Taken as a whole, this competition is an indicator that the ES models utilized by the proposed methodology perform better than the current model and that the proposed model selection process performs adequately at selecting best of these available models. OPP believes the proposed exponential smoothing method for forecasting PCT will better promote EPA's goal of protecting the environment and human health by providing better estimates of pesticide use in future years.

References

Armstrong, J.S. (1984) “Forecast by Extrapolation: Conclusions from 25 Years of Research”, *Interface*, 14, 52-66.
Full text at <http://www-marketing.wharton.upenn.edu/forecast/papers.html>

Armstrong, J.S. (1994) “The Fertile Field of Meta-Analysis: Cumulative Progress in Agricultural Forecasting”, *International Journal of Forecasting*, 10, 147-149.
Full text at <http://www-marketing.wharton.upenn.edu/forecast/papers.html>

Armstrong, J.S. (2001), *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Boston: Kluwar Academic Publishers.

Gardner, E.S. Jr. (Feb 1987), “Short-Range Forecasting”, *Lotus*, Vol. 3, No. 2, 54-58.

Gardner, E.S. Jr. (Mar 1988), “Forecasting with Exponential Trends”, *Lotus*, Vol. 4, No. 3, 27-30.

Gardner, E.S. Jr. (May 1988), “How to Detect Trends and Seasonal Cycles”, *Lotus*, Vol. 4, No. 5, 44-50.

Gardner, E.S. Jr. (Mar 1987), “Using Exponential Smoothing”, *Lotus*, Vol. 3, No. 3, 61-66.

Hahn, J.H., W.Q. Meeker (1991), *Statististcal Intervals: A Guide for Practitioners*. New York: John Wiley & Son INC.

Hamilton, Lawrence C. (1992), *Regression with Graphics*. California: Wadsworth, Inc.

Hyndman, R.J., A.B. Koehler, R.D. Snyder, and S. Grose (2002) "A state space framework for automatic forecasting using exponential smoothing methods". International J. Forecasting, 18(3), 439-454.
Full text at <http://www-personal.buseco.monash.edu.au/~hyndman/papers/>

Hyndman, R.J., A.B. Koehler, J.K. Ord, R.D. Snyder (forthcoming), “Prediction intervals for exponential smoothing state space models.”
Full text at <http://www-personal.buseco.monash.edu.au/~hyndman/papers/>

TreeCrop01-Fungicide03			
	%PCT	ΔPCT	Δ^2PCT
Variance	353.98	75.29	266.97
Index	100%	21%	74%
Trend:	None	Moderate	Strong
		*******	
Year	%PCT	ΔPCT	Δ^2PCT
1990	1		
1991	0	-1	
1992	11	11	12
1993	14	3	-8
1994	23	9	6
1995	26	3	-6
1996	49	23	20
1997	47	-2	-25

RowCrop01-Herbicide01			
	%PCT	ΔPCT	Δ^2PCT
Variance	73.22	6.98	10.34
Index	100%	9%	14%
Trend:	None	Moderate	Strong
		*******	
Year	%PCT	ΔPCT	Δ^2PCT
1987	29		
1988	27	-2	
1989	26	-1	1
1990	26	0	1
1991	26	0	0
1992	25	-1	-1
1993	22	-3	-2
1994	19	-3	0
1995	10	-9	-6
1996	7	-3	6
1997	6	-1	2

Table 1 Gardner-McKenzie Protocol

(a) (b)

Forecasting Summary of Models Predicting PCT for Years 1996 to 2000

Pesticide Crop Combination	Year	PCT	Mean	Mean UPI	SES	SES UPI	LES	LES UPI	DES	DES UPI	Current	Current UPI
Row Crop01-Herbicide01	1987	25										
	1988	27										
	1989	26										
	1990	24										
	1991	26										
	1992	25										
	1993	22										
	1994	19										
	1995	10										
	1996	7	23	41	10	26	2	7	2	12	20	29
	1997	5	23	41	10	26	0	7	0	12	20	29
LES UPI exceeded	1998	5	23	41	10	26	0	7	0	12	20	29
	1999	4	23	41	10	26	0	7	0	12	20	29
	2000	3	23	41	10	26	0	7	0	12	20	29
Row Crop01-Herbicide02	1987	63										
	1988	65										
	1989	68										
	1990	65										
	1991	65										
	1992	64										
	1993	67										
	1994	69										
	1995	67										
	1996	66	65	74	67	79	69	75	69	76	66	74
	1997	71	65	74	67	79	70	75	69	76	66	74
	1998	69	65	74	67	79	71	75	70	76	66	74
	1999	68	65	74	67	79	72	75	71	76	66	74
	2000	70	65	74	67	79	72	75	72	76	66	74
Row Crop01-Herbicide03	1987	12										
	1988	14										
	1989	16										
	1990	17										
	1991	18										
	1992	19										
	1993	19										
	1994	24										
	1995	24										
	1996	22	15	30	24	35	25	34	25	34	20	24
	1997	24	14	30	24	35	27	34	27	34	20	24
	1998	20	15	30	24	35	26	34	26	34	20	24
Current UPI exceeded	1999	21	15	30	24	35	30	34	29	34	20	24
	2000	22	15	30	24	35	32	34	31	34	20	24
Row Crop01-Herbicide04	1987	24										
	1988	25										
	1989	24										
	1990	25										
	1991	29										
	1992	30										
	1993	32										
	1994	32										
	1995	32										
	1996	30	25	39	32	40	34	41	34	41	30	34
	1997	32	28	39	32	40	25	41	25	41	20	24
	1998	25	25	39	32	40	35	41	36	41	30	34
	1999	14	28	39	32	40	28	41	27	41	20	24
	2000	11	25	39	32	40	39	41	35	41	30	34

Table 2 Forecast Comparison

(a)
(b)

(c)

(d)

Mean Absolute Percentage Error						
MAPE = $(100/m)\Sigma\,\lvert e_t/Y_t\rvert$, where e_t and Y_t denote the t^{th} forecast error and observation						
Pesticide Crop Combination	*Mean*	*SES*	*LES*	*DES*	*Current*	
RowCrop01-Herbicide01	396%	104%	94%	94%	334%	Selected Model
RowCrop01-Herbicide02	6%	3%	3%	2%	4%	Minimum MAPE
RowCrop01-Herbicide03	18%	10%	30%	28%	8%	
RowCrop01-Herbicide04	58%	69%	97%	96%	62%	
RowCrop02-Insecticide01	27%	18%	12%	14%	22%	
RowCrop02-Insecticide02	N/A	N/A	N/A	N/A	N/A	
RowCrop02-Insecticide03	155%	69%	55%	82%	98%	
RowCrop02-Insecticide04	54%	30%	17%	17%	43%	
RowCrop02-Insecticide05	163%	125%	52%	59%	139%	
TreeCrop01-Insecticide06	26%	18%	23%	23%	22%	
TreeCrop01-Insecticide07	28%	26%	31%	28%	29%	
TreeCrop01-Fungicide01	30%	28%	36%	34%	28%	
TreeCrop01-Insecticide08	74%	72%	81%	81%	76%	
TreeCrop01-Fungicide02	132%	112%	88%	92%	108%	
TreeCrop01-Fungicide03	72%	41%	25%	23%	64%	
TreeCrop01-Fungicide04	51%	49%	55%	55%	50%	
TreeCrop01-Fungicide05	68%	78%	99%	93%	70%	
Average MAPE:	85%	53%	50%	51%	72%	
Average MAPE Using Propose Methodology:	48%					

Table 3 Accuracy Measurements