Predicting Refill of the Patuxent and Occoquan Reservoirs: A New Interactive Tool based on ARIMA

Prepared by

Mark A. Lorie
Erik R. Hagen

May, 2007

Interstate Commission on the Potomac River Basin
Section for Cooperative Water Supply Operations on the Potomac
51 Monroe St, Suite PE-08
Rockville, Maryland 20850

Report No. 07-4

Purpose and Background

The Patuxent and Occoquan reservoirs in the Washington Metropolitan Area (WMA) are operated to meet a specific target: that they reach 90 percent full by June 1 of each year at a 95 percent reliability (i.e., 95 percent of the time), per agreement by the Water Supply Coordination Agreement of 1982 (WSCA, 1982). The tool discussed in this paper represents an improvement in the method by which water managers can calculate appropriate reservoir withdrawal levels while meeting this reliability standard.

In the late 1990s, ICPRB CO-OP developed an interactive graphical tool to help reservoir operators craft sustainable withdrawal strategies that met this reliability standard. This tool was based on conservative estimates of historical reservoir inflows. Specifically, the 5th percentile flow of the historical record for each month between the current time and the following May were strung together and used as the benchmark series of inflows. This series of flows represented a highly conservative approximation of the 5% probability event and, therefore, a safe benchmark for operations. In the early 2000s, this tool was augmented with an interactive spreadsheet visualization tool that used position analysis. Position analysis is a simulation of reservoir storage given the historical record of streamflows and conducted with an assumed starting storage and various assumptions about reservoir withdrawal rates. This method is similar to work done by Hirsch (1978).

Use of the historical record alone depends on a limited, and specific, hydrologic sequence, and may not provide a sufficiently comprehensive dataset for the analysis. Use of the historical record in a position analysis further includes the assumption that all flows in the historical record are equally likely at any given time. Most hydrologic processes exhibit significant autocorrelations, so recent conditions can provide valuable information as to potential future conditions. An alternative to the historical record is the use of auto-regressive integrated moving average (ARIMA) modeling. (See Box et al, 1994 for details on ARIMA modeling.) Streamflow prediction using ARIMA modeling incorporates autocorrelation combined with information about current conditions to develop flow prediction estimates, thereby providing a forecast conditioned on current hydrology. ARIMA modeling provides the backbone of the spreadsheet tools introduced in this paper. These tools incorporate the latest flow information, develop conditional forecasts of inflow to the reservoir, and allow water managers to develop a reservoir withdrawal strategy that also meets the regional reliability standard. These tools are easy to use and incorporate improvements to the interactive graphical interface.

The model methods and development are discussed using the Patuxent system as an example. The data and results specific to the Occoquan system are presented at the end of this paper.

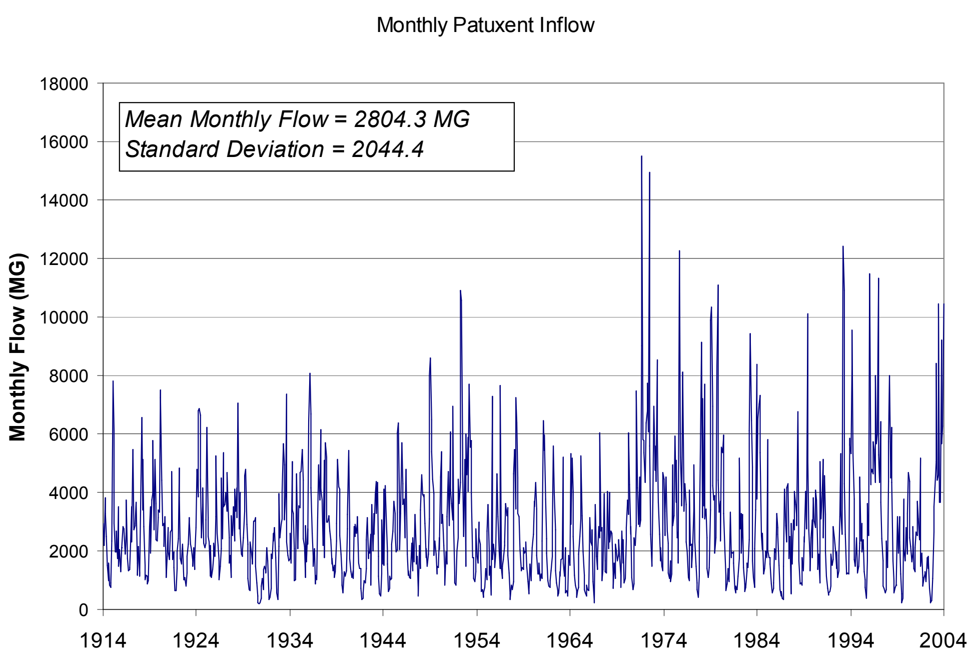
Methods

Data

A dataset of Patuxent Reservoir inflows was developed by ICPRB for the 2005 water supply reliability study (Kame'enui, 2005) using methods developed several years earlier (Hagen et al., 1998). The model used for the reliability study included data from 1929 through 2002 (the period of simulation in the model). For the work reported here, the dataset was extended to cover 1914-2003 using the same methods of the 1998 ICPRB report. The data are based on different USGS gages for different parts of the record. The raw data were adjusted to account for the location, associated drainage area, and other characteristics of those gages. In addition to extending the record and making the necessary adjustments, the data were aggregated into monthly totals for this study. Figure 1 shows the hydrograph and some basic statistics of the Patuxent Reservoir inflows used for this study.

It should also be noted that the two Patuxent River Reservoirs are treated as one reservoir. The two are in series on the river, separated by only 3 to 4 hours travel time, and are functionally operated as one storage unit.

Figure 1: Hydrograph and basic stats for Monthly Patuxent Reservoir Inflows



Initial Data Analysis

The first step in developing an ARIMA model for the Patuxent inflows was to examine the basic structure of the distribution of flows. Most hydrologic datasets are highly skewed and this is the case for the Patuxent inflows (see Figure 2). Most ARIMA methods require the data to be approximately normally distributed, so the Patuxent data were log-transformed. The transformed dataset is approximately normally distributed (see Figure 3) though it does not pass the Kolmogorov-Smirnov goodness-of-fit test (Salas et al, 1980). The distribution of the log-transformed data was assumed to be close enough to normal for purposes of this analysis.

Once the data have been sufficiently normalized, seasonal patterns must be examined. The Patuxent inflows show a strong seasonal pattern, as demonstrated in graphs of the monthly statistics (figure 4). Seasonal patterns can dominate the ARIMA analysis, so the data must be made stationary (i.e., must be transformed so each month can be described with the same distribution). This was done by converting the data to unit normal data according to the following equation:

$$Z(t) = \frac{Y(t) - \check{u}(\tau)}{\check{s}(\tau)}$$

where $Y(t)$ = Log-transformed flow in timestep t
τ = month of the year (1 through 12)
$\check{u}(\tau)$ = monthly mean of the log-transformed inflows for month τ
$\check{s}(\tau)$ = monthly standard deviation of log-transformed inflow for month τ.

The above conversion results in an approximately normally distributed dataset (Z[t]) with mean zero and standard deviation equal to one (i.e., a dataset with a unit normal distribution). This Z(t) time series was used for building the ARIMA model.

Figure 2: Histogram of Monthly Patuxent Inflows

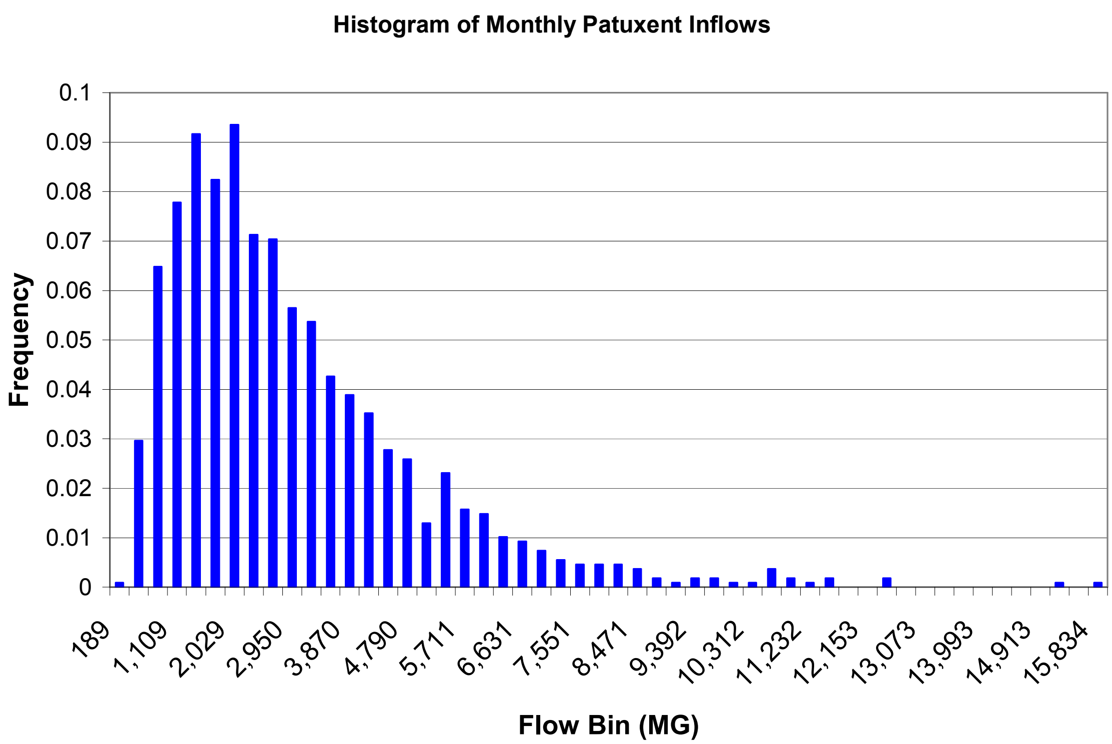


Figure 3: Histogram of Log-transformed Patuxent Inflows

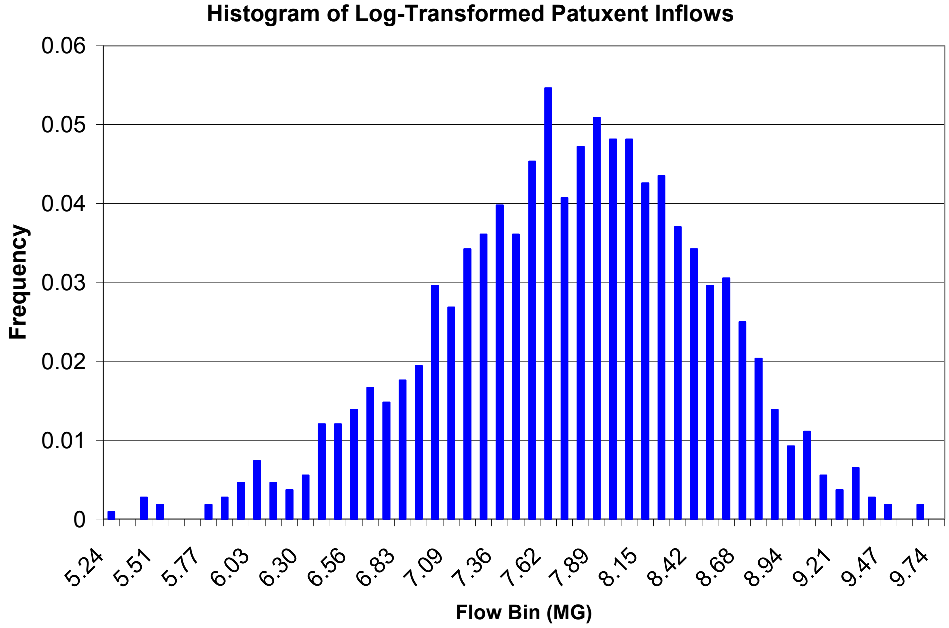
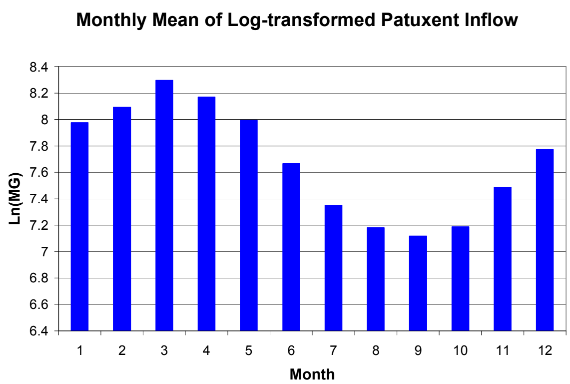
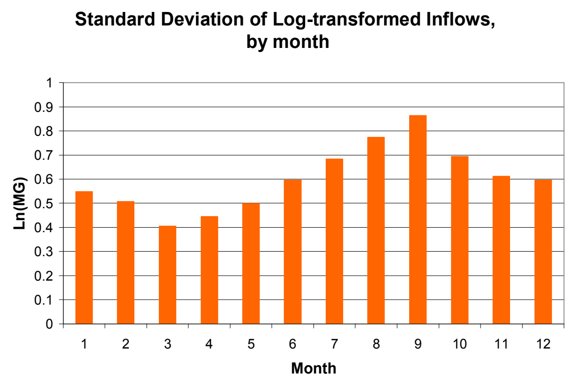


Figure 4: Monthly statistics





Log-transformation and the factoring out of seasonal patterns results in a stationary and approximately normally distributed time series. Autocorrelations in this transformed and normalized time series can be modeled by an ARIMA process. The autocorrelation and partial autocorrelation functions (ACF and PACF, respectively) of this time series can suggest the type of ARIMA model that should be used (Salas et al, 1980). Figure 5 shows the ACF and Figure 6 shows the PACF of the Z(t) time series.

Figure 5: ACF of Z(t)

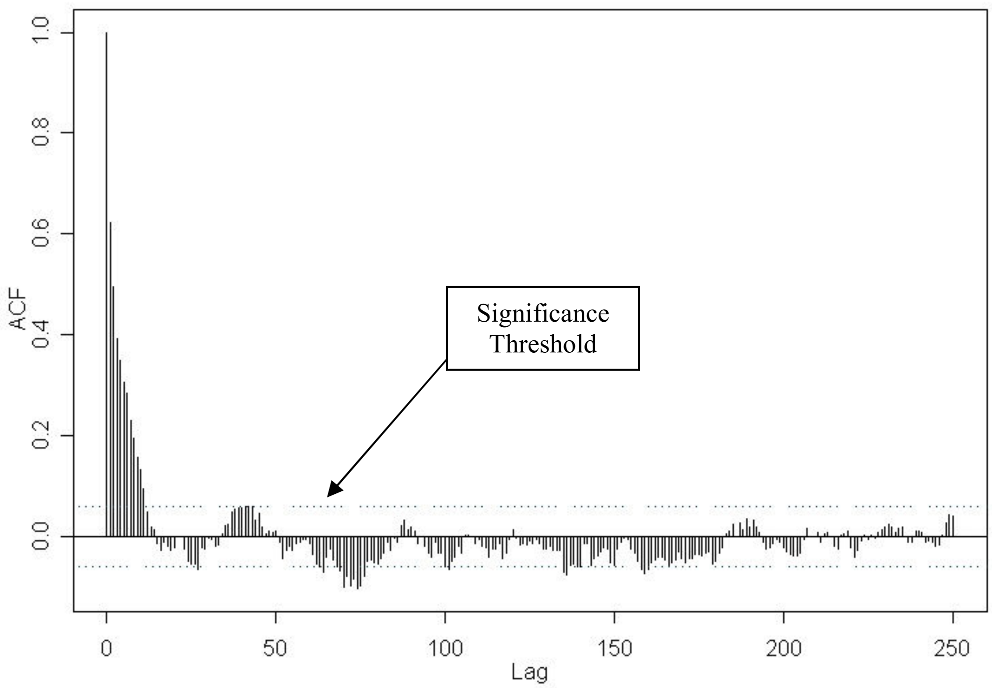
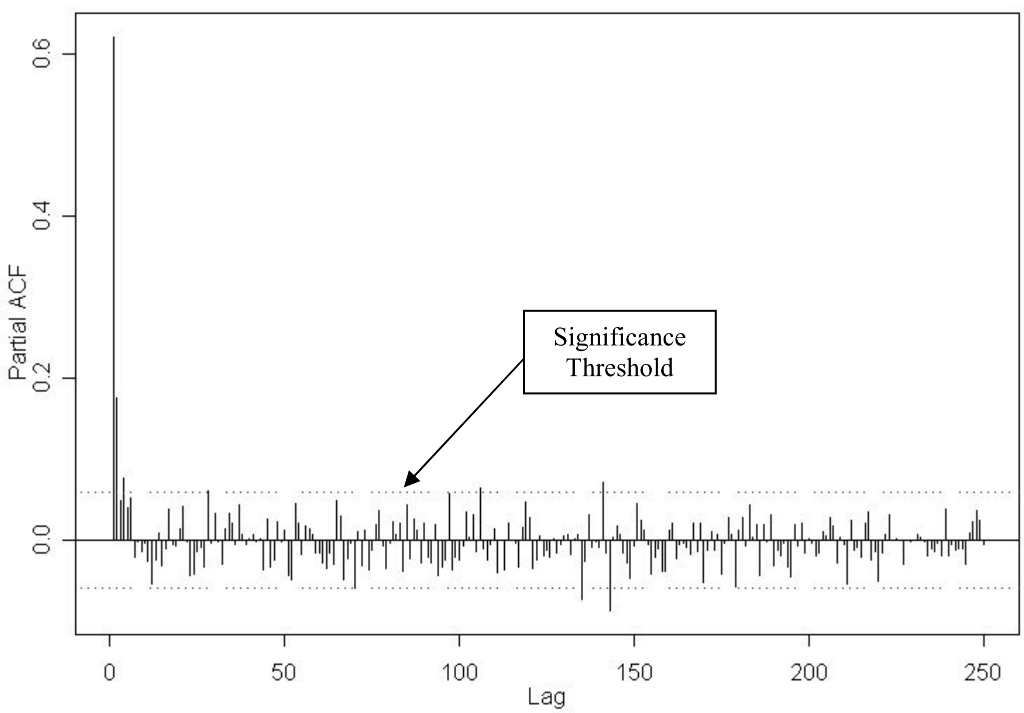


Figure 6: PACF of Z(t)



The ACF of Z(t) shows that there is a strong and significant autocorrelation for the first 11 lags, with the rapid decay in the strength of the correlation. No other lags show a strong correlation and most of them are insignificant (i.e., they fall between the dotted lines). The PACF shows a significant partial autocorrelation for lags 1 and 2, while all others are insignificant. The ACF and PACF can be used to guide the search for an adequate model. Following recommendations in the literature (e.g., Box et al, 1994), the ACF and PACF for Z(t) point to an ARIMA model with 2 autoregressive terms and no moving average component (i.e., ARIMA (2,0,0)[1]. However, ACF and PACF merely suggest a place to start and it is always good to calibrate and compare several models. Therefore, several models similar to ARIMA(2,0,0) are included in the analysis below. Several models, including the ARIMA(2,0,0) model, were calibrated and tested in order to find the form bested suited for Patuxent reservoir operations. Several standard tests were used to compare the models and eventually select one for implementation. The models tested are listed below, along with the tests used to compare them. Details on the characteristics of these models and the listed tests can be found in Box et al (1994) and Salas et al (1980).

Models Tested

- ARIMA(1,0,0)
- ARIMA(2,0,0)
- ARIMA(3,0,0)
- ARIMA(1,0,1)
- ARIMA(2,0,1)
- ARIMA(3,0,1)
- ARIMA(2,0,2)

Comparison Tests

- Akaike Information Criterion (AIC)
- Porte Manteau lack of fit test
- Root mean square error of forecasts of differing time horizons
- Comparison of mean and standard deviation between recorded dataset and stochastically generated datasets

[1] (The standard notation used with ARIMA modeling is as follows: the first term indicates the number of autoregressive terms, the second, whether there is differencing used, and the third whether there is a moving average component. For this series, the need for differencing was eliminated through the standardization process described above.)

Model Calibration

The general form for an ARIMA model is shown below.

$$Z(t) = \sum_p \varphi_p Z(t-p) + \varepsilon(t) - \sum_q \theta_q \varepsilon(t-q)$$

where *$Z(t)$ is the transformed, standardized flow at time t*
$Z(t-p)$ are values of $Z(t)$ lagged p time steps
φ_p are the autoregressive coefficients
$\varepsilon(t)$ are the error or residual terms
θ_q are the moving average coefficients
$\varepsilon(t-q)$ error terms lagged p timesteps

The statistical software package S-Plus was used to compute the parameters for each of the models listed above. Output from S-Plus includes values for the model parameters, the AIC for each model and the model residuals. Table 1 shows the parameters and AIC for each of the models listed above.

Table 1: Coefficients and AIC for Several Potential ARIMA models

Model	φ_1	φ_2	φ_3	θ_1	θ_2	AIC
ARIMA(1,0,0)	.62552	n/a	n/a	n/a	n/a	2521.42
ARIMA(2,0,0)	0.51405	0.17961	n/a	n/a	n/a	2485.84
ARIMA(3,0,0)	0.50537	0.15356	0.0508	n/a	n/a	2483.64
ARIMA(1,0,1)	0.81674	n/a	n/a	0.32903	n/a	2482.13
ARIMA(2,0,1)	1.04659	-0.15928	n/a	0.54872	n/a	2479.45
ARIMA(3,0,1)	1.12258	-0.1785	0.036	0.62765	n/a	2478.96
ARIMA(2,0,2)	1.15983	-0.25461	n/a	0.66235	-0.04407	2481.28

So, for example, the ARIMA(2,0,1) model is

$$Z(t) = 1.04659 * Z(t-1) - 0.15928 * Z(t-2) + \varepsilon(t) - 0.54872 * \varepsilon(t-1).$$

Based on the AIC it appears that ARIMA(3,0,1) is the best model (the lower the AIC the better). However, there are only small differences between the AIC values for most of the models. The ARIMA(1,0,0) is clearly inferior to the others based on the AIC. In addition, the ARIMA(2,0,2) has a higher AIC than the ARIMA(2,0,1) so the additional moving average term does not capture additional information. Therefore, the ARIMA(2,0,2) model is eliminated from further consideration.

Testing Model Adequacy

The Porte-Manteau test was applied to all of these models. The Porte-Manteau test involves calculating the "Q statistic," which is a function of the autocorrelation of the residuals (or error terms) of the model. If there is a substantial autocorrelation in the residuals of a model, then the model is deemed inadequate because there is more information that could potentially be captured by a higher order model. If there is no autocorrelation in the residuals, then the residuals represent a random, so-called white noise process containing no additional information and the model is deemed adequate. If the residuals are truly white noise, then the Q statistic can be described by a chi-squared distribution (Salas et al, 1980). Therefore, this test uses the chi-squared distribution with L-p-q degrees of freedom, where L is the maximum lag considered for the residual autocorrelations, p is the autoregressive order of the model, and q is the moving average order of the model. If Q is greater than the p=.05 critical value from the chi-squared distribution, then there is less than a 5% chance that we would obtain the observed Q if the residuals were actually white noise. This is the standard threshold for accepting/rejecting the adequacy of the model—if there is less than a 5% chance of obtaining the observed Q from an actual white noise process, then there is too much autocorrelation left in the residuals and a higher order model should be used. If Q is less than the p=.05 chi-squared critical value, then the model is deemed adequate. Table 2 summarizes the Porte-Manteau results for several models.

Table 2: Porte-Manteau results for several ARIMA models

Model	Q Statistic	Chi-squared Critical Value	Pass/Fail
ARIMA(1,0,0)	50.66	16.92	Fail
ARIMA(2,0,0)	14.54	15.51	Pass
ARIMA(3,0,0)	13.11	14.07	Pass
ARIMA(1,0,1)	10.14	15.51	Pass
ARIMA(2,0,1)	6.84	14.07	Pass
ARIMA(3,0,1)	5.50	12.59	Pass

The ARIMA(1,0,0) model fails the Porte-Manteau test, so it is eliminated from further evaluation. The other models pass this test, with the higher order models showing better results.

Testing Model Accuracy

The simplest way to test model accuracy is to compare the statistics from synthetic datasets generated using the ARIMA models and the statistics from the calibration sample of the historical recorded dataset. When used stochastically to generate synthetic flow data, ARIMA models should recreate basic statistics like the mean and standard deviation. Table 3 shows the results of this comparison for the four remaining models.

Table 3: Statistics for synthetic datasets (x years of data) vs historical data

Model	Mean flow	% difference	St Dev Flow	% difference
Recorded Dataset (1914-2003)	*2804.32*	*n/a*	*2044.40*	*n/a*
ARIMA(2,0,0)	2760.3	-1.57%	1919.3	-6.12%
ARIMA(1,0,1)	2777.8	-0.95%	1963.7	-3.95%
ARIMA(2,0,1)	2857.9	1.91%	1992.8	-2.52%
ARIMA(3,0,1)	2776.2	-1.00%	1942.1	-5.00%

The mean and standard deviation for each model are within 5% of the values from the historical record, with the exception of the standard deviation for the ARIMA(2,0,0) model, which is 6.12% less than the historical standard deviation. Each of these models reproduce the historical statistics acceptably well.

In addition to summary statistics such as the overall mean and standard deviation, it is useful to compare the seasonal patterns of the generated data with that of the recorded dataset. Figures 7 and 8 show charts of monthly means and standard deviations for the ARIMA models and the recorded dataset. Figures 9 and 10 show errors in the monthly statistics for the ARIMA models as compared to the monthly statistics from the recorded dataset.

Figures 7 and 9 show that each of the models reproduce the seasonal mean flows of the historic data set quite well. The ARIMA(2,0,0) model tends to underestimate flows, while the ARIMA(2,0,1) model tends to overestimate flows. The other models are more mixed between overestimating and underestimating monthly means. The errors are quite small, however.

The results for the seasonal standard deviations are mixed. Each of the models tends to underestimate the standard deviations in most months. The ARIMA(2,0,0) model results in many of the largest errors, including underestimates of nearly 25 percent in September and October.

Figure 7: Seasonal means for Patuxent models

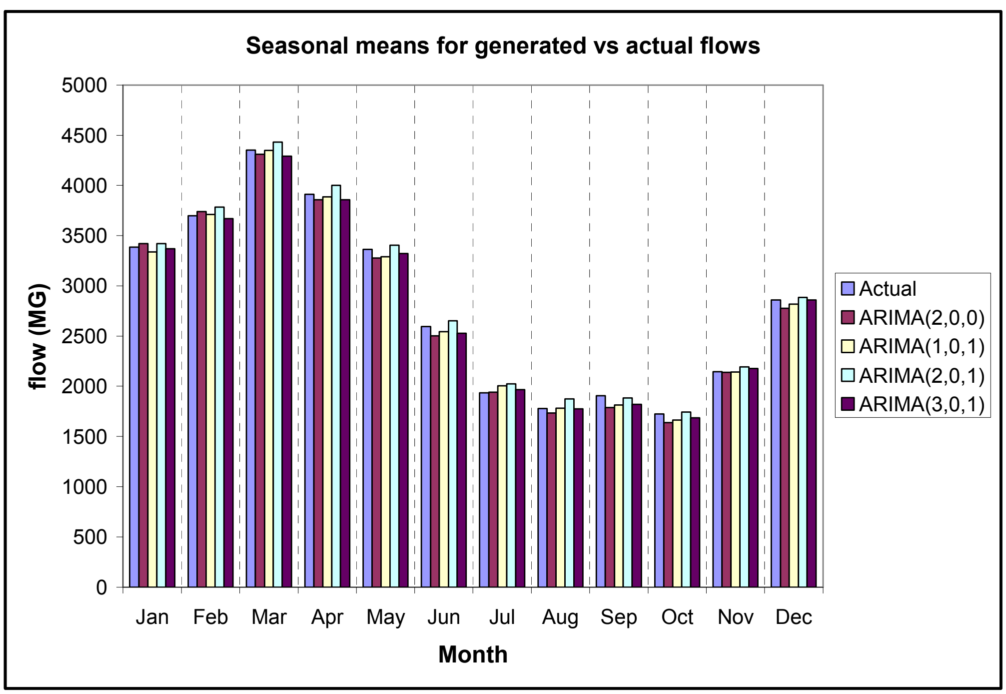


Figure 8: Season standard deviations for Patuxent models

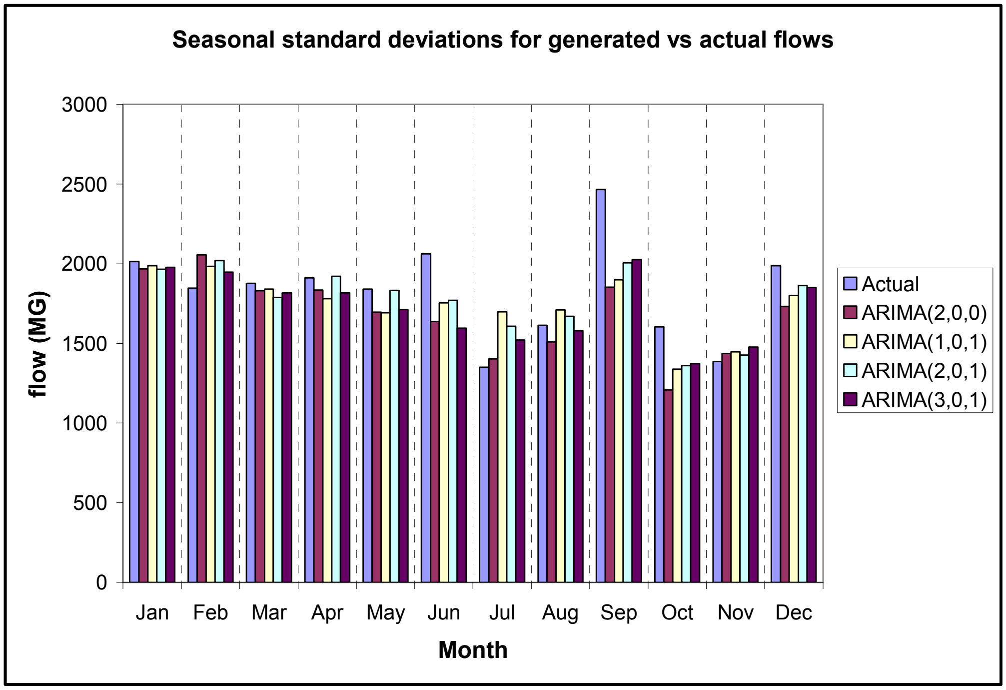


Figure 9: Error in the monthly means for Patuxent models

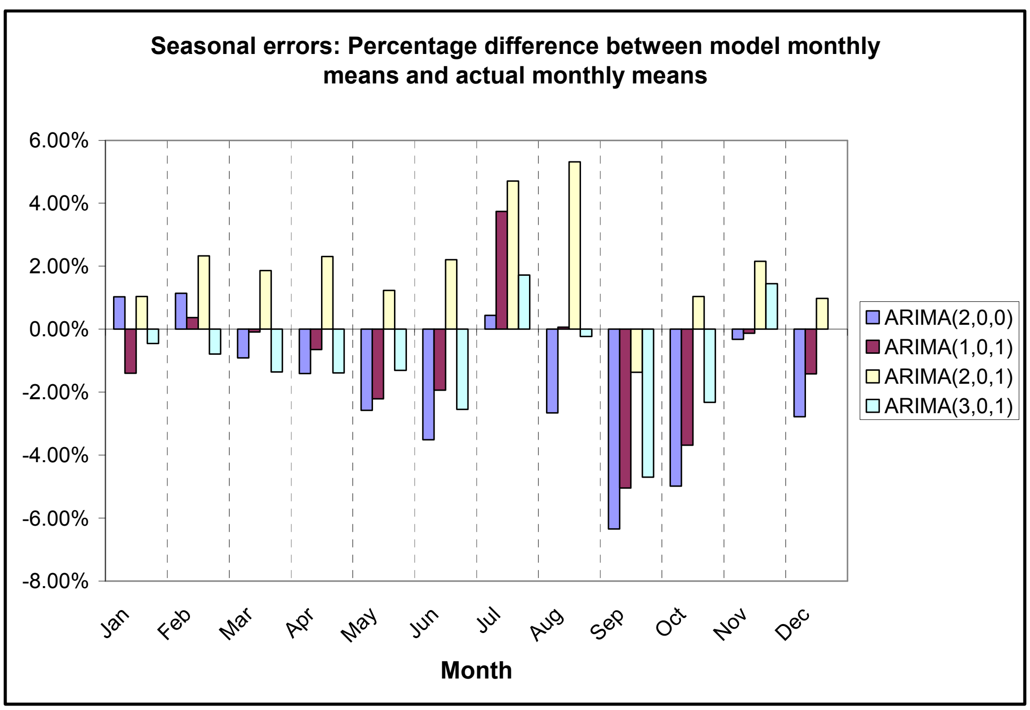
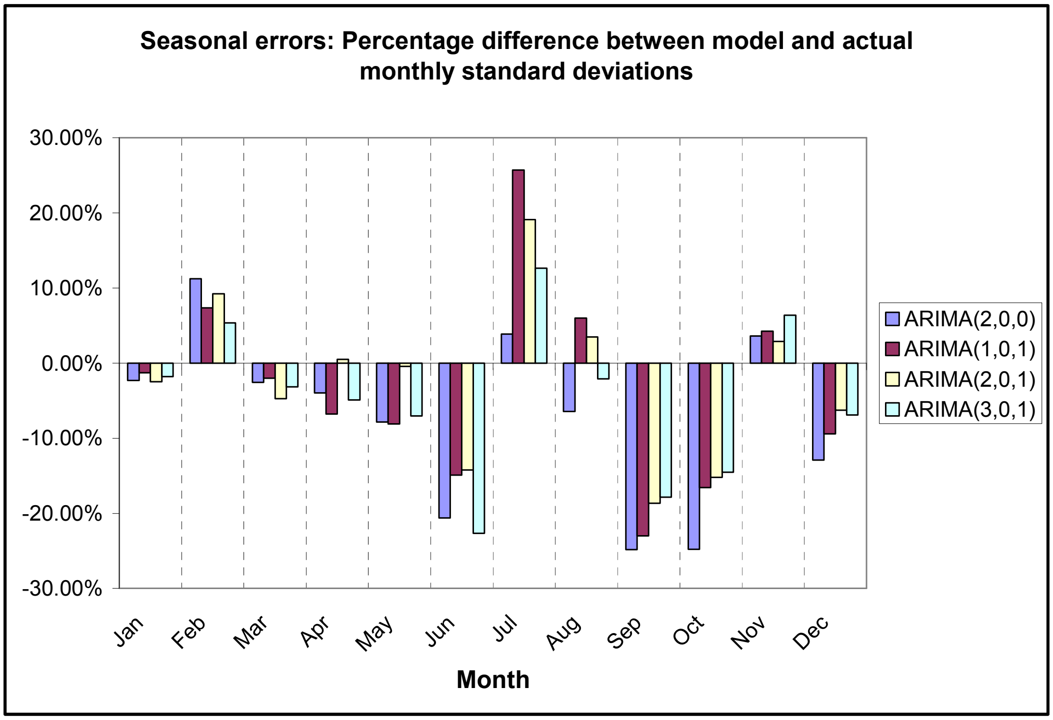


Figure 10: Error in the monthly standard deviation for Patuxent models



Another way to test model accuracy is by measuring forecast error. The models can be used to forecast flows any number of timesteps into the future. This was done with the four remaining models in three different ways. First, each model was used to generate one-month ahead forecasts through the portion of the historical record that was used to calibrate the models (i.e., 1914-2003). Second, also with the data used for calibration, each model was used to generate forecasts for 12 months out. Finally, each model was used to generate one-month ahead forecasts for the portion of the historical record not used in calibration (Jan. 2004-Aug 2006). Table 4 shows the resulting error rates for each type of forecast and each model. For comparison, Table 4 also include errors for a naïve forecast model, which assumes historical average conditions for each month as the forecast.

Table 4: Forecast errors for four ARIMA Models

Model	1-month forecasts1914-2003, RMSE*	12-month forecasts 1914-2003, RMSE	1-month forecasts 2004-2006, RMSE
ARIMA(2,0,0)	1576.8	1823.2	1899.6
ARIMA(1,0,1)	1572.7	1815.1	2208.3
ARIMA(2,0,1)	1569.9	1812.2	1906.0
ARIMA(3,0,1)	1569.8	1812.7	1902.7
Naive Forecast	1843.6	1845.9	1650.1

*Root Mean Squared Error

The differences in error rates between the models are quite small. The ARIMA(2,0,0) results in the highest RMSE when used with the calibration data, but is actually the best model for forecasting recent flows (2004-2006). The ARIMA(2,0,1) shows the opposite pattern—lower error rates for the calibration data, but higher errors for recent data. In general, the differences between the models are small enough that this test is not very useful for choosing one over another.

Relative error is also important for determining the accuracy of a model. While the differences between these models are small, their error rates appear to be significant compared to average flows. The mean absolute percentage error (MAPE) was calculated for each of the types of forecasts listed above. MAPE is the average of the absolute values of the relative error of each forecast expressed as a percentage (Makridakis et al, 2003). Table 5 shows the MAPE for each model and forecast. In addition, Table 5 shows the forecast accuracy of the naïve model for comparison. .

Table 5: MAPE for four ARIMA Models

Model	1-month forecasts 1914-2003, MAPE	12-month forecasts 1914-2003, MAPE	1-month forecasts 2004-2006, MAPE
ARIMA(2,0,0)	36.1%	48.5%	43.5%
ARIMA(1,0,1)	36.0%	48.2%	47.6
ARIMA(2,0,1)	36.0%	48.1%	40.8%
ARIMA(3,0,1)	35.9%	48.1%	43.3%
Naïve model	70.1%	69.9%	-4.4%

*Mean Absolute-Value Percentage Error

As with RMSE, the differences between the models are small, but the results indicate that all of these models will include a significant degree of error when used for forecasting. For example, in a 12 month forecast, the average error for the 12 forecasted flows is nearly 50%. These are significant error rates. However, as will be described later in this report, the model selected for implementation will not be used for generating deterministic forecasts of exact streamflows into the future. Rather, it will be used stochastically to generate many potential streamflow sequences and the resulting statistics used to inform the risk analysis of various operational alternatives. The ARIMA models show a significant improvement over the naïve model in forecasting future flows and they adequately reproduce seasonal flow statistics when used to generate synthetic data.

Model Selection

Based on the results presented here, either the ARIMA(2,0,1) model or the ARIMA(3,0,1) model could reasonably be selected for implementation. The test results presented above show that the differences between these two models is small. Usually, the simpler model would be selected when performance is equivalent, which, in this case, would lead to implementation of the ARIMA(2,0,1) model. However, analysis of synthetic data generated with the two models shows that the ARIMA(3,0,1) model tends to underestimate flows while the ARIMA(2,0,1) tends to overestimate flows. Underestimating flows when planning for drought preparedness and management represents a slightly more conservative approach. Therefore, the ARIMA(3,0,1) is selected for implementation.

Developing a New Refill Forecasting Tool

Developing an ARIMA model for inflows to the Patuxent reservoirs was the first step in developing a new reservoir refill forecasting tool. The ARIMA model helps predict potential inflows over the forecast horizon, but it must be combined with planned withdrawals in order to evaluate the risk of not meeting storage targets. The complete tool should provide an intuitive way to evaluate and visualize the impact of different withdrawal patterns. Excel is used as the platform for the refill forecasting tool, since most water managers are familiar with Excel and since an intuitive user interface could be developed.

The following sections describe how ARIMA forecasts are incorporated into the spreadsheet tool, discuss how the model must be initialized with recent streamflow data, show how the tool can be used to evaluate various withdrawal alternatives, introduce the new spreadsheet tool and interface, and discuss next steps and conclusions.

Using the ARIMA Model

The model is initialized with monthly inflows over the last 12 months. Flows for these last 12 months are known, either from USGS data downloaded automatically by the spreadsheet or from other data as preferred by the user (see description of initialization data below). This 12-month initialization is important for two reasons. First, the purpose of using an ARIMA model for predicting reservoir refill is to forecast potential inflows given current and recent conditions. Initialization provides the model with data describing current and recent conditions.

Second, some period is selected for the initialization (e.g., 12 months back as in this application) and values prior to that period are unknown to the model and must take on assumed values, usually average flows for those months. The longer the initialization, the less the impact assumed initial values have on the results. Both 12-month and 24-month initialization periods were tested for this model. . Comparison of the two initialization periods showed very small differences between the two. Predictions for reservoir refill were almost exactly the same. Predictions of potential inflows differed by less than one percent, with the 12-month initialization consistently resulting in slightly lower predicted flows. Therefore, the 12-month initialization is sufficient and will be used for this model.

Once the model is initialized, it can be used to calculate an ensemble of potential inflows. The initialized Patuxent inflow ARIMA model is used to stochastically generate 1,000 independent inflow sequences over the next 12 months. The term $\varepsilon(t)$ in the ARIMA(3,0,1) equation below represents the random component in the ARIMA process and it is equivalent to the residuals from the fitting process.

$$Z(t) = 1.12258 * Z(t-1) - 0..1785 * Z(t-2) + .036 * Z(t-3) + \varepsilon(t) - .62765 * \varepsilon(t-1).$$

The residuals from the fitting process were analyzed and shown to have a mean of 0.00303 and a standard deviation of 0.76220. Therefore, the stochastic generation of inflow sequences is done using the random number generator in Excel to create 1,000 12-month sequences of independent and normally distributed random numbers with mean and standard deviation equal to those of the residuals from the ARIMA fitting process. Since each random number sequence is different, they can be used in the equation above to produce inflow sequences that are distinct and independent.

Table 6 demonstrates a portion of this process, showing the calculation of two sets of potential inflows. The calculations demonstrated in Table 6 start from the current time step and do not show the initialization calculations. The flows from the three months immediately prior to the start of the calculations shown in the table are known values (3302, 4210 and 2907 MGD). The values for ε(t) are generated normal random numbers with mean 0.00303 and standard deviation 0.76220. Timestep one is equal to the first forecast month. In time step one, the value for ε(t-1) comes from the initialization process—it is the calculated error of a flow forecasted for time zero. Values of ε(t-1) in timesteps two through six are equal to the generated ε(t), but lagged by one time step. Each generated flow becomes an independent variable (i.e., a previous flow) in subsequent timesteps. Figure 11 shows a plot of five different inflow sets, including the two shown in Table 6.

Table 6: Sample of ARIMA(3,0,1) inflow generation

	T	**Flow at t-1 (MG)***	**Flow at t-2 (MG)***	**Flow at t-3 (MG)***	**ε(t), generated from Normal (0.00303, 0.76220)**	**ε(t-1)**	**Generated flow for time=t (MG)****
Inflow Set #1	1	3302***	4210***	2907***	-0.22581	0.253061†	4187.7
	2	4187.7	3302***	4210***	-0.97082	-0.22581	2604.6
	3	2604.6	4187.7	3302***	0.189203	-0.97082	2977.9
	4	2977.9	2604.6	4187.7	0.975958	0.189203	3857.7
	5	3857.7	2977.9	2604.6	0.916413	0.975958	4032.3
	6	4032.3	3857.7	2977.9	1.324024	0.916413	6864.7
Inflow Set #2	1	3302***	4210***	2907***	0.034845	0.253061†	4653.8
	2	4653.8	3302***	4210***	-0.49549	0.034845	3408.7
	3	3408.7	4653.8	3302***	-0.41587	-0.49549	2600.9
	4	2600.9	3408.7	4653.8	0.650382	-0.41587	3132.6
	5	3132.6	2600.9	3408.7	0.615111	0.650382	3028.6
	6	3028.6	3132.6	2600.9	0.35091	0.615111	2708.7

**Note that flows for t-1, t-2 and t-3 are log-transformed and standardized before using them in the ARIMA(3,0,1) equation.*

***Output of the equation is also standardized and log-transformed. Multiplying by the monthly standard deviation, adding the monthly mean, and taking an exponential results in the values shown in the table.*

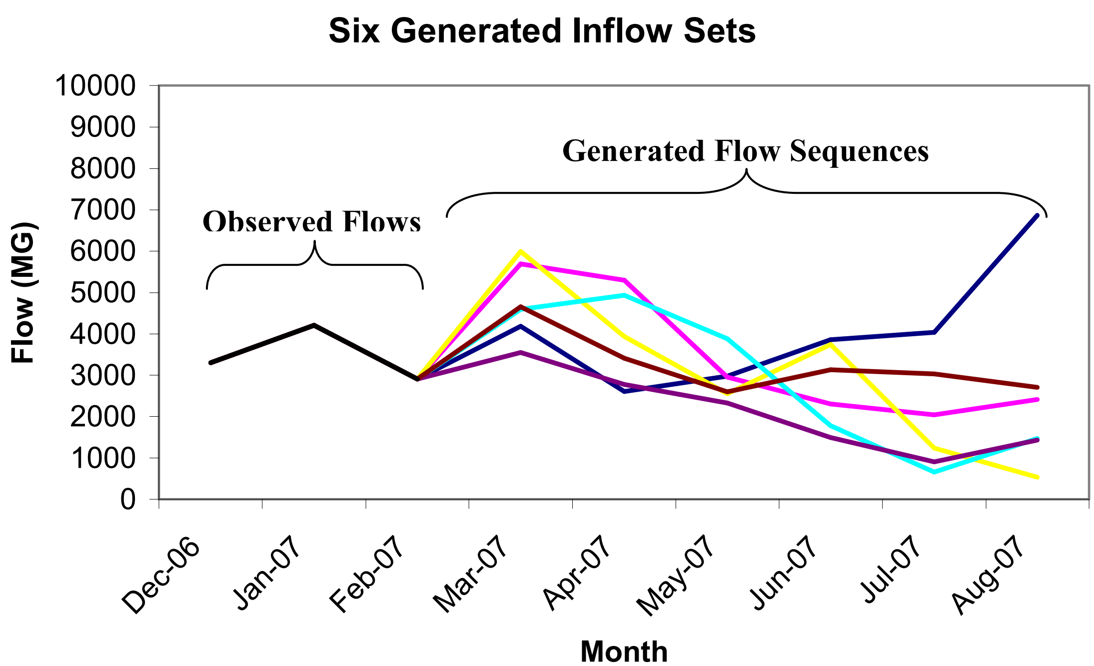
****Observed actual flows*

†*The first value for* **ε(t-1)** *comes from the initialization process.*

Initialization Data

As noted above, the model requires recent flows for initialization. The spreadsheet tool was designed to automatically download appropriate USGS data for the Patuxent River near Unity, MD. These daily data are adjusted to account for drainage area and other characteristics so that they better represent inflows to the Patuxent reservoirs. They are then aggregated into monthly values and displayed on the main interface of the spreadsheet tool. The user can elect to use these USGS-based recent inflows or inflows based on other information, such as WSSC's recorded reservoir inflows. If the user elects to use data other than the USGS-based data provided by the model, they can enter the new numbers in a new column and the model will automatically use those for computations.

Figure 11: Six Example Stochastic Inflow Sets, illustrating how the ARIMA model is used to generate an ensemble of potential future streamflow



Evaluating Withdrawal Alternatives

The spreadsheet tool allows the user to set withdrawals from the Patuxent reservoirs for each of the next 12 months. Withdrawals should be set in order to meet the operational criterion described in the introduction to this report—i.e., that the reservoir reach 90% full by June 1 95% of the time. Meeting that refill target 95% of the time is equivalent to meeting it with a 95% probability in any given year. The refill spreadsheet tool provides two ways to illustrate whether this criterion is met. One is that the tool shows a plot of forecasted storage through next May under an estimate of the five percent probability event. The other reports the percentage of inflow sets for which the storage target is reached given the withdrawals set by the user.

In the first case, the five percent probability event is identified according to reservoir storage at the end of the forecast horizon (i.e., end of May storage, which is equivalent to June 1). The inflow set that results in the 950th largest storage on June represents the fifth percentile event, or five percent probability event. That inflow set associated with fifth percentile storage can change as the user modifies the reservoir withdrawal pattern. This can lead to sudden shifts in the plot of storage as the users tries to identify a sustainable withdrawal pattern. In order to prevent this kind of shift, the tool uses an average of 15 traces as the five percent probability event. Specifically, the tool identifies the set of traces with resulting storage ranked 943 through 957 (i.e., the 950th, along with the seven ranked just above it and the seven just below it). The traces included in this set of 15 occasionally may change as the withdrawals are changed, the but the averaging acts as a smoothing process and prevents sudden shifts in the storage plot. This allows the tool to behave more intuitively while still providing a good estimate of the five percent probability event. The plot also shows storage under the median case and a similar process is used for those calculations.

In addition to plotting storage for the five percent probability event, the tool reports the percentage of years for which the 90% full storage criterion is met. As mentioned above, storage for the forecast horizon is calculated for each inflow set. This tool then calculates the number of years for which storage reaches or exceeds 90% of capacity by June 1. If storage reaches 90% full in 95% of the stochastic inflow sets, each of which is equally probable given current conditions, then there is a 95% probability of meeting the storage target by June 1.

The graphical interface

All these calculations and algorithms were brought together in a spreadsheet tool. The tool includes a graphical interface intended to make the tool more user-friendly (see Figure 12). The interface includes basic instructions with some color coding to point the user to the right places. There is also a link to a more detailed description of the model. The user inputs current storage in the reservoirs (as of the end of the previous month) and the minimum allowable release from the reservoirs to the lower Patuxent River. Recent inflows based on USGS data are shown on the bottom left. These USGS-based numbers are used by the model unless the user opts to enter other values in the adjacent column. Withdrawal rates can be typed in the green column manually, or alternatively they can be set by using the slider bars to the right of the green column. The model calculations are slow, so the automatic calculation feature in Excel is disabled. Therefore, after making all the desired changes for the various months, the user must click on the “Recalculate Storage” button to see the results. Finally, the model alerts the user if the USGS data are out of date. If so, clicking on the “Download New USGS Data” button automatically brings in updated information from the USGS website.

Next Steps

The ARIMA(3,0,1) fit to the Patuxent reservoir inflows may be improved in the future. As noted above, the distribution of the flow data is skewed and required transformation. The log-transform resulted in an approximately normal distribution as seen in Figure 3 above. However, according to the Kolmogorov-Smirnov test, the log-transformed data are not adequately described by a normal distribution. The type of ARIMA calibration described here requires the data to be normally distributed, so the fact that the transformed data failed a test for normality means that there may be some bias in the ARIMA model. This can be addressed in future efforts by applying other transformations to the flow dataset to create a normally distributed dataset. A different transformation was used for the Occoquan analysis (see below) because in that case the data were more severely skewed.

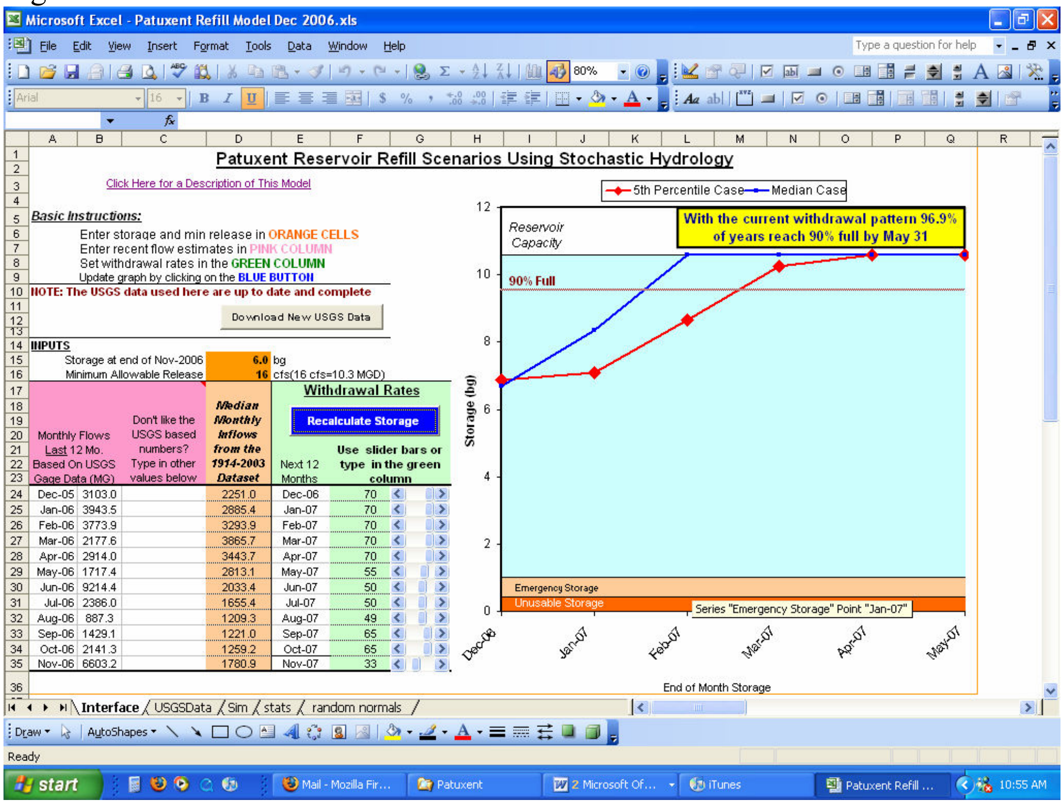
Some factors in the evaluation process can be explored further. For example, a maximum lag of 10 time steps was used in computing the Porte-Manteau Q-statistic. This was an arbitrary choice and other lags might produce different results. Also, additional analysis of forecast error might provide useful information for model selection. Additional analysis might include examining seasonal and longer patterns of forecast error. For example, it might be useful to know whether there are differences between the models in their ability to forecast flows during certain parts of the year or during particular dry or wet periods.

Conclusions

The incorporation of an ARIMA model to forecast probabilities of future Patuxent and Occoquan Reservoir (summarized in subsequent sections of this report) inflows provides a useful management tool for ICPRB CO-OP and the WMA water suppliers. The tool can be used to better understand the probabilities of reservoir refill given current conditions, and the consequences of various withdrawal alternatives. The tool is especially useful in dry periods, since autocorrelation in streamflow will result in lower estimates of future flow given current low flow conditions. Thus, the tool is appropriately conservative and better reflects future scenarios than does a traditional position analysis based on historical inflows alone. The tool is especially useful because the water manager can automatically update the inflows with use of a simple update button and test various withdrawal scenarios in the easy to use interface. The ARIMA forecasting takes place automatically in the background, and requires no specialized knowledge of ARIMA modeling.

The Patuxent and Occoquan reservoirs are an important part of the Washington D.C. area water supply system. The better water managers can predict refill of these reservoirs the more reliable the water supply for the region. Previous tools proved useful during drought exercises and for regular operations. The analysis described here and the tool developed based on that analysis should offer another improvement for regional reservoir water supply management.

Figure 12: Interface of the New Reservoir Refill Tool



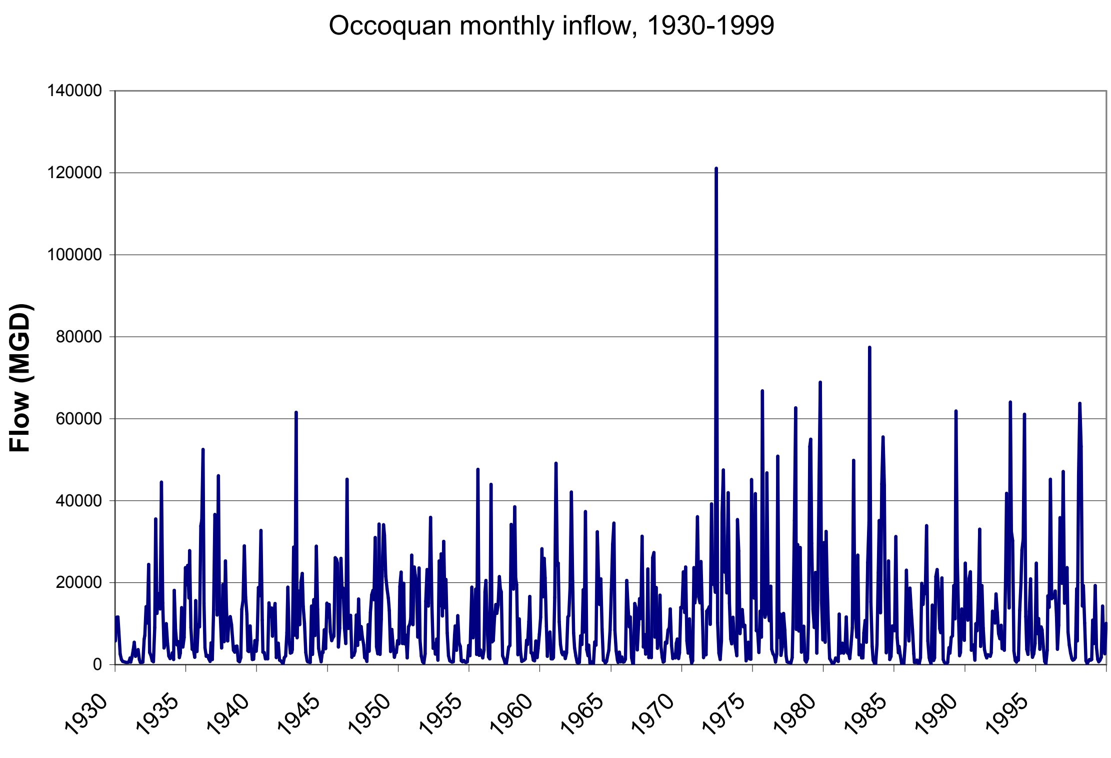
Occoquan Reservoir Refill

A similar tool was developed for Occoquan Reservoir. Occoquan's watershed is much bigger than the Patuxent's and the statistical properties of the inflow data are different. One significant difference in developing the Occoquan ARIMA model was that additional steps were needed to normalize the Occoquan data. The development of the Occoquan refill prediction tool is briefly described below.

Methods and Data

A dataset for Occoquan inflows was developed by ICPRB (Hagen and Steiner, 1998a) for modeling and planning purposes using similar methods to those described for the Patuxent Reservoir. The resulting dataset covers 1929 through 2002. Data for January 1, 1930 through December 31, 1999 were used for the work reported here. The dataset was extended though 2006, resulting in a seven-year dataset (2000-2006) for model verification. ICPRB's daily time series of inflows was aggregated to create a monthly time series of inflows. Figure 13 shows the monthly hydrograph for Occoquan inflows.

Figure 13: Occoquan inflow hydrograph



Normalization

As with the Patuxent data, Occoquan inflows are highly skewed (see figure 14). Unfortunately, log-transformation of Occoquan inflows does not produce a sufficiently normal distribution (see Figure 15). Log-transformation of Occoquan inflows results in a distribution with a skew of -0.48197 Other transformations were investigated in order to produce to a sufficiently normal distribution.

Figure 14: Histogram of monthly Occoquan inflows

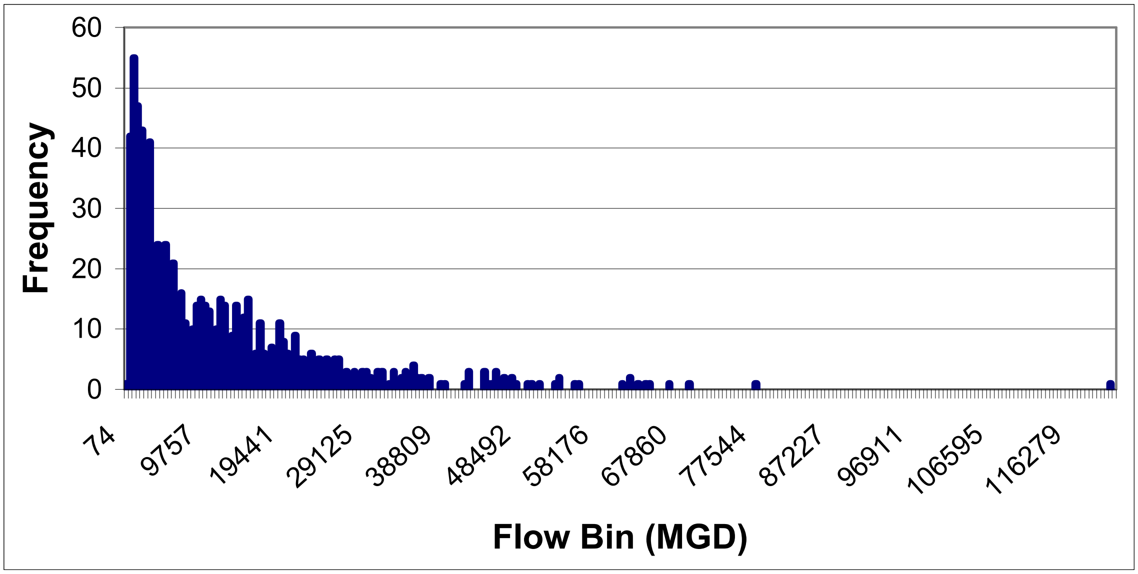
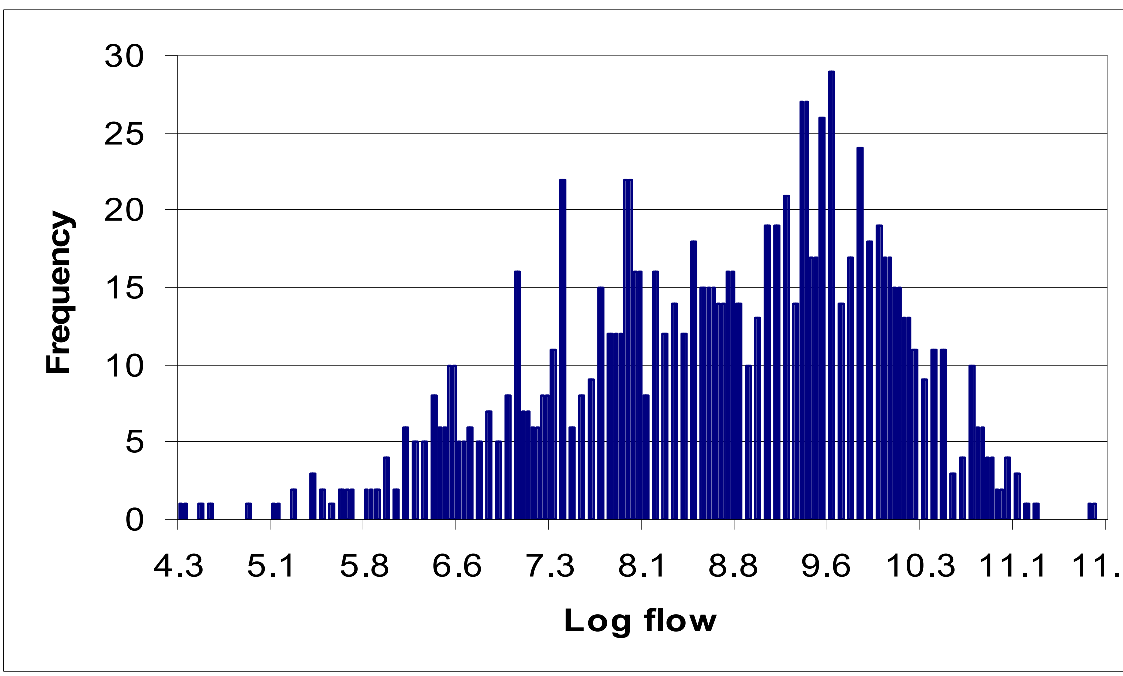


Figure 15: Skewed histogram of log-transformed Occoquan inflows



One common method is the Box-Cox transformation:

$$Y(t) = \frac{(F(t)^{\lambda} - 1)}{\lambda} \qquad \lambda \neq 0$$

where Y(t) is the Box-Cox transformed flow at time t
F(t) is the flow (untransformed) at time t
λ is the transformation parameter.

In the Box-Cox transformation, the transformation parameter λ should be set so that the resulting distribution is as close to normal as possible. A simple optimization model was used to minimize the skew of the transformed data by changing λ. The optimal value of λ was found to be .18019 resulting in a skew that is very nearly zero. Figure 16 shows the histogram of the transformed data with λ=.18019. The resulting distribution is approximately normal. The transformed data were standardized as described above to create a series of data with a unit normal distribution. The standardized data were used in the ARIMA fitting process.

The ACF and PACF of the transformed and standardized Occoquan inflow data are shown in figures 17 and 18. The Occoquan data show positive and significant autocorrelations for the first 6 lags (excluding lag 0), with a rapid decay in the strength of the correlations. The PACF shows significant correlations for only lags 1 and 2. The ACF and PACF suggest ARIMA(2,0,0) as a starting point for calibrating and comparing candidate models. Several models were calibrated and compared. The results are discussed below.

Figure 16: Histogram of transformed data with one value for λ

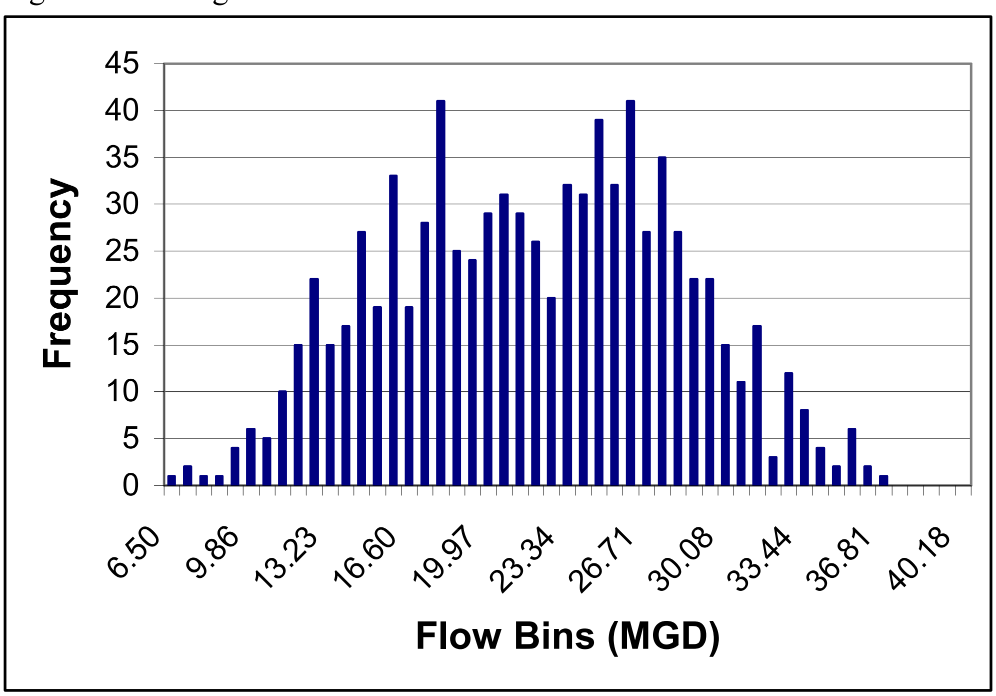


Figure 17: ACF of transformed and standardized Occoquan inflows

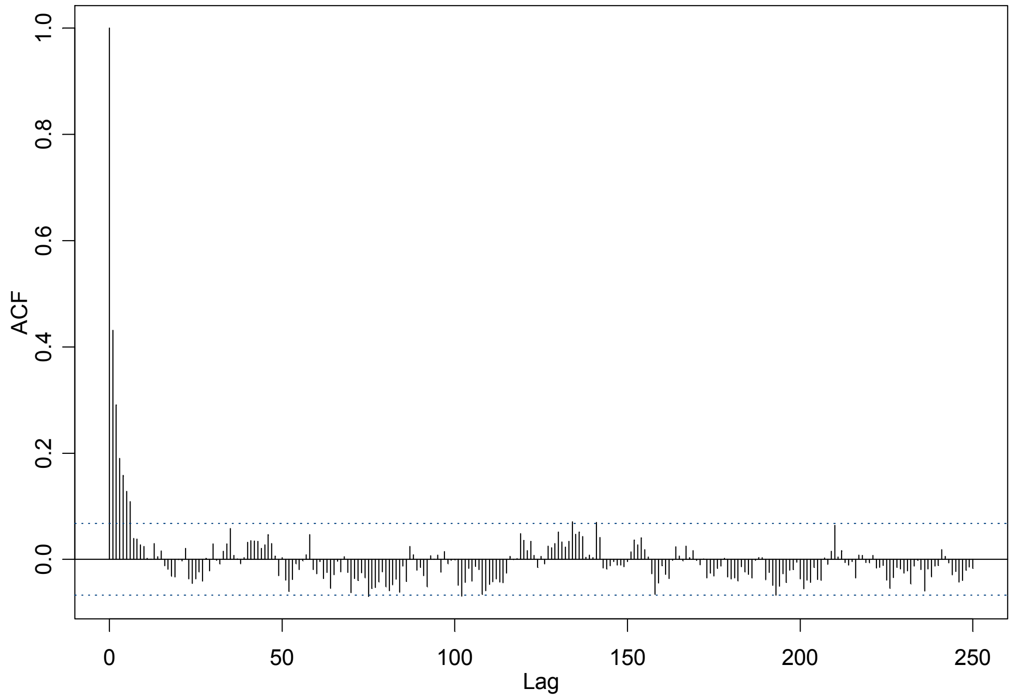
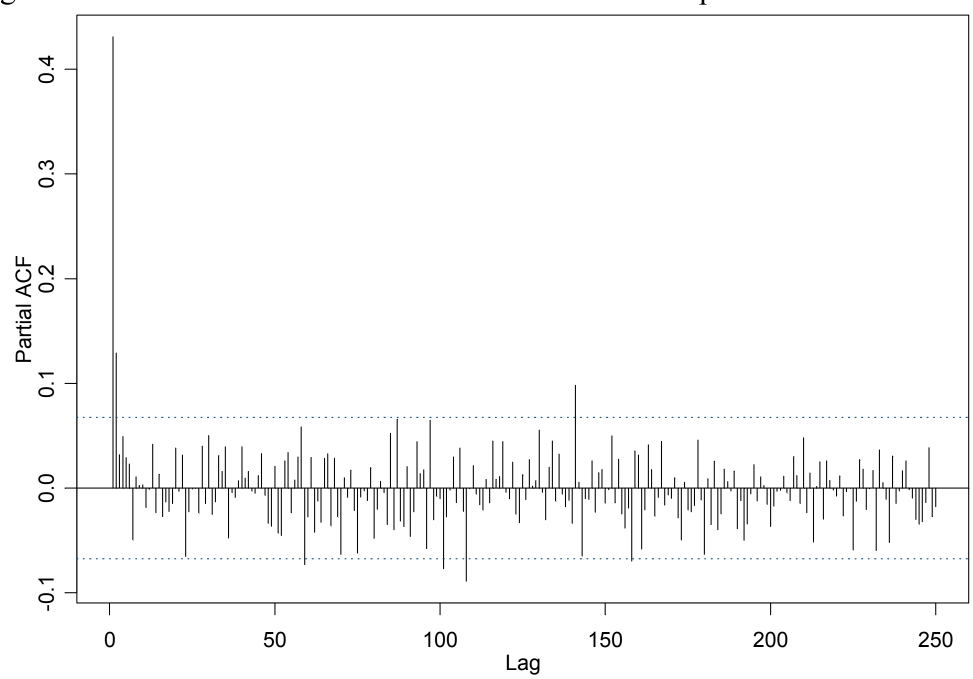


Figure 18: PACF of transformed and standardized Occoquan inflows



Model Calibration and Comparison

Several ARIMA models were calibrated and evaluated using the same methods that were used for the Patuxent ARIMA model. Tables 7 shows the coefficients and AIC values for six models calibrated for the Occoquan data. Table 8 shows the results of the Porte-Manteau test for the same models.

Table 7: Coefficients and AIC for Several Potential ARIMA models

Model	φ_1	φ_2	φ_3	θ_1	θ_2	AIC
ARIMA(1,0,0)	.43098	n/a	n/a	n/a	n/a	2198.44
ARIMA(2,0,0)	0.37540	0.12902	n/a	n/a	n/a	2184.74
ARIMA(3,0,0)	0.37134	0.11634	0.03206	n/a	n/a	2183.93
ARIMA(1,0,1)	0.68667	n/a	n/a	0.32235	n/a	2184.51
ARIMA(2,0,1)	0.84923	-0.08053	n/a	0.48096	n/a	2184.46
ARIMA(2,0,2)	1.60745	-0.62939	n/a	1.25044	-0.29168	2182.10

Table 8: Porte-Manteau results for several ARIMA models

Model	**Q Statistic**	**Chi-squared Critical Value**	**Pass/Fail**
ARIMA(1,0,0)	25.12	16.92	Fail
ARIMA(2,0,0)	6.16	15.51	Pass
ARIMA(3,0,0)	5.47	14.07	Pass
ARIMA(1,0,1)	4.43	15.51	Pass
ARIMA(2,0,1)	3.91	14.07	Pass
ARIMA(2,0,2)	4.30	12.59	Pass

ARIMA(1,0,0) is clearly inferior based on AIC and Porte-Manteau so it is eliminated from further evaluation.

Table 9 shows statistics for synthetic flows that were stochastically generated using the five models still under consideration. The statistics for two of the models—ARIMA(2,0,0) and ARIMA(2,0,1)—deviate from the statistics of the recorded dataset considerably more than the other models. Since the model will be used to stochastically generate streamflow sequences, the results shown in Table 9 suggest that the ARIMA(2,0,0) and ARIMA(2,0,1) models may be inadequate for the refill forecasting tool.

In addition to summary statistics, the seasonal patterns of the generated flows were compared to the seasonal patterns of the recorded dataset. Figures 19-22 show charts of the monthly means, standard deviations and errors. The figures show that all the models deviate from actual flow patterns at various times. The ARIMA(2,0,1) model tends to underestimate mean flows through a greater portion of the year and to a greater degree than the other models. With only a few exceptions, the ARIMA(2,0,2) model tends to have smaller errors than the other models for both monthly means and standard deviations. What is most clear from Figures 20 and 21 is that Occoquan inflows are more difficult to model as an ARIMA process than Patuxent inflows—the errors for the Occoquan models are generally much larger than the errors for the Patuxent models.

Table 9: Statistics for synthetic datasets vs historical

Model	Mean flow	% difference	St Dev Flow	% difference
Recorded Dataset (1930-1999)	*11386.6*	*n/a*	*12923.2*	*n/a*
ARIMA(2,0,0)	12778.75	12.23%	14988.38	15.98%
ARIMA(3,0,0)	11964.73	4.52%	13580.5	5.09%
ARIMA(1,0,1)	12412.21	8.57%	12976.65	0.41%
ARIMA(2,0,1)	10019.98	-11.01%	11262.58	-12.85%
ARIMA(2,0,2)	11204.34	-1.82%	12089.41	-6.45%

Figure 19: Monthly mean flows

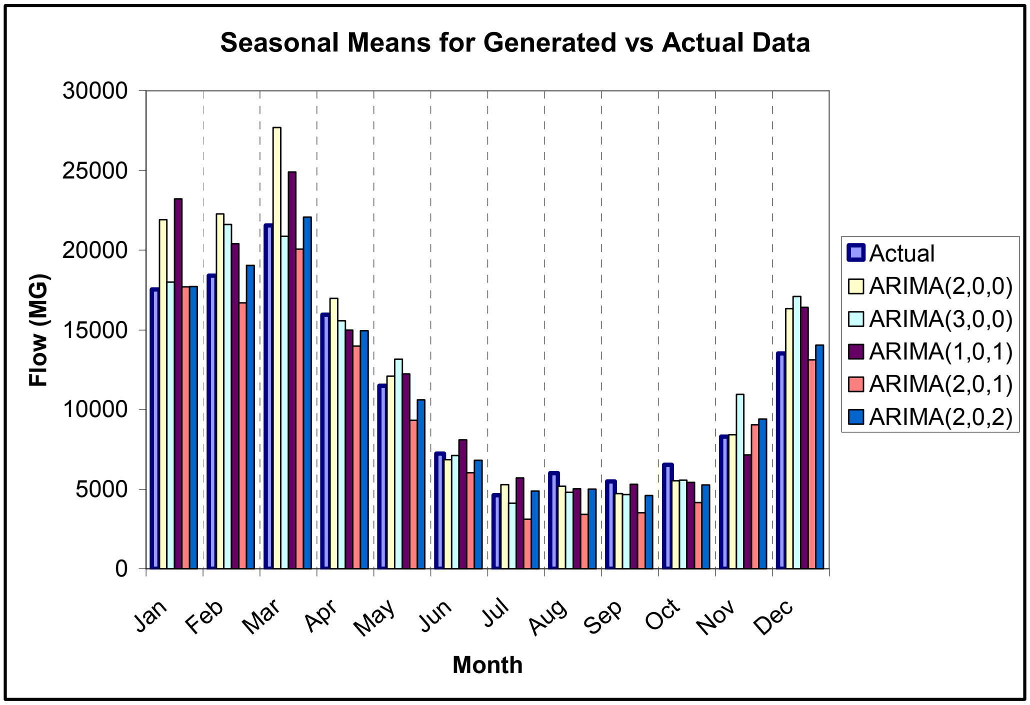


Figure 20: Monthly standard deviations

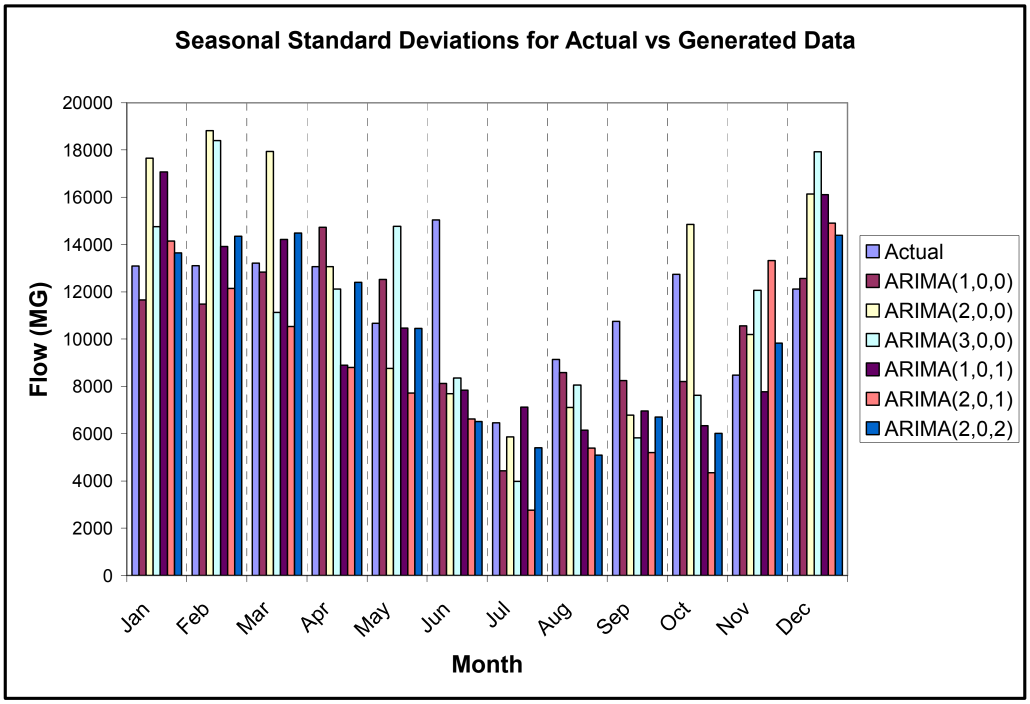


Figure 21: ARIMA Model Errors: Monthly mean flows

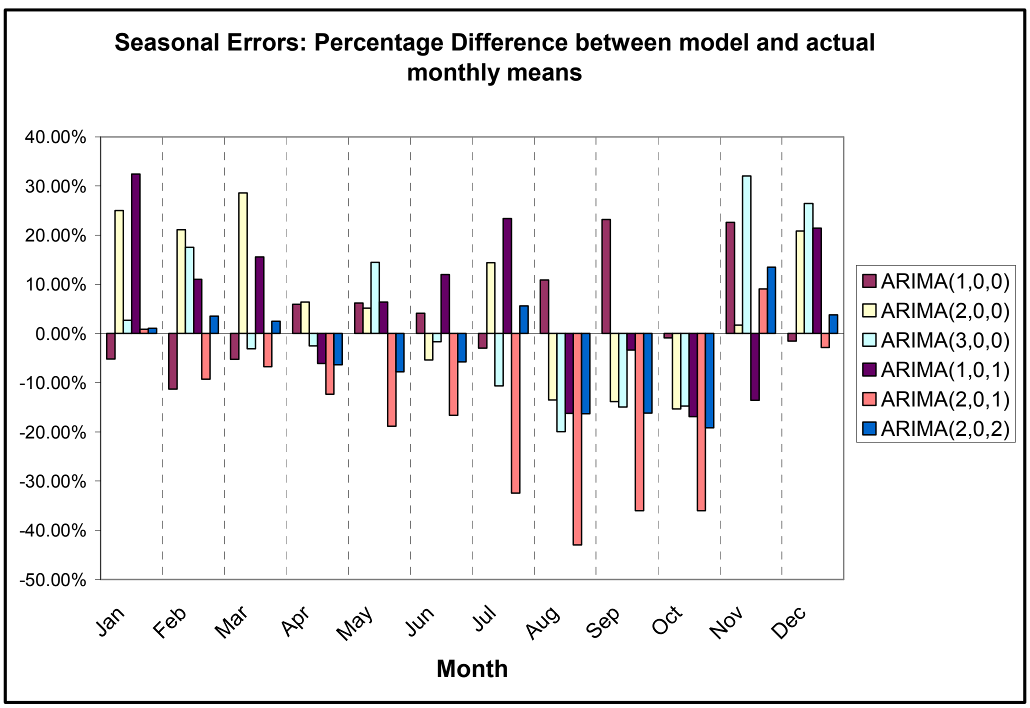
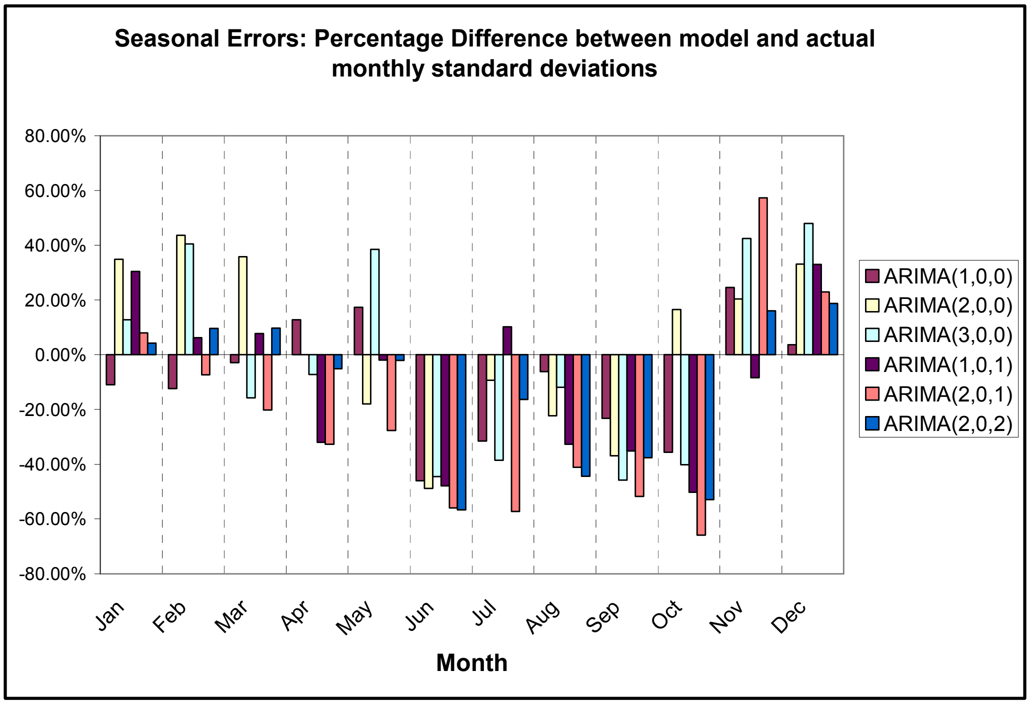


Figure 22: ARIMA Model Errors: Monthly Standard Deviations



The three models with favorable results in Table 9 were tested for forecast error. Table 10 shows 1-month and 12-month RMSE using the calibration data (1930-1999) and 1-month RMSE for 2000-2006 flows. Table 11 shows mean absolute percentage errors for the same three types of forecasts. Both Tables 10 and 11 include forecast errors for the naïve model for comparison.

Table 10: Forecast errors for four ARIMA Models

Model	1-month forecasts, 1930-1999, RMSE*	12-month forecasts, 1930-1999, RMSE	1-month forecasts, 2000-2006, RMSE
ARIMA(3,0,0)	11532.2	12061.8	3946.1
ARIMA(1,0,1)	11423.1	12041.8	3977.5
ARIMA(2,0,2)	11444.9	12471.4	3866.6
Naive Model	11691.8	11686.6	12194.0

*Root Mean Squared Error

Table 11: MAPE for three ARIMA Models

Model	1-month forecasts, 1930-1999, MAPE*	12-month forecasts, 1930-1999, MAPE	1-month forecasts, 2000-2006, MAPE
ARIMA(3,0,0)	109%	143%	126%
ARIMA(1,0,1)	105%	142%	127%
ARIMA(2,0,2)	102%	160%	122%
Naïve Model	274%	275%	348%

*Mean Absolute-Value Percentage Error

The three ARIMA models result in similar forecasting errors. The only notable difference is that the ARIMA(2,0,2) model results in considerably higher errors (both RMSE and MAPE) for the 12-month forecast than the other two models. This complicates model choice for the Occoquan because the ARIMA(2,0,2) model produced more accurate synthetic data, especially in terms of monthly patterns.

Two other things are worth noting about the results shown in Tables 10 and 11. First, the MAPE values confirm the difficulty of modeling Occoquan inflows as an autoregressive process. These models result in MAPE values that are more than twice as large as the MAPE values for the Patuxent models. The primary reason is that the Occoquan is a "flashier" watershed, with a much higher ratio of standard deviation to mean (1.135 for the Occoquan and .729 for the Patuxent).

Second, the MAPE values for the naïve model appear to be inconsistent with the RMSE values. This is especially true for the 12-month forecast where he naïve model results in smaller RMSE values than the ARIMA models, but much larger MAPE values. This can be explained by analyzing the distribution of errors for the naïve model and one of the ARIMA models. Figure 23 shows the histogram of errors for the naïve model and figure 24 shows the histogram of errors for the ARIMA(2,0,2) model. The ARIMA(2,0,2) model results in a greater number of large errors, especially large negative errors. This will push up the RMSE value because the errors are squared. However, the errors for the ARIMA model are centered closer to zero than are the errors for the naïve model. This means that positive and negative errors will balance out more for the ARIMA model, reducing its MAPE value. Indeed, the median error for the naïve model is much larger (3,253 versus 11) and a far greater percentage of its errors are positive (67% versus 50%).

Figure 23: Histogram of errors for the naïve model

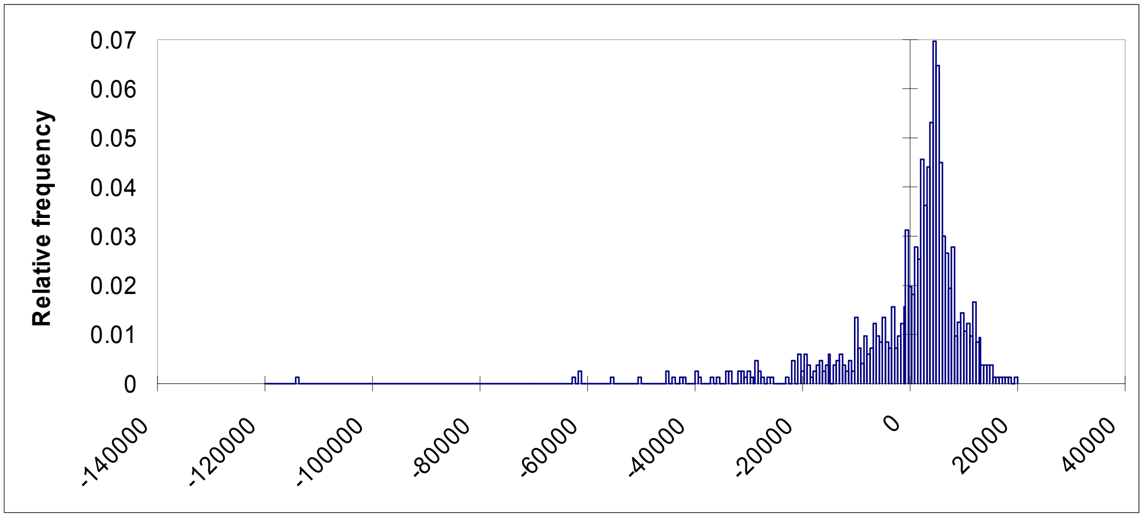
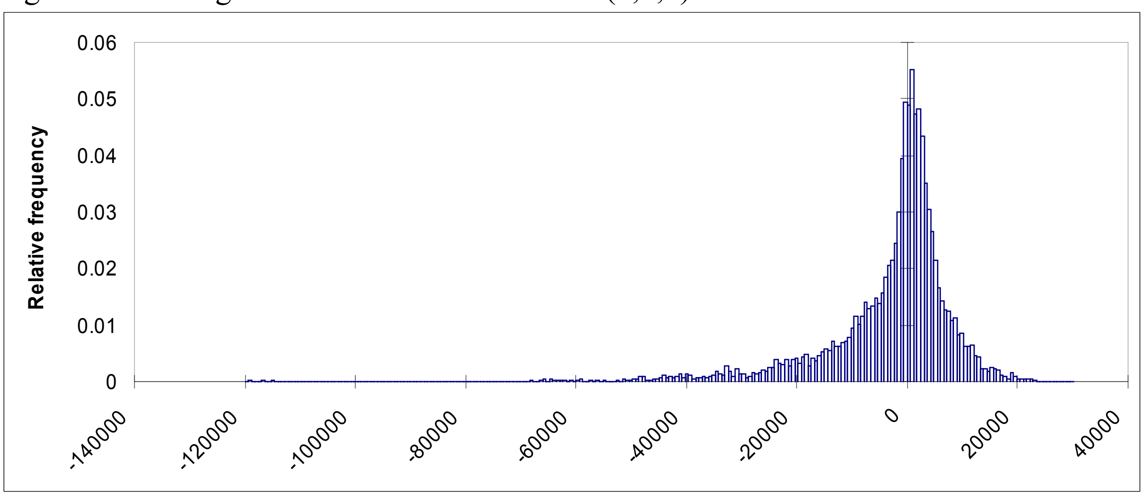


Figure 24: Histogram of errors for the ARIMA(2,0,2) model



It is worth noting that Figure 24 helps confirm the validity of the ARIMA(2,0,2) model. A good forecast model should have minimal bias—i.e., the distribution of its errors should not be skewed and should show central tendency near zero. Figure 24 shows that this is the case fro the ARIMA(2,0,2) model, though there is a slight negative skew. The error distributions for the ARIMA(3,0,0) and ARIMA(1,0,1) model are similar.

Model Selection

The ARIMA(2,0,2) model was selected for implementation in predicting Occoquan Reservoir refill. The results of the ARIMA(3,0,0) and ARIMA(1,0,1) models were similar, but the ARIMA(2,0,2) was selected for two primary reasons. One, it resulted in smaller forecast errors, especially in the 2000-2006 verification flows. Second, flows stochastically generated with the ARIMA(2,0,2) model reproduced seasonal patterns better than the other models. These are important factors since the model is used to stochastically generate potential inflow series given recent conditions.

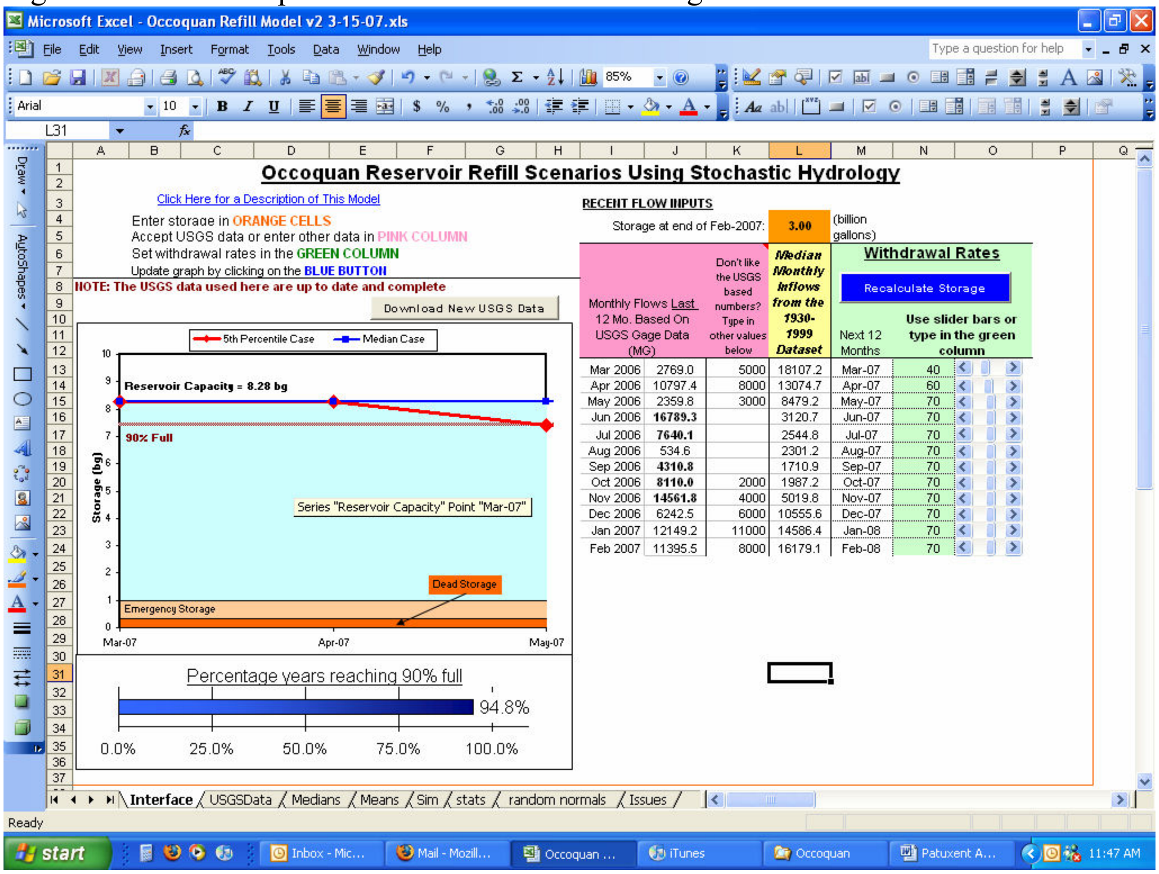
Occoquan Refill Forecasting Tool

The Occoquan ARIMA model was implemented in a similar manner as the Patuxent model to create a refill forecasting tool. The Occoquan tool functions similarly and has a similar interface.

The USGS gage data automatically downloaded by the tool are for Cedar Run near Catlett, Virginia. This gage represents a small portion of the Occoquan watershed (93.4 square miles versus 591.9 for the entire Occoquan watershed) and so the area adjustment method is applied to generate a series of inflows appropriate for the ARIMA model. This result in adequate data for most situations. However, adjusted Cedar Run flows tend to overestimate actual Occoquan inflows during very dry situations. Therefore, during drought situations, it may be beneficial to initialize the Occoquan model with computed inflows based on recorded storage, withdrawals, and releases. The computed data can be entered manually to replace the USGS data for generating flow forecasts.

Other than the subtle difference regarding initialization data mentioned above and the fact that they use different ARIMA models, the Occoquan and Patuxent refill forecasting tools are virtually identical and should be used in a similar manner. The Occoquan interface is configured a little differently, but this does not affect the functioning of the model. The Occoquan interface is show in Figure 25 for information.

Figure 25: The Occoquan Reservoir Refill Forecasting Tool interface



References

Box, George E. P., Gwilym M. Jenkins & Gregory C. Reinsel, 1994. *Time Series Analysis: Forecasting and Control.* 3rd Ed. Prentice Hall International, London.

Hagen, E. R., and R.C. Steiner, 1998. *Occoquan Reservoir Watershed: "Natural" Daily Inflow Development.* Interstate Commission on the Potomac River Basin. ICPRB report 98-3. Rockville, MD.

Hagen, E.R., R.C. Steiner, and J.L. Ducnuigeen, 1998. *Patuxent Reservoirs: "Natural" Daily Inflow Development.* Interstate Commission on the Potomac River Basin. ICPRB report 98-4a. Rockville, MD.

Hirsch, R.M., 1978. Risk Analyses for a Water-Supply System—Occoquan Reservoir, Fairfax and Prince William Counties, Virginia. U.S. Geological Survey, Open File Report 78-452. United States Department of the Interior. 1978.

Kame'enui, A., E.R. Hagen, J.E. Kiang, 2005. Water Supply Reliability Forecast for the Washington Metropolitan Area, Year 2025. Interstate Commission on the Potomac River Basin, ICPRB 05-06, Rockville, Maryland.

Makridakis, S, S.C. Wheelwright, and R.J. Hyndman, 2003. *Forecasting: Methods and Applications*. John Wiley & Sons, Inc, New York.

Salas, J. D., J. W. Delleur, V. Yevjevich, W. L. Lane, 1980. *Applied Modeling of Hydrologic Time Series.* Water Resources Publications, Littleton, Colorado.

Water Supply Coordination Agreement, 1982. Signatories include: The District Engineer Baltimore District COE for the United States of America, the Chairman of the Fairfax County Water Authority, the General Manager of the Washington Suburban Sanitary Commission, The Mayor of the District of Columbia, General Counsel for the Interstate Commission on the Potomac River Basin.