DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOR/ENS/95M-19

SPATIAL TIME-SERIES: POLLUTION PATTERN RECOGNITION UNDER IRREGULAR INTERVENTIONS

THESIS

Samuel A. Wright
Captain, USAF

AFIT/GOR/ENS/95M-19

Approved for public release; distribution unlimited

AFIT/GOR/ENS/95M-19

SPATIAL TIME-SERIES: POLLUTION PATTERN RECOGNITION UNDER IRREGULAR INTERVENTIONS

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Operations Research

Samuel A. Wright, B.S.
Captain, USAF

March 1995

THESIS APPROVAL

STUDENT: Captain Samuel A. Wright CLASS: GOR 95-M

THESIS TITLE: Spatial Time-Series: Pollution Pattern Recognition under Irregular Interventions

DEFENSE DATE: 2 Mar 95

COMMITTEE:	NAME/DEPARTMENT	SIGNATURE
Advisor	Dr. Yupo Chan/ENS Professor of Operations Research	
Reader	Thomas S. Kelso, Lt Col, USAF Adjunct Professor of Space Operations	

Preface

The purpose of this study was to develop alternative means of examining groundwater pollutants, particularly under irregular interventions such as the operation of pumping wells. FERMCO has identified the existence of arsenic pollutants in the groundwater near the site of the groundwater recovery system which pumps groundwater out for remediation purposes. The arsenic is of unknown origin and seems to become more prevalent with the increased operation of the groundwater recovery system. Univariate ARMA and causal transfer function models are constructed, as well as a univariate STARMA and causal transfer function model, and the two approaches are compared. The results are promising, and insights as to the true nature of the system are obtained.

I would like to thank Dr. Yupo Chan, the advisor for this effort, for providing a multitude of methodological possibilities, many of which are reviewed in chapter 2. I also wish to thank Lt. Col. Thomas S. Kelso for his long-distance assistance as my reader. Thanks go to Maj. Dave Coulliette and Dr. Bob Ritzi for their assistance in the analytical portion of groundwater modeling. To those at FERMCO for providing the data and providing answers to site specific questions. Finally, and mostly, I thank my wife Vicki and our children Samantha, Sammy, and ? for providing perspective to my time here.

Samuel A. Wright

Table of Contents

List of Figures

Abstract

The Fernald Environmental Restoration Management Corporation (FERMCO) has noted the introduction of arsenic contamination to groundwater around the area of the groundwater recovery system, which captures groundwater from the ground to prevent further expansion of the South Plume of uranium contamination. The introduction of arsenic occurs during high levels of pumping and is particularly sensitive to the western two of the five pumps.

Auto-Regressive Moving Average (ARMA) and Spatial-Temporal ARMA (STARMA) empirical analyses are used to model the level of arsenic contamination found through time. The intervention of varied levels of pumping is modeled with a transfer function using analytic techniques to create a causal intervention transfer function input series to give physical meaning to the impulse response weights found. Spatial weights employed in the STARMA modeling are also created using analytic, causal methods. Implementation of the temporal and spatial-temporal modeling is performed with an algorithm that iteratively estimates a) the temporal or spatial-temporal relationship and b) the intervention transfer function in an attempt to estimate the true model parameters as closely as possible.

Results from the temporal modeling suggest a clear physical interpretation of the causal relationship between a) a particular level of pumping and b) the amount of arsenic to be found at the site. Results from the spatial-temporal modeling suggest there is at least a correlative relationship between a) a particular level of pumping and the effect of the pumping on a site of interest's neighbors and b) the amount of arsenic to be found at the site of interest. The models presented may be employed in the forecasting of arsenic levels at various monitoring well sites due to a given change in the operation level of the groundwater recovery system.

SPATIAL TIME-SERIES: POLLUTION PATTERN RECOGNITION UNDER IRREGULAR INTERVENTIONS

I. Introduction

1.1 Introduction to the Problem

One of the final tasks involved in the decommissioning of the Fernald Plant is the remediation of uranium contaminated ground-water. Five pumping wells exist to assist in this effort, however, the use of these wells has increased the presence of arsenic in the ground water, possibly of off-site origin. The intent of this research is to provide an empirical model for the propagation of this ground water arsenic pollution which can give short-term forecasts for use by FERMCO in pollution level monitoring. Specifically, the effect of the use or disuse of the existing remediation wells or of prospective remediation wells may be quantified. In this way, the use of the current or of future remediation wells may be planned so as to minimize the collective effect of the wells in terms of the introduction of additional arsenic to the ground water. In this thesis, "pollution" will refer to elevated levels of arsenic concentration, unless otherwise noted.

Conventionally, numerical and analytical approaches such as those used with SUTRA, MODFLO, and SWIFT software have been taken to model the flow of ground-water. Knowledge of how pollutants travel through the ground water has allowed the super-imposition of pollutants onto the ground water flow models, resulting in a model for the propagation of a pollution plume. This approach has been applied with some success to the propagation of the plume of uranium polluted ground water. However, due to the complex geometry of the physical plant and the remediation of polluted ground

water over time, calibrating the large numbers of parameters required by these methods makes the methods difficult to implement for this problem. In the process of the remediation of the uranium contaminated ground water, an increased level of arsenic was presented in the ground water near the remediation wells. Due to the complex flow of ground water near the remediation wells and the fact that little is known of the source of the arsenic pollutants, prediction of the flow of arsenic within this area creates many difficulties with the conventional numerical and analytical approaches.

Therefore, an empirical approach using spatial time series methods to model the arsenic contamination and the effect the remediation wells have on the introduction of arsenic to the ground water system is considered. Spatial time series methods can provide insight as to the ground water arsenic flow as well as to the effect the remediation pumps have on the ground water arsenic flow without the complexity of required of an analytic study of the ground water flow.

Tasks involved in creating this empirical model for ground-water pollution propagation include 1) performing a spatial-temporal analysis (data analysis across both space and time dimensions) of the polluted areas to include intervention analysis (a study of the effect of intervening events which could impact the spatial-temporal analysis) based on the varying scheduling of the remediation wells, 2) delineating questionable data values for their possible enhancement, and 3) integrating the first two tasks into a usable, iterative model to perform short-term forecasting of the propagation of ground-water pollution.

1.2 Scope

Next, a temporal and spatial-temporal analysis is performed on the data to establish the propagation pattern of the pollution plume through time. To perform the spatial-temporal analysis of the polluted areas, a heterogeneous ground composition can be

assumed. This assumption is necessary since pollution propagation cannot be assumed constant under varied ground composition types. For instance, at the Fernald Plant, the ground water flows through the Great Miami Aquifer, a large underground channel which has walls of impenetrable bedrock and an interior composition of sand and gravel in most places and clay in other places. The use of this site-specific knowledge may be accomplished with a simple geographic information system (GIS) which contains the densities of the ground composition of each pixel of interest, as well as information concerning the direction of ground water flow at the pixel. With this information, the spatial-temporal modeling is accomplished to determine the characteristics of specific pollution plume propagation.

Through time, various abrupt events which could affect the propagation of the pollution plume occur. These events include the changing status of the remediation pumps used to remove polluted ground water. At various times, the five remediation well pumps have been either off or at some specified level of pumping. Intervention analysis techniques are used to model these abrupt interventions in the spatial-temporal model. The result of this analysis is the marginal effect of an active remediation pump. This information may be used to assist the future operation scheduling of the current pumps or to tests the possible effects an additional remediation pump could have.

The objective of this analysis is to remove questionable data values so future iterations of the spatial-temporal analysis may occur with as clean data as possible. This could lead to the identification of a homogeneous model which would model the non-intervention level of arsenic at any of the well sites in the study equally as well. Though the number of usable "pixels" for use in an image classification procedure is limited in this research, the procedure for performing this analysis is outlined as a proposal for use with other studies possessing more robust data sets.

Finally, forecasts of the pollution plume area may be constructed. These forecasts may be used to determine the effectiveness of remediation efforts and the effect of potential additional resources required to complete the task. Forecasts may also be compared to actual future contaminate levels in the vicinity of the pollution plume to determine the model's effectiveness and to adjust the model's calibration as necessary.

1.3 Main Research Objective

The main objective of this thesis is to develop a means with which to provide reasonable short-term forecasts to FERMCO concerning the propagation of ground water arsenic pollution under the irregular intervention of a varying level of ground water remediation via the south plume recovery system. While other ground water modeling techniques are effective, this site-specific and data-oriented approach used in concert with what is known of the analytic nature of the system is more flexible in its simplicity and use of intervention analysis to determine the marginal effect through time of each remediation pump. For this reason, this approach should be viewed as complementary to other on-going efforts.

1.4 Secondary Research Objective

The secondary objective of this thesis is to present methodology for the delineation of questionable data points for their possible removal from the analysis in order to obtain as accurate a spatial-temporal model as possible, as well as to develop a homogeneous model form representing the natural arsenic flow in the entire area of interest around the groundwater recovery site. Though the number of sites is limited for this analysis, this analysis could lead to the identification of invalid data, the knowledge of which could be used to further refine the spatial-temporal analysis, and to refinements in a homogeneous spatial-temporal model for the region, thereby improving the model's quality.

1.5 Research Product Deliverables

The final product delivered to FERMCO includes 1) this research thesis including accomplished tasks, extension possibilities, and recommendations for increasing the usefulness and validity of such modeling, 2) an estimated model concerning a) the baseline spatial-temporal flow of arsenic through the immediate area of the groundwater recovery system, b) the effect the intervention of operating the groundwater recovery system pumps has on the spatial-temporal portion of the model, and c) the delineation of polluted from unpolluted sites, and 3) FORTRAN code which may be used a) to update this model through recalibration of parameters or through the inclusion of one or more model extensions or b) to provide forecasts based on existing circumstances or based on hypothesized circumstances such as operating the groundwater recovery system pumps at different levels or determining the possible effect of additional pumps.

II. Literature Search and Review

2.1 Introduction

This chapter describes the current state of technology in regards to this research. Merits of the methods discussed are determined with respect to this effort. The tasks for which modeling methods are evaluated include 1) performing temporal and spatial-temporal analyses of the contaminated areas of groundwater, to include intervention analysis of the effect of operating the south plume groundwater recovery system and 2) delineating questionable data values for their possible enhancement. This second task may also be used to delineate polluted areas from unpolluted areas on a pixel map, and within the polluted areas, delineating various levels of pollution

Tasks involved in creating such a model for ground-water pollution propagation include 1) performing temporal and spatial-temporal analyses of the polluted areas to include intervention analysis, 2) delineating questionable data for enhancement on a pixelized map, and 3) integrating the first two tasks into a usable model to perform short-term forecasting of the propagation of ground-water pollution. This chapter reviews current literature which presents methods applicable to the first two tasks to determine specific methods applicable for temporal and spatial-temporal modeling of the groundwater pollution including intervention analysis and to delineate questionable data points for their possible enhancement in order to obtain a better and possibly homogeneous model for the entire area of interest. A review of possible model validation techniques is also presented.

2.2 Temporal Modeling of Pollution Plume with Intervention Analysis

Various potential methods for performing three separate tasks in the temporal modeling of the arsenic pollution plume are reviewed. First, a strictly temporal analysis

will be conducted on the well sites of interest. Applicable methodology used to perform this analysis is presented in section 2.2.1. Next, a spatial-temporal analysis is to be accomplished for comparison to the strict temporal analysis. These spatial-temporal analysis methods are described in section 2.2.2. Finally, intervention analysis is required to determine the effect of the operation of the ground water recovery system pumps as on the level of arsenic found at the various well sites. Potential methods for use in intervention analysis are found in section 2.2.3.

2.2.1 Temporal Modeling

Forecasting the state of a variable based on that variable's past states is a straight-forward process and has been well documented by Box and Jenkins [5], Kashyap and Rao [17], and Makridakis et al [23] using autoregressive integrated moving averages models. The original Box and Jenkins methodology is found useful in modeling discrete systems where measurements of the system's state occur at evenly spaced intervals of time [5:1]. Makridakis et al describe the use of Box and Jenkins procedure of modeling univariate time-series data. The ARIMA (Auto[R]egressive Integrated Moving Average) procedure is shown to be an effective means with which to forecast based on history [23:358-363]. The ARIMA procedure alone, however, is limited in its usefulness without modification. For instance, a procedure for handling data irregularities and a method for studying the spatial relationship between pixels as well as the temporal relationship, could prove useful. Following are relevant extensions to the original Box and Jenkins procedure.

Tiao and Box present a method for the modeling of multiple time series [32]. This method allows understanding of the dynamic relationships between various time series. Also, modeling multiple time series will improve the accuracy of the forecasts [32:802]. A transfer function is presented as a means to model multiple time series. The transfer function presented implies that the state of one specific pixel relies only on its own past;

that the state of another pixel depends only on its own past and the past and present of the first pixel; that the state of a third pixel depends only on its own past and the past and present of the first two pixels; and so on [32:802-803]. This may not be practical for modeling groundwater data, as the pixels in all directions affect a central pixel. Tiao and Box go on, however, to present a vector ARMA model which allows for the required feedback among pixels [32:803]. This method separately models each pixel based on its relationship to its neighbors and based on its relationship to the past of itself and its neighbors.

2.2.2 Spatial-Temporal Modeling

Pfeifer and Deutsch [27] and Chan [7] present methodology to analyze spatially correlated data using ARIMA as a basis. "The STARMA (Space-Time Auto[R]egressive Moving Average) model examines the relationship between the current observation . . . as a linear combination of past observations as well as observations at neighboring sites" [7:29]. Pfeifer and Deutsch describe a three-stage iterative procedure for solving STARMA models: 1) identification of a tentative STARMA model, 2) estimation (fitting) of the tentative STARMA model, and 3) diagnostic checking of the model [27:35-36]. This model could be employed to analyze multiple pixels of groundwater pollution data based on past observations of each pixel and nearby pixels, while maintaining the simplicity of a univariate ARMA model.

Various orders of neighbors may be considered when estimating a STARMA model. For instance, the zeroth order neighbor is considered the pixel under consideration itself. The first order neighbors consist of those pixels closest to the pixel under consideration. Second order neighbors are further away, and so on. For use in the STARMA model, each of the lth order neighbors is consolidated into a single data set using spatial weights which sum to unity. Represented on a normal pixel grid, the neighbors of any one order

are of equal distance from the target pixel. If each of the neighbors in the same order can be assumed to have the same relationship to the target pixel, Pfeifer and Deutsch suggest the use of equal weights if possible, not only because they are beneficial in modeling regularly spaced homogeneous systems, but also because they serve as a pattern for more general weighting schemes [27:40].

The regular grid used in this research, however, is sparse in that the number of sampling wells accounts for only about ten percent of the total number of pixels, and these wells are not evenly spaced. For this reason, a spatial weighting scheme is devised based on site-specific information. One way to assess the degree of the autocorrelation among all the sites is to perform a spatial autocorrelation analysis. Black suggests that spatial autocorrelation exists when the value of a random variable (such as arsenic level) at a particular location depends upon the values of the random variable at other contiguous locations [4:207]. He then gives examples of the use of Moran's I statistic in evaluating such spatial and network autocorrelations [4:209]. Getis suggests a difference between spatial autocorrelation, which merely suggests correlation between contiguous sites, and spatial interaction, which implies a more causal relationship between sites [10:1269]. He then describes a general spatial statistic which may be used in either spatial autocorrelation or spatial interaction models [10:1274-1275]. Tobler gives an alternative formulation for spatial interaction modeling, however the conservation rule required for implementation is not realistic in the context of this research, as the movement of arsenic form an origin does not necessarily arrive at a destination, since arsenic measurements are made not in terms of absolute arsenic, but in terms of arsenic concentration in ground water. Also, the sparsity of the monitoring wells allows arsenic to flow between the wells, thereby breaking the conservation rule, as well [33:693].

A Geographic Information System (GIS) is considered to deal with site-specific information because the large amounts of data being collected require a common

reference for applying effective modeling techniques. Lang's article provides a synopsis of how GIS has and will affect the water industry and quotes the president of a software firm: "We're headed toward the full integration of . . . modeling [and] GIS . . . to keep track of and monitor enormous volumes of data" [21]. The use of a simple, even ad-hoc GIS makes the model formulation for this problem generic, allowing its use in other situations given a relevant site-specific GIS of the same form.

Greene shows the use of unequal spatial weights due to an irregular grid pattern [13:58]. She suggests that spatial weights be selected to represent a physical property of the various pixels and also shows that a first or second order neighbor does not have to be considered in the traditional sense of a regular square grid in order to be effective [13:58-59].

2.2.3 Intervention Analysis and Data Irregularity

Box and Tiao also describe a method of intervention analysis which applies a transfer function to model the occurrence of an event which affects a time series [6:70]. Several different causal intervention responses are given based on what is known of the true response [6:72]. In this research, the nature of the response to a particular intervention, such as the nature of the increase of arsenic in the ground water due to a change in the operations of the remediation pumps, is not necessarily known. An analysis of the residuals of a time series fit through the intervention coupled with knowledge of the effect the operation of the remediation pumps has on ground water flow could provide the required causal relationship to give a desired transient response function.

Even when little is known of the nature of an intervention, Pankratz suggests the use of an intervention transfer function [26:268]. In this method, a transfer function which allows for transient behavior of an intervention effect is estimated empirically from the data which can then be added to the existing ARIMA model. Again, knowledge of the

effect the operation of the remediation pumps has on ground water flow may provide insight to the required causal relationship.

Also, methods for handling multiple interventions which occur at different times and compound interventions which occur at the same time are presented [26:272-273]. The different levels of remediation pump operation may be considered multiple interventions in this research which could be modeled using this simple additive intervention technique.

Greene illustrates the use of intervention analysis when little is known about the nature of the intervention through the employment of the fractile method based upon a survey of experts to determine the extent of an intervention [13:99-107]. The fractile method was used to determine the extent of the intervening event in terms of the form and shape of the transfer function [13:100]. However, more analytical means should be used if anything is known about the nature of the underlying system.

Viessman, et al provide a method of analyzing the effect of a pumping well at a site of interest [36:322-325]. They also extend the analysis to consider multiple pumps in a well field, as is the case in the groundwater recovery system [36:325-326]. The use of simple analytical groundwater modeling techniques applied to the input series in an intervention transfer function will increase confidence in the eventual calibration of that transfer function. Such analysis may also be used to provide spatial weights for the spatial-temporal modeling.

Irregular data sets in which, for instance, observations are made at different time intervals or certain observations are missing are considered by Harvey and Pierse [15], by Kohn and Ansley [20], and by Aoki [2]. Harvey and Pierse describe the application of the Kalman filter to an ARIMA-type problem represented in state-space form. This application makes it possible to estimate both the required parameters in the ARIMA model when some of the observations are missing and the values of the missing

observations [15:125]. Kohn and Ansley present a similar application of the Kalman filter, but their application allows for any pattern of missing data while the application by Harvey and Pierse does not [20:752].

An extended and thorough presentation of the state space formulation of the ARIMA model class, known as structural time series, is given by Harvey [14]. Harvey shows that the simplest structural time series models have a corresponding ARIMA representation but that the structural form is much more flexible [14:12]. Structural time series models may include one or more explanatory variable besides the time series variable. Among other things, these explanatory variables may be dummy variables used to explain interventions introduced into the system [14:397]. The impact of the interventions may then be assessed through the application of the Kalman filter [14:400].

Applications of the Kalman filter are given in the works of Porter [27] and Gallagher [9]. Porter provides an explanation of a Kalman filter from *An Introduction to Kalman Filtering with Applications* by Miller and Leskiw:

> The basic idea behind Kalman filtering is to combine the *system model* (the differential equation governing the flight path of the vehicle) and the *measurement model* (the observations made on vehicle position) in an optimum fashion in order to obtain the best estimate of the position of the vehicle at any time $t_1 > t_0$. [27:5]

For the case of time heteroscedasticity among a set of data, that is, the variance of a response varies over time, Rochon presents a method in which the standard deviation is allowed to change over time [30:778].

For small to moderate-sized time series, Stoffer and Wall suggest that ARMA estimates and estimates generated by the state space model Kalman filters are suspect in terms obtaining asymptotic results [31:1024]. Their method employs the nonparametric Monte Carlo bootstrap which is based on the Gaussian maximum likelihood estimator, which has favorable asymptotic properties. At the same time, however, the family of

distributions for the model must be known in order to employ this technique, making it undesirable for many time series analyses [31:1024].

2.3 Delineation of Regularly or Irregularly Pixelized Data

Many methods exist for delineating members of a set from one another, but in terms of image classification, these methods are generally applied to images with regular pixelized grids. Care must be taken, therefore, if these methods are to be applied to non-regular pixels as the sampling wells at the Fernald site are situated. Either a method must be modified to be able to handle such a situation, or a transformation must be made to induce "regularity" to the grid so the methods can be applied directly. McGee presents a kriging method which is a statistical "method of best linear unbiased prediction of spatial data" [24:x]. Specifically, kriging may be used to predict "the gray value of all pixels which are spatially distributed on an image of gray values, whether or not the data previously existed at that point" [24:1-2]. Performing this type of analysis in order to realize a "filled-in" pixel map with which to determine contours would be an acceptable use of such a technique. However, completing historical pixel maps for their use in spatial-temporal analysis will only serve to add additional error to model parameters and ultimately, forecasts. Whether the data are kriged or not, the available pixel "gray values" indicating a particular level of contamination must be delineated, either into unpolluted and polluted classes, or into several classes representing various levels of contaminants.

Walsh and Burk suggest "The classification of land into various categories is an integral component in analyzing many natural resource issues"; further, "Methods for estimating land area need to be relatively unbiased, accurate, and efficient" [37:281]. The relevance of land area classification to ground-water pollution delineation is clear. One purpose of this research, therefore, is to find the least biased, most accurate and efficient delineation technique for use on large models.

Benabdallah and Wright demonstrate that selecting a contiguous set of cells from a set of candidate cells that best represents a specified set of goals within specific constraints can be accomplished using either linear or non-linear modeling techniques. An important feature of the Benabdallah and Wright algorithm is in its delineation of non-inferior solutions. The authors have presented algorithms for solving small-sized multiple subregion allocation problems and a heuristic method of solving the problem for medium-sized problems. The authors include several extensions to previous modeling attempts, resulting in five new subregion allocation formulations. The most complex algorithm is then tested, and results, in terms of computational efficiency, are given. Finally, possibilities for uses of the algorithms presented and for extensions to the current model formulations are given [3:24-38]. The Benabdallah and Wright algorithms, while useful for small problems, are too computationally demanding for operational use.

Reed attempted to formulate a network representation of the Benabdallah and Wright subregion allocation model for "segmenting large numbers of satellite imagery pixels into multiple regions" [29:5]. However, "based on the tremendous increase in branch-and-bound iterations required to reach the solution, any savings in computation time would soon be overcome by further branch-and-bound iteration increases as larger problems were encountered" [29:94].

Amrine met with similar limited success in implementing the Benabdallah and Wright model, but he was able to combine more than one image to enhance the correct delineation of the ground truth through multicriteria optimization [1:xiii]. Still, Amrine admits that "improvements to this model are necessary" in terms of reformulation as a network with side constraints problem and in terms of a reformulated objective function [1:xiii].

Gonzalez and Woods present a logit split (delineation of areas based on two or more modes) model using an algorithm based on the Bayes classifier for Gaussian pattern

classes. This method defines a surface of separation between the two classes of pixels (polluted and unpolluted ground-water) [12]. This method is an improvement to the Benabdallah and Wright algorithm in that it is computationally efficient, however, the method does not necessarily provide contiguous subregions after the delineation. Also, both the Benabdallah and Wright method and the Gonzalez and Woods methods require some degree of knowledge about the ground truth in order for accurate modeling.

McLachlan presents a method for performing a contextual allocation of pixels which is also more computationally efficient than that of Benabdallah and Wright. The Iterated Conditional Modes (ICM) algorithm is "one of the standard algorithms for the segmentation of images" [25:428]. The method is contextual because it delineates generally contiguous areas from one another, thereby reducing the effect of random noise present in the data. This data "smoothing" will also allow for the classification of complex images with more than one pollution plume present. This model has advantages over the Benabdallah and Wright formulation in that 1) it is contextual in nature, 2) little has to be known about the ground truth in order to apply the algorithm, and 3) it can be applied to large problems with relatively small computational requirements.

Klein and Press present a Bayesian method for performing the classification of spatial data [19:844]. This adaptive Bayesian classification (ABC) method is similar in theory to ICM, however, ABC uses a formal predictive Bayesian classification of the area of interest as its initial starting point, then iteratively reclassifies all of the labels based on an empirical Bayesian updating algorithm [19:844]. In several tests performed by Klein and Press, the ABC algorithm performed marginally better than the ICM algorithm in terms of mean percent of correct classifications per scene and in terms of having generally smaller variances [19:850]. However, the test was performed on a scene of the authors' choice, and the results were not conclusive.

Walsh and Burk also suggest two methods for performing pixel delineation founded on the likelihood principle and Bayesian reasoning, respectively. The thrust of the article, however, is to show how delineation can be improved by careful calibration of the remotely sensed data to what is known about the ground truth [37:281]. In other words, the more information one has of the ground truth, the better the delineation of the pixels will be. Since little is known of the ground truth in the instance of this study, and in fact the ground truth is dynamic in this study, these methods are not likely candidates for a method of pollution delineation.

Kaufman and Rousseeuw demonstrate an optimal method for delineating a smooth data set into k clusters [18]. The k-Medoid Method partitions pixels into k clusters which minimize the sum of the distance each pixel is away from the pixel representing the cluster to which it is assigned [18]. In this formulation, the distances can be represented by the differences in gray values from one pixel to the next. For a noisy data set in which a degree of contextuality would be useful for identifying more contiguous subregions, the distance can be represented by some combination of gray value difference and the actual distance pixels are away from each other. The result of such an analysis would be a contour plot of k gray values over the given area. This method is attractive in that it provides an "optimal" solution to the image classification problem, and in that it may be solved using a network algorithm, thereby reducing computation time.

2.4 Model Validation

The final step in the modeling procedure will be validation. Lin, et al. present validation of an analysis performed on remotely sensed data of soil moisture over a heterogeneous watershed [22]. The article shows that while forecasts concerning soil moisture were not completely accurate, they were generally accurate, indicating that

remotely sensed data could be effective in correcting the initial state of a hydrologic model [22:159].

2.5 Summary

The states of technology of two integral tasks in performing empirical pollution plume propagation modeling 1) the temporal and spatial-temporal modeling of the contaminated groundwater areas to include intervention analysis due to varied use of the south plume groundwater recovery system and 2) the delineation of questionable data values for their possible enhancement are reviewed. For performing the spatial-temporal modeling, ARMA and STARMA models are reviewed. Transfer functions and Kalman filtering are presented for dealing with interventions. For performing the delineation of polluted pixels, both non-contextual and contextual classifications are given. A validation method performed on similar analysis is also reviewed.

The number of and varied nature of techniques currently available should allow for the successful temporal modeling of spatially correlated groundwater pollution. The accuracy of the obtained model is uncertain, however, and will depend on several factors including the quality of data. The application of techniques to perform the modeling in a timely manner will also be an important measure of effectiveness for such a model.

III. Methodology

3.1 Introduction

This research effort combines two separate modeling techniques to show the one possibility for handling a multi-faceted problem such as this. First, a spatial-temporal analysis of the available data is performed in order to make predictions pertaining to the level of arsenic pollution found at the various well sites. Univariate ARMA model descriptions of the sites of interest will be compared to univariate STARMA modeling in order to determine the validity of including spatial aspects to the model. Of particular interest in this study is the effect the intervention of operating the groundwater recovery system pumps has on the arsenic level at the well sites. The methods used to determine both the short-term forecasts pertaining to arsenic level and the effect of the intervention of operating the groundwater recovery system pumps are detailed in section 3.2.

Second, the delineation of questionable data points may be accomplished. Although the delineation in this research would only cover a relatively small number of sites, the concepts are extended easily to larger data sets. The methods pertaining to this delineation of questionable data points are found in section 3.3.

3.2 Temporal Modeling of Pollutants with Intervention Analysis

Methods for performing three separate tasks used in the temporal modeling of the arsenic pollution are presented. First, a strictly temporal analysis will be conducted on the well sites of interest. Methods used to perform this analysis are detailed in section 3.2.1. Next, a spatial-temporal analysis is accomplished for comparison with the strict temporal analysis to determine the validity of using such an approach in this type of problem. The spatial-temporal analysis methods are described in section 3.2.2. Finally, intervention analysis is accomplished to determine the effect of the operation of the

groundwater recovery system pumps has on the level of arsenic found at the various well sites. Methods employed in this analysis are found in section 3.2.3.

3.2.1 ARMA Temporal Modeling of Pollutants

The ARMA model can be used to predict the "gray-value" of a polluted pixel based on the pixel's past history. The general form of the ARMA model is

$$z_t = \sum_{k=1}^{p} \phi_k z_{t-k} - \sum_{k=1}^{q} \theta_k a_{t-k} + a_t \qquad [1]$$

where z_t is a member of the time series of interest, p is the order of autoregressive parameters, ϕ_k are the p autoregressive coefficients, q is the order of the moving average parameters, θ_k are the q moving average coefficients, and a_t is a member of the series containing the model residuals (also seen as ε_t).

I will illustrate this model with an example of an ARIMA (1, 0, 1) model. This designation is a standard ARIMA (p, d, q) representation where p is the autoregressive order, d is the order of differencing, and q is the moving average order. (The backward shift operator works on z_t as follows: $Bz_t = z_{t-1}$.)

$$(1-\phi_1 B)z_t = (1-\theta_1 B)a_t \qquad [2]$$

The term on the left side is the autoregressive component, the term on the right side is the moving average component. Representing the model in its reduced form yields

$$z_t = \phi_1 z_{t-1} - \theta_1 a_{t-1} + a_t \qquad [3]$$

where the only calibrations required are the autoregressive parameter ϕ_1 and the moving average parameter θ_1.

Differencing of the time series may be necessary to induce stationarity in the mean. The determination of whether differencing is necessary or not will be covered in more detail in section 3.2.1.1. First-order differencing in the above example would take the form of a (1 - B) term on the left side of the backshift notation equation, resulting in slightly more algebraic manipulation, but the number of coefficients requiring calibration remains the same.

Model building using the ARMA model class may proceed in the same manner Pfeifer and Deutsch describe in their three-stage iterative procedure for space-time modeling. The first stage of identifying a tentative model involves the analysis of the data structure specifically checking for autocorrelations in the time series. The second stage requires fitting the tentative model by estimating the model's required parameters. The third stage involves a diagnostic check of the tentative model in order to determine the adequacy of the fit model. If the model fit proves adequate in this stage, the model is complete. If the model fit is inadequate, the steps must be repeated in a second, or subsequent, iteration [26:35-36]. Sections 3.2.1.1 through 3.2.1.3 will describe each of these three stages in more detail.

3.2.1.1 Identification of an ARMA Class Tentative Model

Determining the required numbers of autoregressive, differencing, and moving average parameters to include in a model is a somewhat subjective task. Two statistics which assist in the process, however, are the sample correlation and sample partial autocorrelation coefficients. The sample autocorrelation coefficient r_k is an estimate of the direction and strength of the relationship among observations in a single time series, separated by k time periods [25:36]. This estimate is usually calculated as

$$r_k = \frac{\sum_{t=1}^{n-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n}(z_t - \bar{z})^2} \qquad [4]$$

The partial autocorrelation coefficient is similar to the autocorrelation coefficient, except that in computing it, the roles of all intervening random variables, z_{t+k-1}, z_{t+k-2}, . . ., z_{t+1} are simultaneously taken into account [25:38]. A time series has many autocorrelation and partial autocorrelation coefficients, one for each value of *k*. Since an observation is lost each time *k* increases by one, however, the maximum useful value of *k* is somewhat less than *n*. The common rule is to use a maximum *k* of *n*/4, then, where *n* is the length of the time series [25:35].

Several guidelines have been developed for the use of determining the amount of differencing required and the numbers of autoregressive and moving average terms required in an ARMA model. First, If the mean of the time series is stationary through time, the sample autocorrelation function (a list of the sample autocorrelation coefficients from $k = 1, 2, \ldots,$ max k) will tend to decay quickly toward zero [25:37]. If this is not the case (sample autocorrelation coefficients do not quickly tend toward zero), first or higher order differencing may be performed in an attempt to force stationarity in the mean. (Second order differencing is accomplished by differencing the first order differenced data, and so on.)

In general, the autocorrelation function for an autoregressive process of order *p* will tail off while the partial autocorrelation function will cutoff after lag *p*. For moving average processes, however, the autocorrelation function will cutoff after lag *q*, while the partial autocorrelation function tails off. For a mixed process with autoregressive order *p*

and moving average order q, the autocorrelation function should exhibit a mixture of exponential behavior and sine waves after the first q - p lags, while the partial autocorrelation function is dominated by a mixture of exponentials and damped sine waves after the first p - q lags [5:175]. Of course, the sample autocorrelation and sample partial autocorrelation functions may or may not exhibit such obvious trends. Because of this problem, the selection of the autoregressive and moving average orders is a very subjective process, which may need to be accomplished iteratively in order to determine a sound representation of the time-series data.

3.2.1.2 Estimation of ARMA Class Model Parameters

The parameters for the univariate ARIMA class model may be fit using any of a number of software packages. In order to provide somewhat standard code, the modeling here is accomplished with a FORTRAN program running IMSL subroutines [16].

3.2.1.3 Diagnostic Testing of an ARMA Class Model

The residuals of the fit model should exhibit the behavior of white noise. White noise is characterized by a constant mean of zero and a constant variance from zero of σ^2. All covariances should be zero, as well as autocovariances at non-zero lags [26:43]. A tentative model may fail in two ways. First, the model may inadequately represent the observed correlation of the process. If this is the case, the residuals will show significant correlation, which could potentially be modeled with additional autoregressive or moving average parameters. Second, too many parameters may be included in the model, resulting in estimated coefficients which are not statistically different from zero [26:43]. A candidate model which fails neither test may be usable in short-term forecasting of the time series, assuming no model with less calibrated parameters is equally as worthy.

3.2.2 STARMA Spatial-Temporal Modeling of Pollutants

The STARIMA model is used to predict the "gray-value" of a polluted pixel based on the pixel's past history, as well as the past history of the first and second-order neighbor pixels. The STARIMA model is an extension of the ARIMA model, except accounting for the pixel's *l*th order neighbors. Instead of periodic seasonal fluctuations in the data found in a SARIMA (Seasonal AutoRegressive Integrated Moving Average) model, we see periodic spatial fluctuations in the data. The general form of the STARMA model follows. L works as the backshift operator B, except on the spatial component. For instance, $L^{(l)}z_t$ is the lth order neighbor of z_t.

$$z_t = \sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}L^{(l)}z_{t-k} - \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}L^{(l)}a_{t-k} + a_t \qquad [5]$$

where λ_k is the spatial order of the kth autoregressive term and m_k is the spatial order of the kth moving average term [7].

However, a flexible model may be applied which may be estimated using ordinary ARIMA estimation methods: STARIMA $(p, d, q)^s$ (λ, D, m). In this model, s is the number of spatial orders included in the modeling (whether modeled as significantly affecting the zeroth order or not), p is the number of autoregressive components (temporal lags from 1 to p), d is the degree of differencing (temporal lag differencing), q is the number of moving average components (residual temporal lags from 1 to q), λ is the number of spatial autoregressive components (spatial lags from 1 to λ), D is the degree of spatial differencing, and m is the number of spatial moving average components (residual spatial lags from 1 to m). I will use illustrate this model with an STARIMA $(1, 1, 1)^4$ $(1, 1, 1)$ model:

$$(1-\phi_1 B^4)(1-\Phi_1 B)(1-B^4)(1-B)X_t = (1-\theta_1 B^4)(1-\Theta_1 B)e_t \qquad [6]$$

The first term is the autoregressive component (the backshift of four is due to the form of the input data stream, which will be explained shortly), the second term is the spatial autoregressive component, the third term is the first order differencing, the fourth term is the first order spatial differencing, the first term on the right hand side is the moving average component, and the second term on the right hand side is the spatial moving average component. After algebraic manipulation, the final model form is

$$\begin{aligned} X_t = {} & (1+\Phi_1)X_{t-1} - \Phi_1 X_{t-2} + (1+\phi_1)X_{t-4} - (1+\phi_1+\Phi_1+\phi_1\Phi_1)X_{t-5} \\ & +(\Phi_1+\phi_1\Phi_1)X_{t-6} + \phi_1 X_{t-8} + (\phi_1+\phi_1\Phi_1)X_{t-9} - \phi_1\Phi_1 X_{t-10} \\ & +e_t - \Phi_1 e_{t-1} - \phi_1 e_{t-4} + \theta_1\Theta_1 e_{t-5} \end{aligned} \qquad [7]$$

This model presents a wealth of flexibility while only requiring the estimation of four parameters.

Data required for performing this specialized form of STARIMA analysis must follow a specific form. The form is that of a string of values, the first representing the zeroth order neighbor (the data itself) at time t, the second represents the first order neighbor group at time *t*, and so on, until all desired spatial neighbor classes are accounted for. Then, the data representing time $t + 1$ are similarly accounted for, then time $t + 2$, etc. until all time periods are included.

Should more than one *l*th order neighbor exist, the data representing each neighbor must be consolidated into one point representing the group of *l*th order neighbors using weights which sum to unity [7]. If the data across the *l*th order neighbor are homogeneous, the weights can be as simple as 1/(# *l*th order neighbors). Pfeifer and

Deutsch suggest the use of equal weights not only because they are beneficial in modeling regularly spaced homogeneous systems, but also because they serve as a pattern for more general weighting schemes [26:40]. Due to the system of irregularly spaced well sites in this research and the fact that groundwater flow at the site is patterned and not consistent through space, the use of equal spatial weights would be problematic. If equal weights cannot be used, the weights should reflect some physical property of the system, such as distance or accessibility [26:36].

An empirical method of determining spatial weighting is proposed. First, the spatial-temporal correlations between all sites of interest are accomplished. A significant level of correlation between the time series' of spatial neighbors will suggest the neighbors should be considered first order neighbors in the STARMA model. The level of correlation, then, is used as the spatial weight, which must be normalized such that the sum of all weights for a particular lth order equal 1. Should significant correlation occur between sites at temporal lags other than a zero lag, further analysis should be performed to determine if the correlation could be consistent with the system being modeled, making a case for an lth order term at the temporal lag indicated, or if the correlation is merely a spurious result of the data. While this approach may be used if nothing is known of the nature of the system, creating spatial weights based on *a priori* knowledge of the system will provide a more causal approach.

Since the required weights should, if possible, reflect some physical aspect of the system being modeled, it is proposed that spatial weights are based on the lateral distance between the groundwater flow lines of the various monitoring wells. The natural flow at the area in question is known, but this natural flow is skewed by the operation of the groundwater recovery system pumps. Since significant data are not available during periods of non-use of the groundwater recovery system, flow lines must be constructed using information about the natural flow and how the nominal use of the groundwater

recovery system affects the natural flow. Flow lines are constructed normal to lines of equal potential. Potential is expressed as drawdown, or the amount the piezometric surface is reduced at a point. The natural piezometric surface is known (see Figure 4.33). The piezometric surface including the effects of nominal groundwater recovery system usage is this natural surface with the drawdown effect of the pumping wells subtracted off (see Figure 4.34).

The steady-state drawdown of the piezometric surface around a pumping well in an unconfined aquifer may be expressed as

$$h_0^2 - h^2 = \frac{Q}{\pi K} \ln(\frac{r_0}{r}) \qquad [8]$$

where h_0 is the original height of the water table, h is the height of the water table after pumping, Q is the flow rate of the well, K is the hydraulic conductivity (a measure of the ease of groundwater flow through the prevalent ground composition), r_0 is the distance from the pumping well to a point at which drawdown is considered negligible, and r is the distance from the pumping well to the point of drawdown interest [35:325].

The steady-state composite effect of several pumps operating is additive and may be expressed as

$$h_0^2 - h^2 = \sum_{i=1}^{n} \frac{Q_i}{\pi K} \ln(\frac{r_{0i}}{r_i}) \qquad [9]$$

where the total drawdown is written in terms of Q_i, the flow rate of well i, K is the hydraulic conductivity, r_{0i} is the distance from pumping well i to a point at which drawdown is considered negligible, and r_i is the distance from pumping well i to the point

of drawdown interest [35:326]. The first assumption made for applying this model is that the system modeled exhibits radial flow through an unconfined aquifer. Though the aquifer is unconfined, the flow pattern is more of a uniform flow field, since the piezometric surface in the Great Miami Aquifer has a gradient toward the Great Miami River. Viessman suggests, however, that if the piezometric surface is slight, the equations may still be employed [35:324]. It is also assumed the value of K is constant throughout the system. There exist many unquantified heterogenetic properties in the Great Miami Aquifer which are not modeled. However, across relatively large areas, the assumption is necessary. Though drawdown is not very sensitive to the specific value of r_0, r_0 is calculated using the Theis equation. The value is proportional to the level of operation of the pump. In other words, if a pump is not used, the distance at which the effect of the pump is negligible is zero.

After over-laying the effect of the nominal level of use of the groundwater recovery system on the natural piezometric surface, flow lines are graphically constructed normal to the lines of equal potential. The spatial weights, then, consist of the scaled lateral differences between flow lines constructed through the various well sites of interest. (For an example of this, see Figure 4.34.) These weights could change over time as the operation of the groundwater recovery pumps change, though the particular order neighbor one site is to another would require remaining constant. This approach is not taken, however, since the effect of the intervention is being modeled with the transfer function.

Determining which neighbors of the site of interest will be lth order neighbors is another task which involves some amount of subjectivity. But first, the site of interest must be determined. An adjacency matrix is constructed between all points. If groundwater flow from one well could conceivably affect the groundwater at another well, due to lying along the same flow line or dissipating pollutants which could traverse

flow lines, the first well is potentially influential to the second. The second well cannot be influential to the first, since the flow line emanating from it is traveling away from the first well. The nearest influential wells i of each well j is included in the adjacency matrix, element (i, j) as a one. This matrix describes the set of neighbors reachable in one step. The set of neighbors reachable in two steps is defined by squaring the adjacency matrix. The sum of the matrix and the squared matrix define the matrix of neighbors reachable in one or two steps. The column with the most entries, then, represents the well with the most influential neighbors. This site is then selected as the site of interest.

The first-order neighbors are selected as those which exhibit the smallest lateral distance (between flow lines) to the site of interest. Other sites will constitute the second-order neighbors. The weights of the members of each first and second-order neighbors are scaled to sum to unity so the stream of data for a particular lth-order neighborhood remains indicative of the values of its members.

With the data set sanitized to fit STARMA's needs, the regular ARMA procedure may be followed to arrive at the coefficients. Again, Pfeifer and Deutsch's three stage iterative procedure for space time modeling will be used to determine the most parsimonious model.

3.2.2.1 Identification of a STARMA Class Tentative Model

The identification of a STARMA class tentative model is similar to the process of identifying an ARMA class tentative model, except the combined time series will be considered in determining the values for p, q, λ, and m.

3.2.2.2 Estimation of STARMA Class Model Parameters

The estimation of a STARIMA class tentative model is performed similarly to the process of estimating the ARIMA class models, using IMSL subroutines for performing parameter estimation on all four parameter types [16]. Parameters relating to single lags

are interpreted as spatial in nature, while parameters estimating coefficients for lags which are evenly divisible by *s* are interpreted as temporal in nature.

3.2.2.3 Diagnostic Testing of a STARMA Class Model

The residuals of the fit model should again exhibit the behavior of white noise. All covariances should be zero, as well as autocovariances at non-zero spatial or temporal lags [26:43]. Tentative models may fail in the same two ways as ARIMA models. First, the model may inadequately represent the observed correlation of the process. If this is the case, the residuals will show significant correlation either spatially or temporally, which could potentially be modeled with additional spatial or temporal autoregressive or moving average parameters. Second, too many parameters may be included in the model, resulting in estimated coefficients which are not statistically different from zero [26:43]. A candidate model which fails neither test may be usable in short-term forecasting of the time series, again assuming no model with less calibrated parameters is equally worthy.

3.2.3 Intervention Analysis

An intervention is an identified event which can affect a time series. Modeling a particular intervention is only necessary if that intervention actually does influence the time series of interest. In this research, the proposed interventions consist of varying the usage of the groundwater recovery system pumps which, it has been suggested, has influenced the level of arsenic found in the groundwater at nearby sampling wells. Specifically, the increased use of western-most pumps appears to have the effect of increasing the presence of arsenic in the groundwater at areas to the south and south-west of the groundwater recovery system pumps (see Figure 4.34).

3.2.3.1 Intervention Transfer Function Model

To perform an analysis of interventions, Pankratz suggests a model of the form

$$Y_t = C + f(X_t) + N_t \qquad [10]$$

where Y_t is the time series, C is the constant term associated with the stochastic disturbance N_t (which may be autoregressive), and $f(X_t)$ is a transfer function which accounts for the effect of the intervention time series X_t on the time series of interest Y_t [25:148].

Since the effect of the intervention may depend upon both present and past values of X_t, it is common to write the intervention $f(X_t)$ in terms of the backshift operator as

$$f(X_t) = v(B)X_t \qquad [11]$$

Here, $v(B)$ is a set of weights known as the impulse response weights. These weights explain how a unit change in the intervention time series X_t is translated (transferred) to the time series of interest.

The combined effect of the C and the N_t terms has been accounted for in the sections on ARIMA and STARIMA modeling, but the effect of the intervening X_t time series representing changes in the level of pumping water from the groundwater recovery system requires form and magnitude. Several forms for the $v(B)$ are possible, but a general form of

$$v(B) = \frac{\omega(B)B^b}{\delta(B)} \qquad [12]$$

can account for a wide variety of intervention patterns [25:158]. In this form, the B^b term accounts for a lag time between the occurrence of an intervention and the start of that intervention's effect on the time series, thereby accounting for transient behavior in the effect of a pump on a site of interest. If $b = 0$, the effect begins immediately, while a $b = 1$ implies the effect to the intervention begins in the time period following the occurrence of the intervention. The $\omega(B)$ term accounts for unpatterned spikes (those not part of a decay pattern) as well as decay start-up values. In other words, given no decay, the $\omega(B)$ represents the entire effect of the intervention. The order of the $\omega(B)$ is h. The $\delta(B)$ term allows the effect of the $\omega(B)$ term to last, however. The $\delta(B)$ represents the decay pattern, with an order of r. For $r = 0$, decay is instant, $r = 1$ represents simple exponential decay, and higher orders of $\delta(B)$ can be used to account for compound exponential decays or for damped sine waves [25:159-160]. With the quantity of data available for this research, a $\delta(B)$ term is not likely to be obtained with confidence. For this reason, only a $\omega(B)$ term representing the spike effects through time the intervention has on the time series will be obtained. If the transient effect seen in the $\omega(B)$ term shows an obvious decay pattern, a $\delta(B)$ term could be constructed from it.

Multiple interventions may be handled by including M of the intervention time series in the model. The final form for the effect of all interventions imposed on the time series of interest is represented as

$$Y_t = C + \sum_{i=1}^{M} \frac{\omega_i(B)B^{b_i}}{\delta_i(B)} X_{i,t} + N_t \qquad [13]$$

However, if similar interventions are expected to take the same form, even with different magnitudes, these interventions may be modeled as a single intervention time series X_t taking into account the various levels of intervention. Such *a priori* modeling

takes into account knowledge of the system, which 1) increases confidence in the obtained transfer function model and 2) validates that the system is in fact operating as anticipated.

The orders (b, h, r) may be estimated based strictly upon the available or upon knowledge of the system being modeled. Identifying the form of the intervention effect by hypothesis (based on system knowledge) is desirable, however. The model may then be refined to include effects yet unknown about the system based upon the data, which leads to increased knowledge of the system. Strict use of data to identify the form of the transfer function may point to the use of an intervention transfer function which defies intuition. In this instance, the identification of an intervention effect which defies causal reasoning would not likely prove effective in forecasting the time series.

3.2.3.2 Intervention Transfer Function Model Building

There are several methods available for building the transfer function of an intervention into an ARIMA formulation of a time series. This research proposes an iterative method which is both data-based as well as analytical in order to form a meaningful transfer function which also corresponds to the actual data. First, the theoretical effect the groundwater recovery system has on the groundwater at the various monitor well sites throughout the time of the data collection should comprise the X_t intervention time series. Construction of this time series may be accomplished much as the determination of spatial weights has been.

The composite effect of the groundwater recovery system on a site of interest is analyzed over time. The effect over time may then be scaled to reflect zero at baseline time periods (no interventions), any other scaling applied to the rest of the time series values would only be reflected in the calibration of the transfer function in the univariate case. However, no other scaling is accomplished in an order to compare the effect of

pumping from one monitoring well site to another. Similarly, in applying the scheme to the spatial-temporal model, care is required to ensure a constant scaling, and thus, a homogeneous transfer function calibration.

Knowledge of the origin of the arsenic contaminants imposed on this time series could also aid in the construction of a homogeneous transfer function which could be used as part of any of the univariate ARIMA models, or as part of a spatial analysis. Though little is known of the origin of the arsenic currently, knowledge pertaining to direction and concentration could be gained through the univariate ARIMA modeling, which could then be applied to later model refinements.

The iterative process employed in this research is used to obtain reasonable estimates of the effect of the intervention based on what is known of the system, while allowing the effect of the intervention to be modified by the data. The process follows: First, an ARIMA model is fit to the time series. This model may be fit to the entire time series or to strictly pre-intervention data. The result of fitting to the entire time series is that the effect of the intervention is already being accounted for by the time series modeling. This could lead to highly erred estimates, particularly if the effect of the intervention is significant. The danger in modeling strictly pre-intervention data is that the number of usable data points decreases. These estimates may also be significantly incorrect. Too much emphasis is not placed on these estimates, however, as they will be updated.

The next step is to estimate the $v(B)$ weights based on the residuals of the estimated ARIMA model just fit. If only pre-intervention data values are used in the preliminary ARIMA modeling, these "residuals" should consist of the difference between forecast values and actual data values through the intervention. The orders (b, r, h) may be determined then by comparing the estimated set of weights, or impulse response function, to theoretical impulse response functions [25:184-189].

With the determination of a candidate model, the model's parameters may be estimated, and the model is checked for fit. Much as Pfeifer and Deutsch's three stage iterative procedure for STARMA modeling, if a poor fit is assessed, the process is accomplished again, except that here, a better knowledge of the actual ARIMA parameters are used to enhance the preliminary estimate of the ARIMA model, thereby producing residuals during the intervention period which better reflect the actual effect of the intervention. This fine tuning is accomplished iteratively until the final model fit is adequate, or until the marginal improvement of performing another iteration becomes sufficiently small.

3.3 Delineation of Questionable Data Values

The ICM algorithm allocates each pixel individually to a subregion based on its posterior probability of belonging to a particular subregion given its value and the values of the neighboring pixels. The algorithm requires two assumptions: 1) given the true gray value of each pixel, the observed gray value of that pixel is independent of observed gray values of other pixels with the same true gray value and 2) the true image is a realization of a locally dependent Markov random field with a probability distribution that is dependent on some β.

Each pixel is allocated, then, on the basis of $\text{pr}\{\mathbf{Z}_j = \mathbf{z}_j \mid \mathbf{x}, \mathbf{z}_{(j)}\}$ where $\mathbf{z}_{(j)}$ is the vector of true colors of the pixels around, but not including j.

For a binary problem, allocating pixels to one of two subregions, such as polluted versus unpolluted pixels, MacLachlan proposes a logit specification:

$$\text{pr}\{Z_{1j} = 1 | \mathbf{z}_{(j)}\} = e^{\beta u_{1j}} / (e^{\beta u_{1j}} + e^{\beta u_{2j}}) \qquad [14]$$

where u_{1j} is the number of j's neighbors belonging to subregion 1 and u_{2j} is the number of j's neighbors belonging to subregion 2. Based on this, it follows that

$$\log[\mathrm{pr}\{Z_{1j} = \mathbf{x}, \mathbf{z}_{(j)}\} / \mathrm{pr}\{Z_{2j} = 1 | \mathbf{x}, \mathbf{z}_{(j)}\}] = -(x_j - 0.5) / \sigma^2 + \beta(u_{1j} - u_{2j}) \qquad [15]$$

Assuming equal posterior probabilities, this reduces to the manageable form

$$x_j < \frac{1}{2} + \beta\sigma^2(u_{1j} - u_{2j}) \qquad [16]$$

As seen, a β of 0 produces a non-contextual formulation, while increasing β increases the contextual bias. The algorithm is performed iteratively, however it has been shown to converge very quickly. There is a trade-off between β and σ^2 shown in this equation The σ should be small enough to prevent greatly overlapping regions, and β will need to be adjusted, then, to affect some real change in the desired amount of contextuality present..

In performing the ICM algorithm to values located in a regular square grid, proximity may be considered a factor when determining the allocation of a pixel. Specifically, the distance between the pixel in question and its first order neighbor on an ordinary square grid is unity, the distance between that pixel and its second order neighbors is $\sqrt{2}$. Taking this into account, an inverse relationship between distance and importance in determining the allocation of some central pixel may be assumed. Normalized weights are obtained by scaling the above relationship so the sum of all neighbors of a pixel is still 8, as it is for any interior pixel. The weights obtained in this manner are 1.1716 for first order neighbors and 0.8284 for second order neighbors. The sum over all of an interior pixel's neighbors is 4 x 1.1716 + 4 x 0.8284 = 8 and the first

order neighbors are 1.1716 / 0.8284 = $\sqrt{2}$ times as important as second order neighbors in determining allocation of a pixel.

In this weighted ICM algorithm, the weighted difference in the number of pixels is scaled appropriately and added it to 0.5 to determine a comparison value for the standardized gray value. For instance, if a pixel has 2 first order neighbors assigned to subregion 1 and 2 assigned to subregion 2 and 2 second order neighbors assigned to each, as well, the context makes no difference, as the weighted sum is 4, meaning that 0 is added to 0.5 to comprise the "compare" value. In the extreme case, however, 4 first order neighbors allocated to subregion 1 and 0 second order neighbors allocated to subregion 1 would, in the unweighted algorithm, result in the same conclusion that context makes no difference. However, using a weighting scheme, the difference between u_{1j} and u_{2j} is 1.3728 and the algorithm will increase the probability that the pixel belongs to subregion 1 (the exact amount depends upon β and σ).

This research effort does not use a regular square grid. As described in chapter 2, kriging could be accomplished to "fill-out" such a grid, but doing so imposes an estimation error to every pixel other than the pixels for which ground truth is known. Performing a pixel delineation technique to such a data set would be ill-advised.

Instead, the ICM method may be modified to accommodate an irregular grid. The modification is similar to the method of dealing with the edge pixels of a regular grid. These edge pixels don't have enough neighbors to sum to eight, but the relative importance between first and second order neighbors remains the same. The only difference with these pixels is that, since nothing is known about the value of pixels not on the grid but adjacent to the edge pixel, the possible amount of contextuality applied is less than that for interior points.

Similarly, all neighbors of a pixel on an irregular grid may be weighted relative to each other and normalized to a value relative to the sum of the inverse distances of all

neighbors considered. In this manner, for example, an undue amount of contextuality may not be artificially imposed upon a pixel with only one neighbor. This method will provide more meaning and consistency to the value of β since the sum of the inverse distances provide a stable meaning as opposed to the relatively arbitrary value of 8 assumed in the original algorithm. The distances used for this analysis may be the same as those which are used to determine the spatial weights for the STARMA analysis.

IV. Results and Analysis

4.1 Introduction

The applications of methods employed for performing for the temporal and spatial-temporal modeling of the arsenic pollution are presented. The algorithms used and the applications of these algorithms are described. First, the univariate temporal modeling of each of the arsenic monitoring wells, to include intervention analysis, is given in section 4.2. Spatial-temporal modeling of the system including intervention analysis is described in section 4.3. A comparison of strict temporal modeling and spatial-temporal modeling is presented in section 4.4.

4.2 Temporal Modeling of Pollutants with Intervention Analysis

In this section, each of the well sites of interest is modeled separately, with no interaction or helpful correlation assumed between sites. The algorithm used in the development of each site model is presented in section 4.2.1. The data used in the application of this algorithm is shown in section 4.2.2. The actual modeling effort is located in section 4.2.3.

4.2.1 Algorithm for the Temporal Modeling of Pollutants with Intervention Analysis

This section describes the algorithm used to perform the temporal modeling of the arsenic contamination, to include the application of intervention analysis. The algorithm is based on the methodology discussed in sections 3.2.1 and 3.2.3. The algorithm is iterative in order to develop the best possible ARMA and transfer function model possible.

The algorithm consists of six distinct steps. 1) The time series data not including observations at periods affected by interventions is collected. 2) The autocorrelation

function and the partial autocorrelation function are determined, and the time series and the autocorrelation and partial autocorrelation functions are plotted for analysis. The analysis determines if differencing the data is necessary for obtaining stationarity prior to ARMA modeling. If differencing is required, the data are differenced and step 2 is repeated. If differencing is not required, the values of p and q to be used in modeling this stationary process are determined. 3) Autoregressive and or moving average coefficients are calculated. 4) The autoregressive and moving average coefficients are assumed to represent the entire noise series (time series minus effect of interventions) in order to calculate the impulse response weights of the interventions. 5) If first time through, or if stopping criteria are not met, form a new time series by subtracting the effect of the calculated transfer function impulse response weights from the time series Return to step 2 with this new time series and repeat process until both the orders of p and q are the same as the previous iteration and the coefficients associated with the autoregressive parameters, the moving average parameters, and the impulse response weights are within some specified tolerance from the previous iteration. 6) Translate the final impulse response weights to the transfer function parameters. If the model fit is adequate, the model of the form

$$Y_t = C + \sum_{i=1}^{M} \frac{\omega_i(B)B^{b_i}}{\delta_i(B)} X_{i,t} + N_t \qquad [17]$$

where N_t is some ARMA(p,q) process is complete for the monitor well site of interest.

4.2.2 Data for Temporal Modeling of Pollutants with Intervention Analysis

The groundwater recovery system has been in operation since August 27, 1993. Standard analysis of many potential contaminants was performed in early September 1993, during which increased levels of arsenic were discovered in wells south of the

western-most pumps of the groundwater recovery system. To that point, significant levels of arsenic had not been discovered in any wells near the groundwater recovery system. The exact source of the arsenic is unknown. However, since the use of the groundwater recovery system has introduced a west-to-east component to the prevailing north-to-south groundwater flow, the source is generally considered to exist somewhere to the west of the monitoring wells which have seen an introduction of arsenic. To date, there has not been an increase in arsenic levels at the sites of the groundwater recovery wells.

Data has been collected on increased arsenic contamination levels found in monitoring wells since September 1993. The first three observations were taken during sequential weeks, then only one observation was recorded until December 8, 1993. Since that time, weekly observations have been made. Though observations are made on various weekdays, they are generally in the middle of the week. In order to assume a homogeneous data gathering, therefore, the assumption is made that observations are collected on each Wednesday. This assumption is not unrealistic as the majority of data points were gathered on or near Wednesday. Also, there are no seasonalities expected in the data with periods less than a week.

Most of the samples collected were analyzed for arsenic both before and after filtering sediment from the groundwater. The analysis here is accomplished on the unfiltered observations since this data provided a more complete database (analyses on filtered groundwater were not completed after August 22, 1994).

On many observation dates, analysis was accomplished on more than one sample per monitor well. When this is the case, the relevant observations are averaged to obtain a single value for the date. A high variance among observations taken at a single well on a specific date is seen in certain cases, but generally, the observations are relatively consistent.

The data for most of the time at six of the seven wells from which observations are being made is not exact at arsenic levels below a threshold of 10.0 micrograms per liter (μg/l). Though undesirable, 10.0 μg/l is used as the level of arsenic contamination in these cases, as the percentage of observations using this threshold is so large. Since arsenic concentrations above 10.0 μg/l are recorded as well, particularly during the "intervention" of increased pumping rates in the groundwater recovery system pumping wells, use of these values is not expected to significantly affect the analysis. When observations are made with a threshold lower than 10.0 μg/l, the exact value is used. However, one well site, well 2900, and samples recorded from two of the groundwater recovery system wells, wells 3924 and 3925, exhibit nearly no values above the threshold of 10 mg/l. These wells are not modeled with the univariate ARMA algorithm but may be used in the STARMA algorithm.

Arsenic data at monitor well 2636 used a higher threshold of 100.0 μg/l in approximately fifty percent of all observations. Since no values were recorded above 100.0 μg/l at this well, using 100.0 μg/l for these observations would misrepresent values at the well. Since the data are not complete, this well is not modeled using the univariate ARMA algorithm. Only applying a non-zero spatial weight when a usable data value is available, however, will allow this data to be used in the STARMA modeling.

4.2.3 Temporal Modeling of Pollutants South of the Groundwater Recovery System

Applying the FORTRAN code in Appendix A allows the user to proceed step-by-step through the process detailed above. The procedure is followed for the three wells with relatively meaningful data, wells 2128, 2548, and 2625. The remainder of this section will detail the development of the intervention input time series used to model the transfer function. Analysis and results from the modeling of the three applicable wells are found in sections 4.2.3.1, 4.2.3.2, and 4.2.3.3, respectively.

The time series selected as the intervention input time series should reflect the relative effect the intervention has on each point of interest through time. One applicable measure is the amount of drawdown a particular pumping scenario has at the site of interest. Since the baseline case is the pumping scenario in which all groundwater recovery system pumps are collecting water at some nominal rate (due to a lack of data available during periods when no pumping occurs), the difference in drawdown between a particular pumping scenario at a site of interest and the drawdown due to the nominal pumping scenario at the site will determine the magnitude of the intervention time series for the site at a given time. The drawdown due to the pumping wells could be combined with the natural flow of the groundwater to construct the intervention time series. However, the natural flow is not an intervention and is therefore not included in the time series. Since drawdown is a function of a site's distance to each of the pumping wells, each site will have a unique intervention input time series.

From chapter 3, the combined effect of pumping wells in an unconfined aquifer is

$$h_0^2 - h^2 = \sum_{i=1}^{n} \frac{Q_i}{\pi K} \ln\left(\frac{r_{0i}}{r_i}\right) \qquad [18]$$

Solving this equation for the drawdown, $s = h_0 - h$ yields

$$s = h_0 - \sqrt{h_0^2 - \sum_{i=1}^{n} \frac{Q_i}{\pi K} \ln\left(\frac{r_{0i}}{r_i}\right)} \qquad [19]$$

where h_0 is the thickness of the aquifer given a fully penetrating well. Even though the pumping wells are not fully penetrating, the estimates obtained should be reasonable considering the goal of obtaining relative magnitudes of interventions. Q_i is the pumping

rate at the *i*th pump. The groundwater recovery system pumps have operated at levels from 300 to 550 gallons per minute. The value used for hydraulic conductivity, K, is 387 feet per day, the median value employed in drawdown models as given in the South Plume Groundwater Recovery System Pump Test Work Plan [34:2-12]. r_i is the distance from the *i*th pump to the site of interest and r_{0i} is the distance at which effects of the pumping well are considered negligible. Values for r_{0i} are determined using principles of Theis, who developed flow relationships for unsteady flow. This research utilizes these relationships for unsteady flow in determining the distance at which drawdown is negligible, but incorporates the steady-state flow relationships given above to determine actual drawdown. The equation for unsteady flow drawdown is

$$s = \frac{Q^i}{4\pi K h_0} W(u) \qquad [20]$$

where $W(u)$ is known as the well function of u, given as the infinite series

$$W(u) = -0.577216 - \ln(u) + u - \frac{u^2}{2 \times 2!} + \frac{u^3}{3 \times 3!} \cdots \qquad [21]$$

with u being defined as

$$u = \frac{r^2 S}{4 K h_0 t} \qquad [22]$$

where S is the specific yield, assumed in this case to be 0.2, and t is the time since pump operation. A value of three days is assumed for time, since there are generally at least

three days between a pump operation scenario change and the collection of samples for arsenic testing.

To solve for r_{0i}, equation 20 is solved for $W(u)$ using what is considered a negligible value of s. Since there are five pumping wells adding to the drawdown at any given point, this research assumes a negligible value for the entire groundwater recovery system to be one half of an inch. One fifth of that, 0.008333 feet, is then considered negligible for a single pump. Since solving the well function of u for u is not practical, it is generally assessed from a table. Figure 4.1 shows the relationship between u and $W(u)$ using 14 terms of the infinite series. The figure is accurate for all required values of $W(u)$.

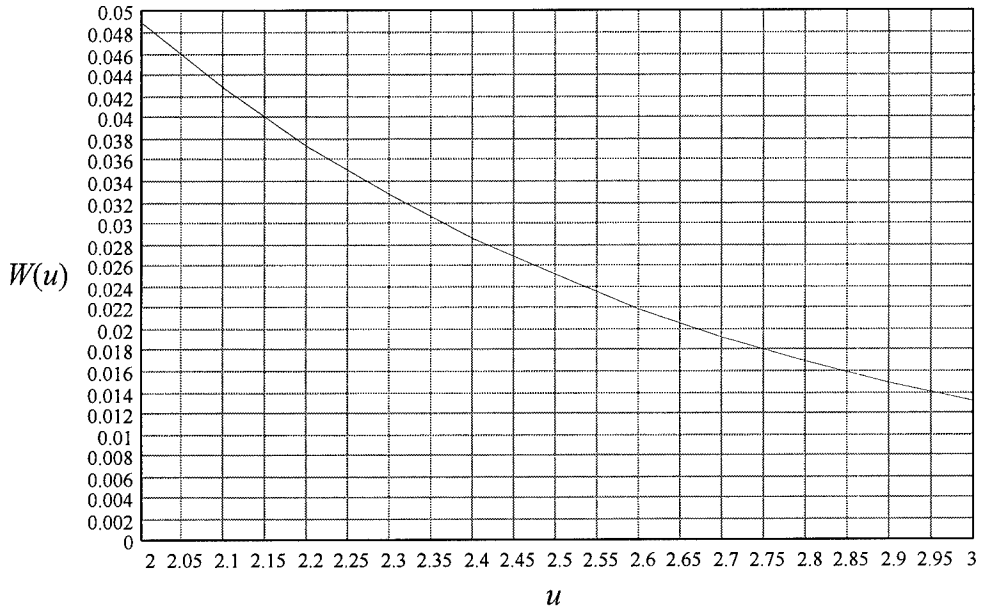


Figure 4.1 Values of u Corresponding to Levels of $W(u)$

After a value of u is determined, equation 22 is solved for r which becomes r_{0i}, the distance at which the effect of the pumping at a well site is negligible. The value of r_{0i} is considered zero for a pump which is not in operation, but this modeling yields values

from 1496 feet at 300 gallons per minute to 1636 feet at 550 gallons per minute, the range of values required in this research.

Application of equation 19 gives predicted drawdown for any pumping scenario. The values obtained using this method were consistent in form with those produced in the South Plume Groundwater Recovery System Pump Test Work Plan, except approximately 7.5 times larger [34:2-12]. Therefore, all values of s, the drawdown at a site of interest due to a particular pumping scheme are scaled to be consistent with those predicted by FERMCO. This scaling will not affect the temporal analysis, except to make the magnitude of the transfer function weights more meaningful.

The pumping schemes employed at the groundwater recovery system are shown as a matrix in Figure 4.2. The top row shows the level of operation in gallons per minute for the most western groundwater recovery system pump well 3924 through the six distinct pumping scenarios. The other four groundwater recovery system pump wells are similarly displayed.

$$\begin{matrix} \text{Recovery} \\ \text{well} \end{matrix} \begin{bmatrix} 400 & 300 & 450 & 400 & 0 & 0 \\ 400 & 300 & 0 & 0 & 450 & 450 \\ 400 & 300 & 550 & 550 & 550 & 550 \\ 400 & 300 & 0 & 0 & 0 & 0 \\ 400 & 300 & 500 & 550 & 550 & 0 \end{bmatrix}$$

Pumping scenario

Figure 4.2 Levels of Operation in Gallons per Minute of 5 Recovery Wells during 6 Pumping Scenarios

Employing equation 19 and scaling the drawdowns yield the drawdowns in feet at any point of interest. A contour plot of these drawdowns for the nominal pumping scenario of 300 gallons per minute per well is given in Figure 4.3. The five focus points

at the top of the figure represent the five pumping wells, and the contours at these points are contours of the greatest magnitude of drawdown.

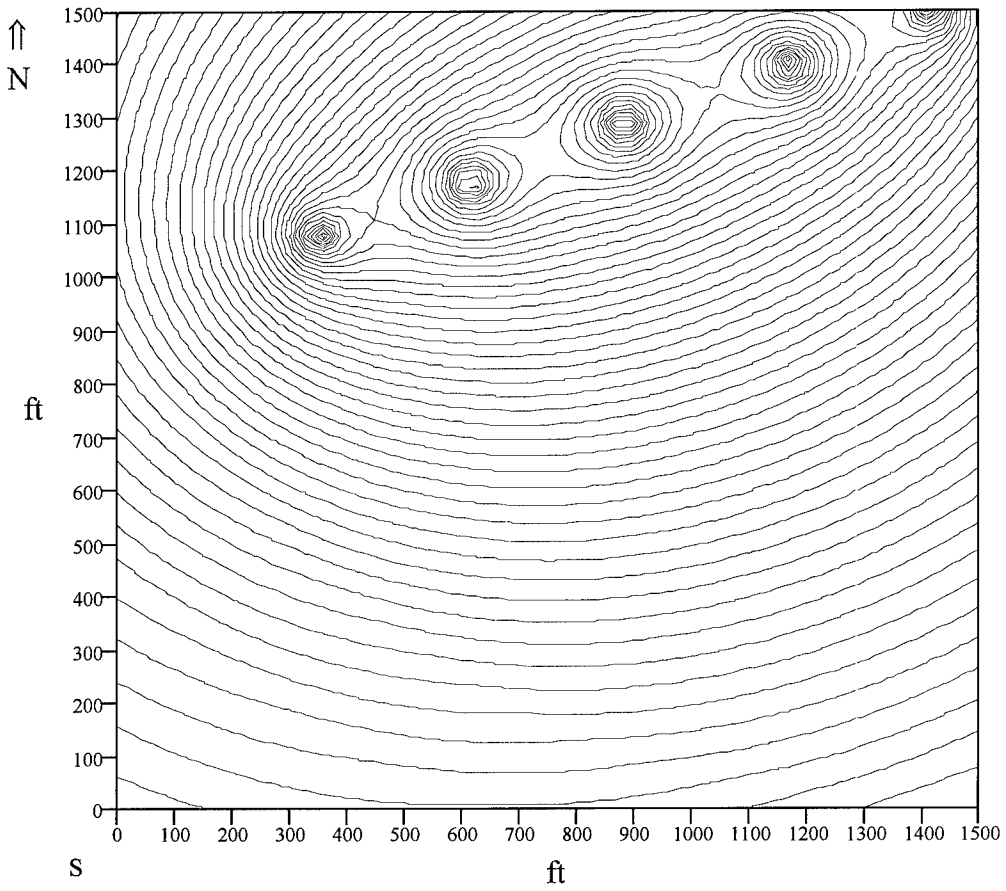


Figure 4.3 Contours of Predicted Equal Drawdown during Nominal Pumping

The predicted drawdowns at each of the seven monitoring wells through each of the six pumping scenarios are shown in Figure 4.4. The first five rows represent strict monitoring wells which are located generally south of the two most western pumping wells shown in Figure 4.3. The last two rows represent the drawdown at the most western pumping well and the pumping well adjacent to it, as shown in Figure 4.3.

Monitor well

0.258	0.174	0.201	0.186	0.179	0.168
0.145	0.091	0.116	0.108	0.119	0.104
0.31	0.213	0.24	0.221	0.207	0.193
0.192	0.126	0.151	0.14	0.141	0.133
0.546	0.385	0.404	0.381	0.379	0.328
1.276	0.901	1.155	1.04	0.437	0.383
1.459	1.026	0.574	0.553	1.337	1.222

Pumping scenario

Figure 4.4 Predicted Drawdown in Feet at 7 Monitor Wells during 6 Pumping Scenarios

Next, the drawdowns are scaled to reflect the nominal case of all pumps operating at 300 gallons per minute. This nominal pumping scenario is shown as the second scenario in the figures since the first scenario, 400 gallons per minute per pump was an intervention which introduced arsenic flow to the monitoring wells. Sufficient data were not collected at this pumping scenario to perform meaningful analysis, but the intervention effect is shown for comparison. Figure 4.5 shows the final intervention input time series elements. The intervention input time series for each monitor well is constructed with the value for each pumping scenario for as many time periods as that pumping scenario occurred.

Monitor well

0.083611	0.0	0.026672	0.01132	0.004572	0.0
0.054236	0.0	0.024929	0.017494	0.028614	0.013704
0.097334	0.0	0.027031	0.008724	0.0	0.0
0.066048	0.0	0.025008	0.013254	0.014831	0.006425
0.160906	0.0	0.018272	0.0	0.0	0.0
0.374547	0.0	0.253342	0.138981	0.0	0.0
0.433068	0.0	0.0	0.0	0.310877	0.195461

Pumping scenario

Figure 4.5 Intervention Input Time Series Elements for 7 Monitor Wells during 6 Pumping Scenarios

If the intervention effect was less than that of the nominal case, 0.0 is used as the intervention effect. This is done since significant arsenic presence (specifically values much greater than the usual threshold of 10 μg/l) is not common during the nominal case, and a negative intervention would suggest arsenic levels to exist less than 10 μg/l, which is not generally tested. The top three rows in Figure 4.4. are intervention time series values for the three monitoring wells for which the univariate ARMA modeling is accomplished. The 0.0 values in the second column constitute the first 26 observations of each intervention time series. Values from the third column are located in the 27th through the 30th positions in the intervention input time series; the fourth column values are used in positions 31 through 34; the fifth column values are used in positions 35 through 40; the sixth column values are used in the 41st, and last, position.

4.2.3.1 Program Flow and Univariate Temporal Modeling of Well 2128

Monitor well 2128 is located approximately at coordinates (350,450) on Figure 4.3. The levels of arsenic concentration found at well 2128 are shown in Figure 4.6. The data points prior to time zero correspond to measurements made prior to the nominal pumping rate of 300 gallons per minute per pump. The points, though significant, are not consistent enough to perform temporal modeling. The intervention input transfer function weights are given in row 1 of Figure 4.5. The difference between two time periods is a week. Time period 1 corresponds to December 8, 1993. The last time period, $t = 41$, corresponds to September 13, 1994.

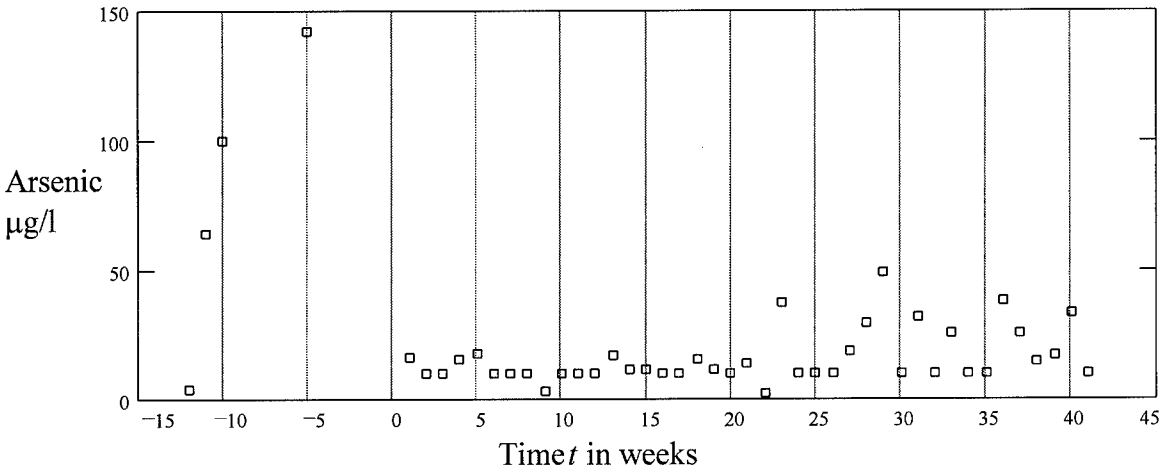


Figure 4.6 Arsenic Level Versus Time for Well 2128

The time series used in the temporal modeling of the arsenic is the same as shown above, except using only positive t values. This time series is shown in Figure 4.7.

The first prompt in the interactive ARIMA and causal intervention transfer function program is how many observations (starting at $t = 1$) of the time series occurred prior to any intervention. The first estimate for autoregressive or moving average terms are based on this number. In this case, the first 26 time periods occurred at the nominal pumping scenario of 300 gallons per minute per well. However, if the effect of the interventions are not strong, indicating that the entire time series occurred prior to the intervention may reduce the number of iterations required to converge. In this case, using 26 pre-intervention values allowed the program to converge to four decimal places in 12 iterations, while using all 41 observations immediately allowed the program to converge in 11 iterations.

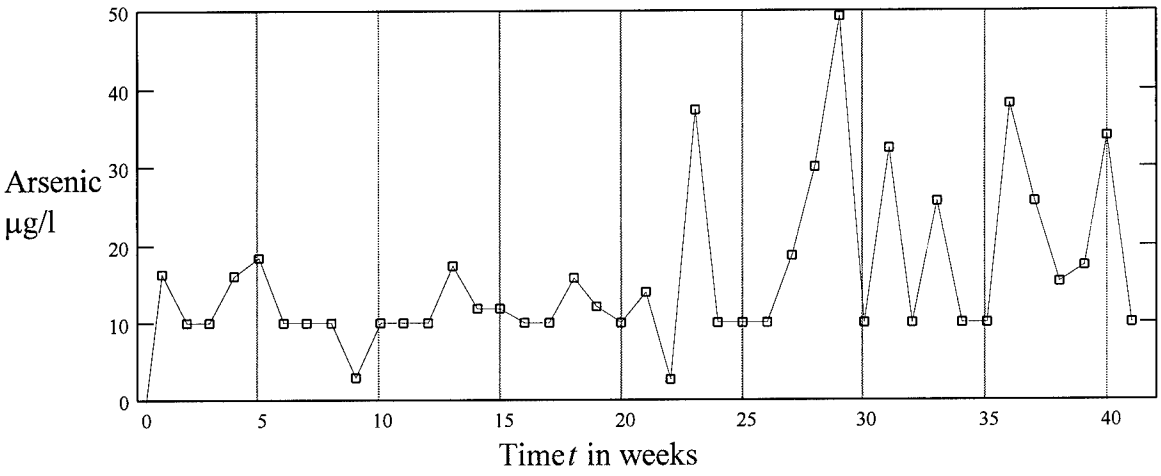


Figure 4.7 Time Series for Temporal Modeling of Well 2128

The next prompt is for which well is being modeled. There are three choices given: well 2128, 2548, or 2625, corresponding to the three wells for which consistent, meaningful data were collected throughout the period from December 8, 1993 though September 13, 1994.

The sample autocorrelations and sample partial autocorrelations are then calculated. The time plot, and plots of the autocorrelations and partial autocorrelations are displayed. The autocorrelation plot shows the standard errors for the various lags, as well, for determining the significance of the autocorrelations. The partial autocorrelations are compared against their standard errors, as well, as determined by $s = n^{-1/2}$ [25:38]. Actual plots from the program are shown in Appendix B, but plots with greater fidelity are shown in Figures 4.8 and 4.9 for the autocorrelations and partial autocorrelations, respectively.

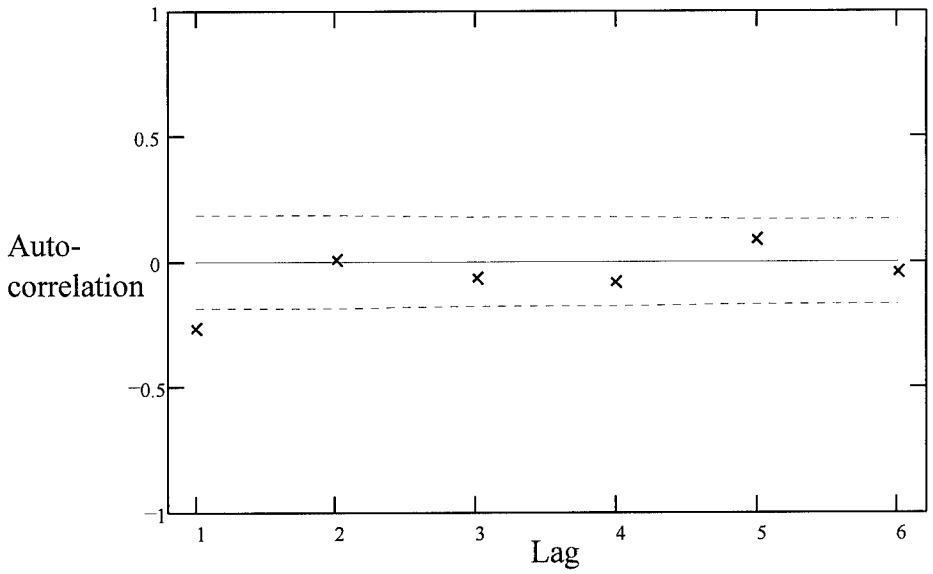


Figure 4.8 Autocorrelations for the Non-Intervention Period for Well 2128

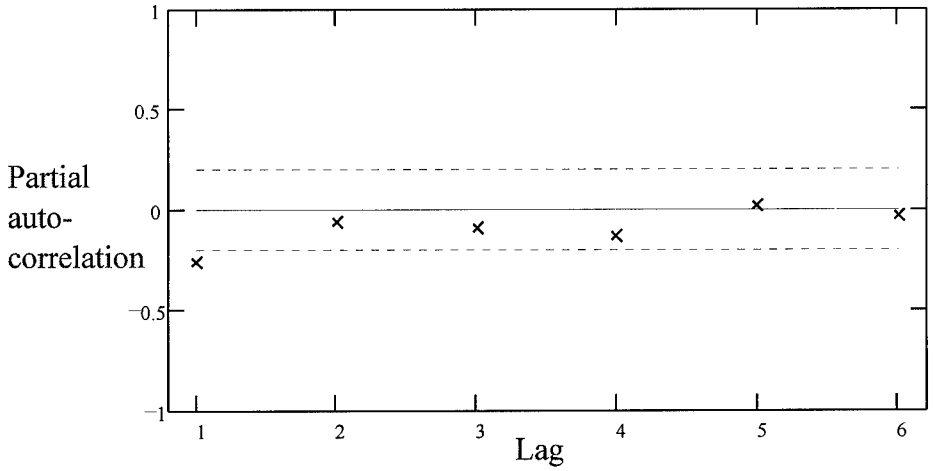


Figure 4.9 Partial Autocorrelations for the Non-Intervention Period for Well 2128

The user is then asked whether differencing should occur. If the user wishes to perform differencing, the differencing occurs on the time series as well as the intervention input time series [25:177]. Inspection of this time series suggests that the series is approximately stationary in the mean prior to any intervention, and the autocorrelation

function decaying quickly toward zero supports this conclusion, as well. Therefore, no differencing is warranted.

Next, the first tentative model form is established. This preliminary estimate should contain as few parameters as possible, since the estimate is not likely to reflect well the final selected model. Since both the autocorrelation and partial autocorrelation functions have one significant spike, an AR(1) or MA(1) model could be selected. Since neither the autocorrelation nor the partial autocorrelation functions show clear exponential decay toward zero, the choice between AR(1) and MA(1) is subjective. The AR(1) model is selected in this instance.

Parameter coefficients for the selected model are then calculated. These estimates are then used as prewhitening parameters in the transfer function computation, which is accomplished on the entire time series, using the intervention time series as the input time series and the entire (41 observations) time series as the output time series. The impulse response weights are calculated and displayed, along with the autoregressive and moving average parameter coefficients. Only four impulse response weights are calculated for this research since it is not expected that the effect of an increased level of pumping would be significant after three weeks, the four weights being applied to the level of intervention at zero time (immediately) and one to three time periods later. The first iteration for well 2128 yields autoregressive coefficient ϕ_1 = -0.2653 and impulse response weights ν_0 = 677.544, ν_1 = 669.263, ν_2 = 522.313, and ν_3 = 297.427. Since the intervention input time series reflects the difference in drawdown at a site of interest from the nominal pumping scheme to the various other pumping schemes measured in feet, if the drawdown calibrations are correct, the physical interpretation of the impulse response weights is the expected increase in the level of arsenic concentration measured in μg/l due to a one foot increase in the drawdown at a site from the drawdown level at the nominal pumping scheme of 300 gallons per minute per pumping well. Again, since the

prewhitening parameter coefficients were computed using only the non-intervention time series, it is not expected that these first obtained impulse response weights will accurately reflect the final model.

Finally, the program asks whether the user would like to perform another iteration. Iterations in this research were accomplished until prewhitening parameter coefficients converged to one within ten-thousandths on successive iterations. If the user specifies another iteration should occur, the predicted noise series which consists of the original time series minus the expected impact of the intervention based on the impulse response weight estimates is used as the time series from which the prewhitening estimates are based. The transfer function then uses the original time series with the newly estimated prewhitening parameter coefficients to determine the new impulse response weights. The process is then repeated until the user is satisfied with the convergence of the parameter coefficient estimates. The number of iterations required for convergence varies, but is around twelve for the time series used in this research.

In order to assist the convergence of the algorithm, it is useful to select consistent model forms from one iteration to the next. For instance, the second iteration for the modeling of well 2128 shows exponential decay in the autocorrelations, but two spikes in the partial autocorrelations, indicating an AR(2) model form. This determination requires some amount of judgment, since actual autocorrelation and autocorrelation functions rarely exhibit the precise shapes of their associated theoretical functions. The AR(2) form selected here is consistent with the first AR(1) model in that it is an extension of the first. The concern is that switching between AR and MA models on successive iterations may hamper solution convergence of the algorithm.

The final iteration of the temporal modeling for well 2128 displays autocorrelations and partial autocorrelations shown in Figures 4.10 and 4.11, respectively.

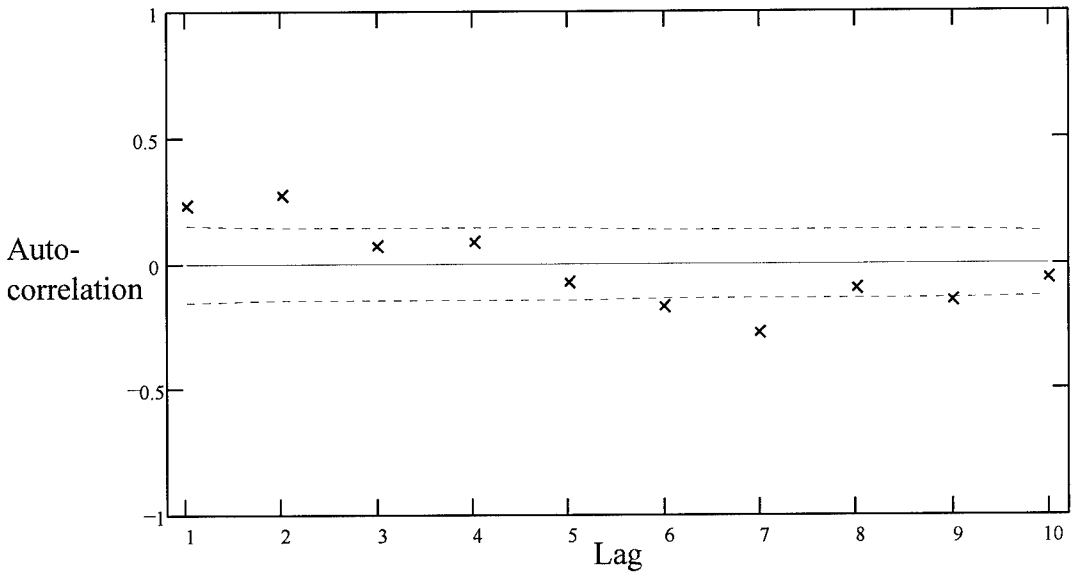


Figure 4.10 Autocorrelations for Well 2128

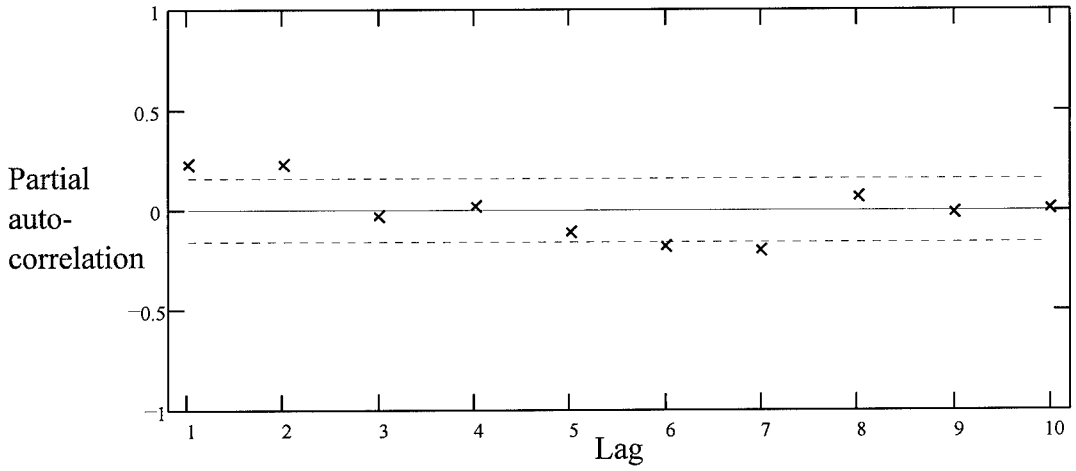


Figure 4.11 Partial Autocorrelations for Well 2128

The autocorrelation plot is interpreted as a sinusoidal pattern, likely dampening, though enough data are not available to make that judgment. The partial autocorrelation plot is judged to have two spikes at lags one and two. This combination of autocorrelation and partial autocorrelation plots suggests an AR(2) process and the final autoregressive coefficients estimated are $\phi_1 = 0.1776$ and $\phi_2 = 0.2344$. However,

different interpretations of the autocorrelation and partial autocorrelation plots are possible. The first two autocorrelations could be considered spikes, thereby suggesting a moving average process, MA(2). The algorithm was run fitting this tentative model. The MA(2) model parameter coefficients are approximately equal to the AR(2) coefficients, but opposite in sign. The problem with the fit is in the estimation of the transfer function impulse response weights. The AR(2) weights behave as expected, dying completely out by the fourth term: $\nu_0 = 522.875$, $\nu_1 = 599.798$, $\nu_2 = 392.735$, and $\nu_3 = -1.473$. The MA(2) weights, however, suggest a strong influence even at the third week after the intervention occurs: $\nu_3 = 98$. Since the MA(2) model does not provide any variance reduction over the AR(2) model, MA(2) is no longer a candidate model.

Another possible interpretation of the autocorrelation plots is that of a mixed process. Several mixed process models were fit, but the ARMA(2,2) provided the most insight. In only two iterations of fitting the ARMA(2,2) model, it became apparent that only two of the four estimated coefficients would be at all significant, the ϕ_2 term representing the two-lag autoregressive parameter and the θ_1 term representing the one-lag moving average term. The estimated coefficients are $\phi_2 = 0.263$ and $\theta_1 = -0.138$ with transfer function impulse response weights of $\nu_0 = 510.920$, $\nu_1 = 600.550$, $\nu_2 = 397.387$, and $\nu_3 = -18.681$ and a constant term of 5.50. The model shows very similar response weights as the AR(2) model estimated, and the random shock variance of the estimated models is nearly equal (120.6). Given the equal model fit, the AR(2) model is selected as the most parsimonious since only two coefficients are estimated, as opposed to the ARMA(2,1), $\phi_1 = 0$ model, in which the ϕ_1 term still forces the loss of a degree of freedom.

After convergence in the program, the final residuals are plotted along with their autocorrelations and partial autocorrelations. Figure 4.12 shows the residual plot, while Figures 4.13 and 4.14 show residual autocorrelations and residual partial autocorrelations, respectively.

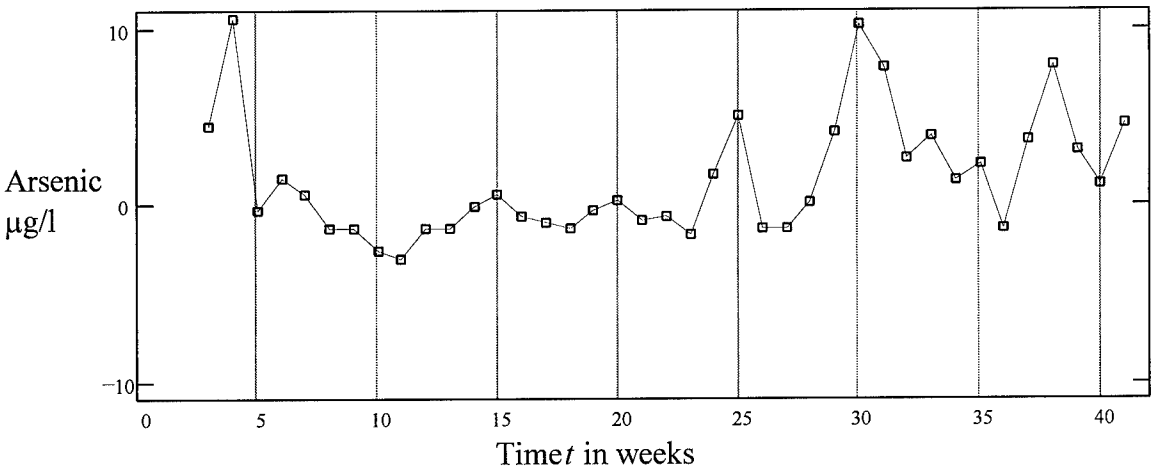


Figure 4.12 Residuals for Well 2128

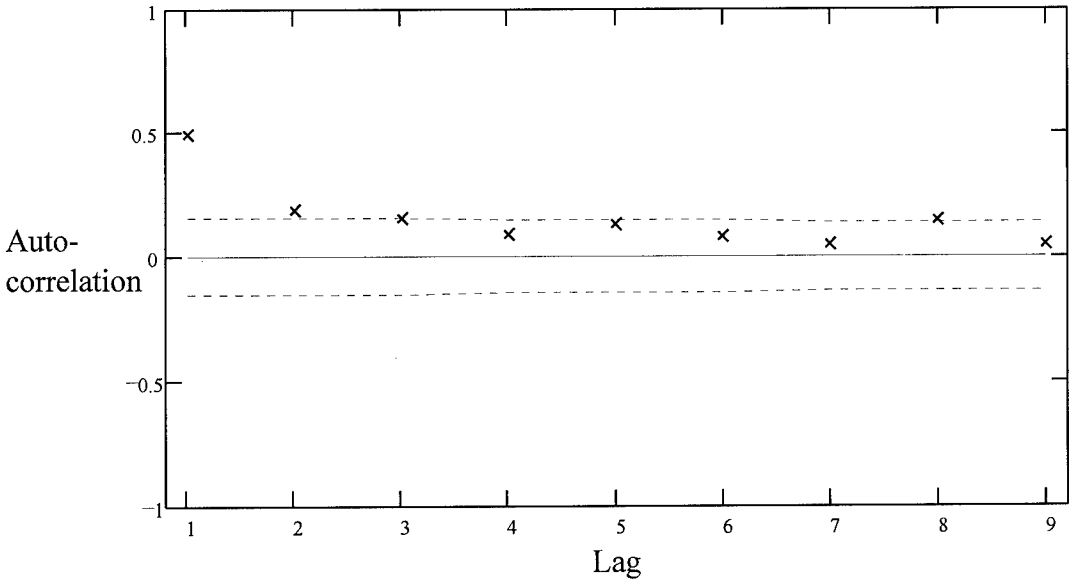


Figure 4.13 Residual Autocorrelations for Well 2128

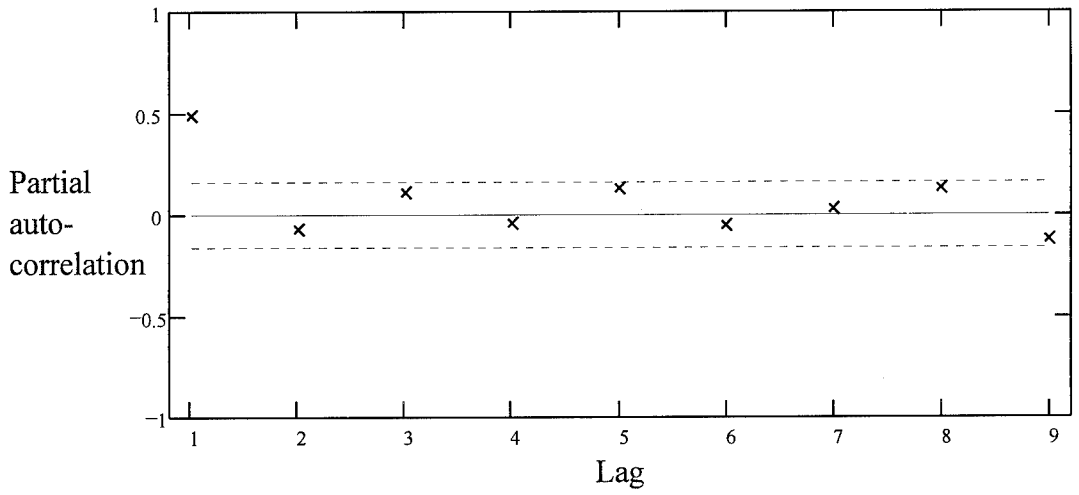


Figure 4.14 Residual Partial Autocorrelations for Well 2128

Figure 4.12 suggests the variance is greater during periods of intervention (*t* greater than 26) than periods prior to the intervention. This is likely due to two reasons. First, the arsenic concentration threshold of 10.0 µg/l produced many concentration values of 10.0 µg/l, thereby reducing the variance for the pre-intervention series. Second, the data series reflecting a nominal level of arsenic concentration is likely to have a small variance compared to the same data with additional arsenic imposed upon it from some outside source. This is likely because the amount of additional arsenic imposed on the system contain its own variability. The autocorrelations and partial autocorrelations similarly suggest that additional parameters could be warranted. However, it has been shown that additional parameters do not aid this model. The other candidate model, ARMA(2,1) with $\phi_1 = 0$, exhibits similar residual autocorrelation and partial autocorrelation patterns. The most parsimonious model for well 2128, therefore, is the AR(2) model

$$Y_t = 5.50 + 0.1776Y_{t-1} + 0.2344Y_{t-2} + 522.875X_t + 599.798X_{t-1} + 392.735X_{t-2} + a_t \quad [23]$$

where Y_t is the original time series, and X_t is the intervention time series. The complete model has an r^2 of 0.745, accounting for 74.5 percent of the total variance of the original time series with the model as estimated. The model is relatively simple to forecast since past values of a_t are not required to be determined, as in a moving average model.

4.2.3.2 Univariate Temporal Modeling of Well 2548

Monitor well 2548 is located approximately at coordinates (890,150) on Figure 4.3. The levels of arsenic concentration found at well 2548 are shown in Figure 4.15. The data points prior to time zero correspond to measurements made prior to the nominal pumping rate of 300 gallons per minute per pump. Again, these points are not consistent enough to perform temporal modeling. The intervention input transfer function weights are given in row 2 of Figure 4.5. The difference between two time periods corresponds to a week. Time period 1 is December 8, 1993. The last time period at $t = 41$ corresponds to September 13, 1994.

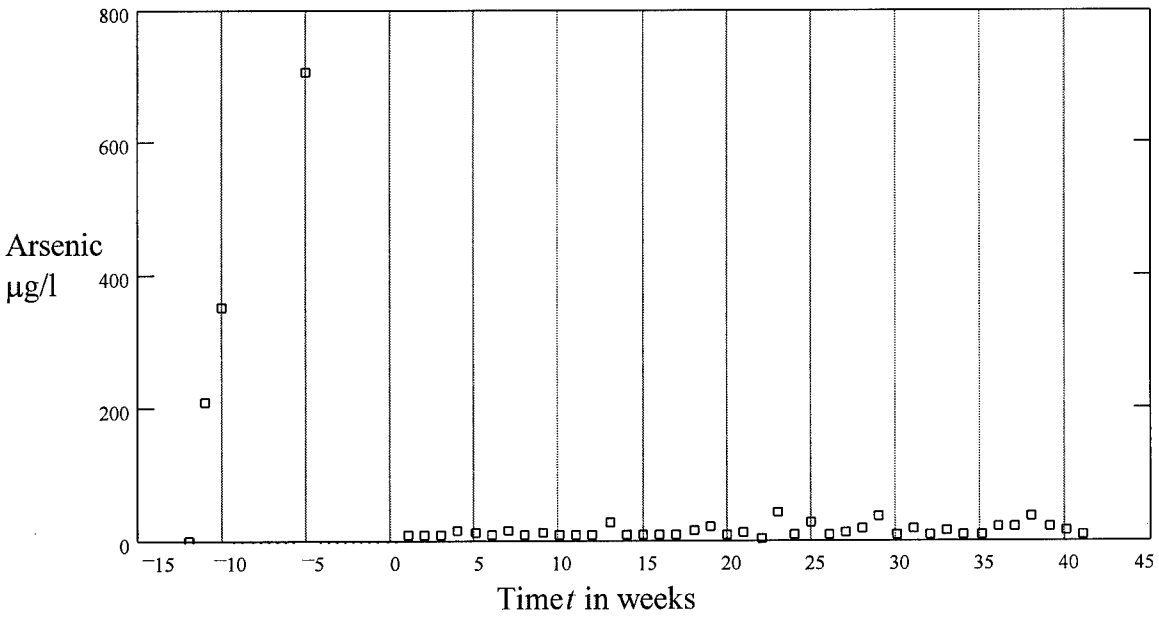


Figure 4.15 Arsenic Level Versus Time for Well 2548

The time series used in the temporal modeling of the arsenic is the same as shown above, except using only positive t values. This time series is shown in Figure 4.16.

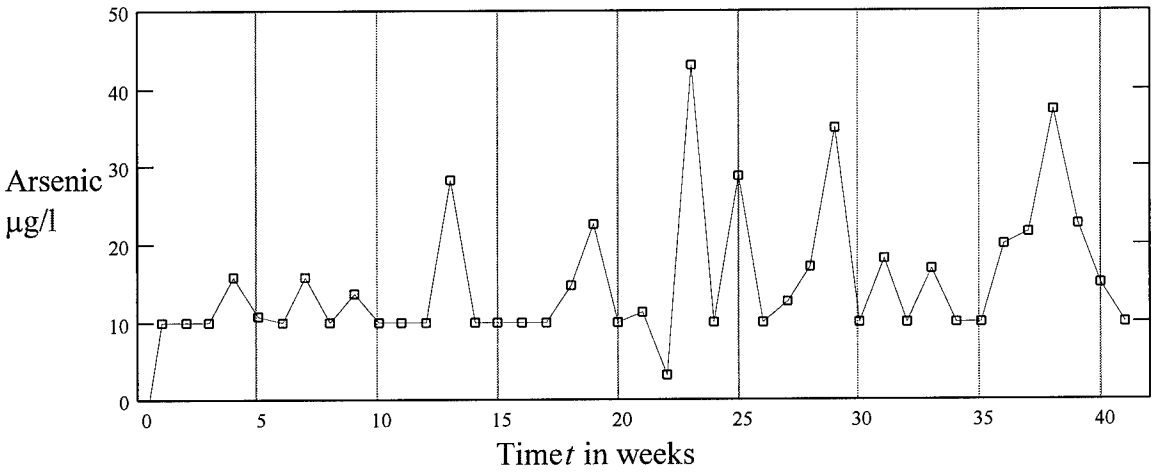


Figure 4.16 Time Series for Temporal Modeling of Well 2548

The algorithm yields the following autocorrelation and partial autocorrelation functions shown in Figures 4.17 and 4.18, respectively, for the pre-intervention time periods.

Inspection of this time series suggests that the series is approximately stationary in the mean prior to any intervention, and the autocorrelation function decaying quickly toward zero supports this conclusion, as well. Therefore, no differencing is warranted.

The first tentative model form is established next. Since the partial autocorrelation function has one significant spike while the autocorrelation function seems to behave as a sinusoid, an AR(1) model is selected.

The first iteration for well 2548 yields autoregressive coefficient $\phi_1 = -0.3215$ and impulse response weights $\nu_0 = 259.713$, $\nu_1 = 242.855$, $\nu_2 = 186.457$, and $\nu_3 = 86.204$. Again, since the prewhitening parameter coefficients were computed using only the non-

intervention time series, it is not expected that these first obtained impulse response weights will accurately reflect the final model. At this point, subsequent iterations are performed.

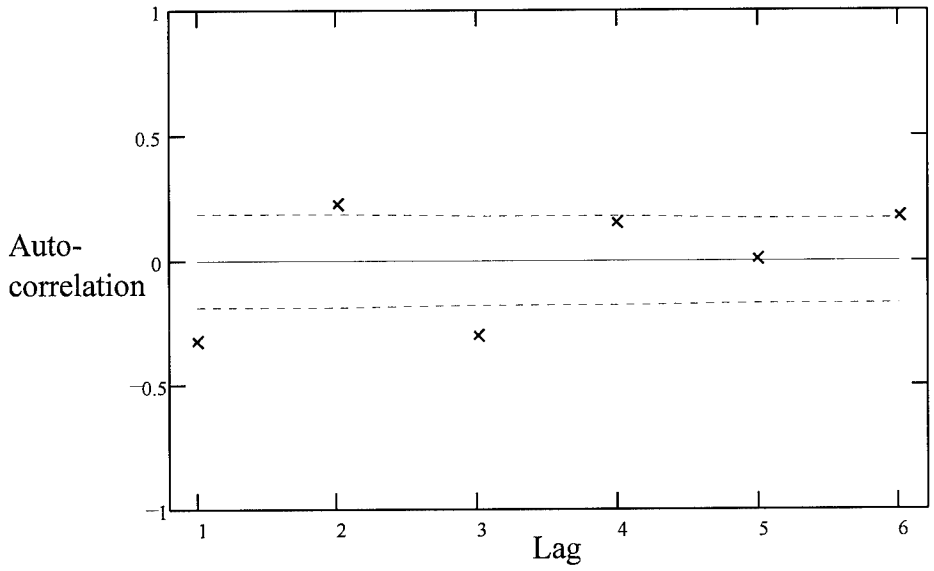


Figure 4.17 Autocorrelations for the Non-Intervention Period for Well 2548

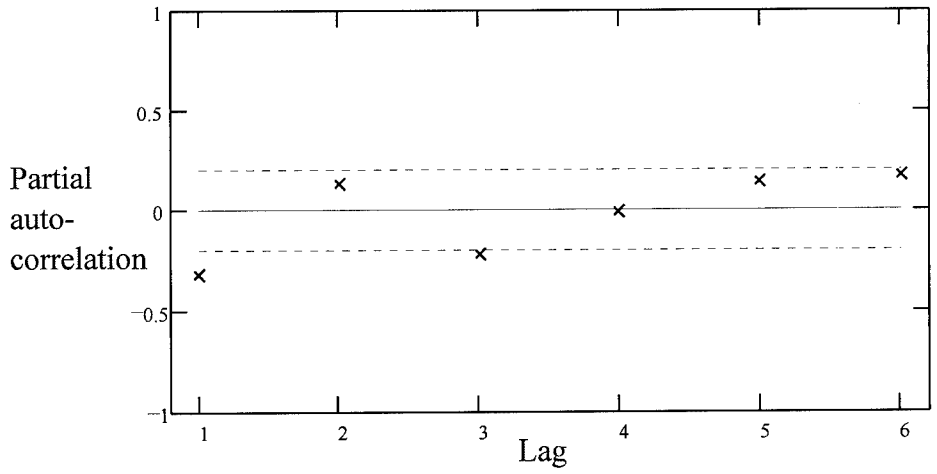


Figure 4.18 Partial Autocorrelations for the Non-Intervention Period for Well 2548

The final iteration of the temporal modeling for well 2548 displays autocorrelations and partial autocorrelations shown in Figures 4.19 and 4.20., respectively. Similar to modeling for well 2128, successive iterations for well 2548 suggested the inclusion of another autoregressive term.

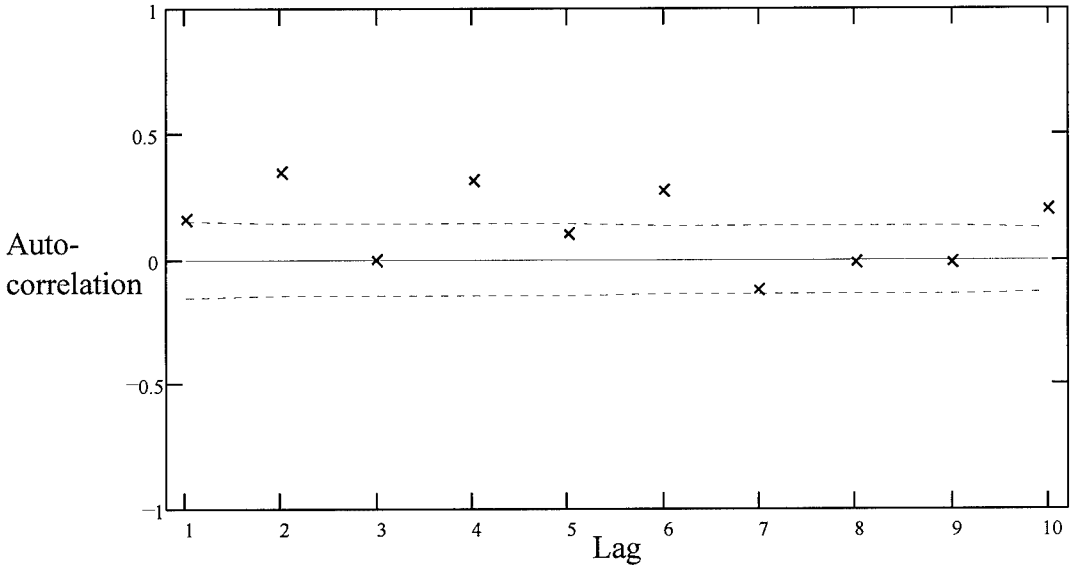


Figure 4.19 Autocorrelations for Well 2548

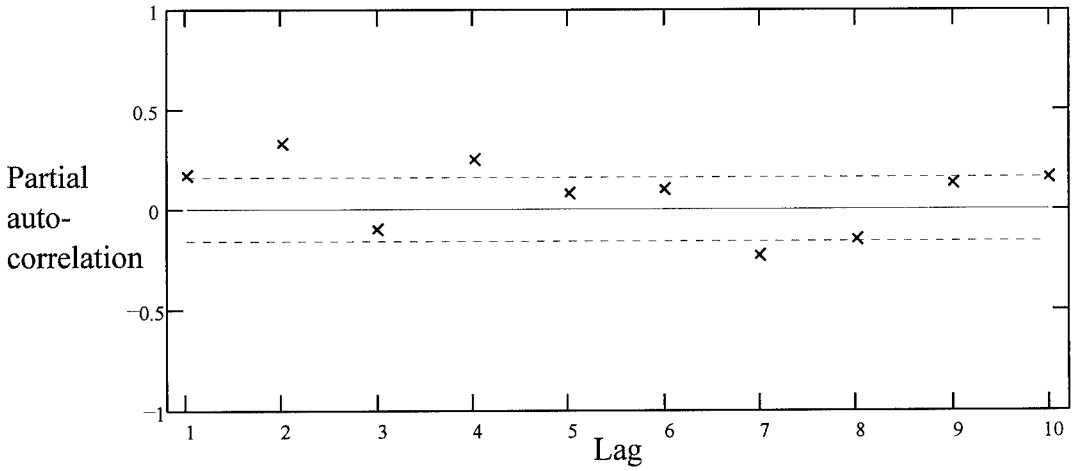


Figure 4.20 Partial Autocorrelations for Well 2548

The autocorrelation plot is interpreted as exponentially dampening. The partial autocorrelation plot is judged to have spikes through lag two. This combination of autocorrelation and partial autocorrelation plots suggests an AR(2) process and the final autoregressive coefficients estimated are $\phi_1 = 0.1104$ and $\phi_2 = 0.3313$. Again however, different interpretations of the autocorrelation and partial autocorrelation plots are possible. The first two autocorrelations could be considered spikes, thereby suggesting a moving average process, MA(2). The algorithm was run fitting this tentative model. The MA(2) model parameter coefficients are again approximately equal to the AR(2) coefficients, but opposite in sign. There was no problem with the fit for the estimation of the transfer function impulse response weights this time. The AR(2) weights behave as expected, again dying out by the fourth term: $\nu_0 = 180.974$, $\nu_1 = 283.278$, $\nu_2 = 214.292$, and $\nu_3 = -29.137$. The MA(2) weights this time, however, are nearly identical to the AR(2) weights. Higher order models do not produce better results in this case and are not considered. The AR(2) and MA(2) models being nearly equal, the AR(2) model is suggested as the most parsimonious both for its simplicity in forecasting and since it is possible the entire system follows a homogeneous modeling scheme.

After convergence in the program, the final residuals are plotted along with their autocorrelations and partial autocorrelations. Figure 4.21 shows the residual plot, while Figures 4.22 and 4.23 show residual autocorrelations and residual partial autocorrelations, respectively.

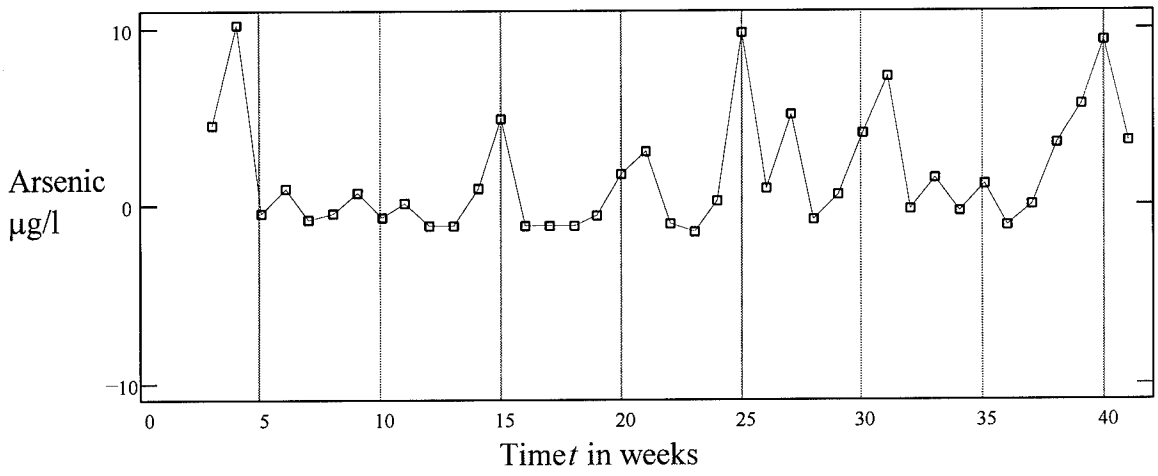


Figure 4.21 Residuals for Well 2548

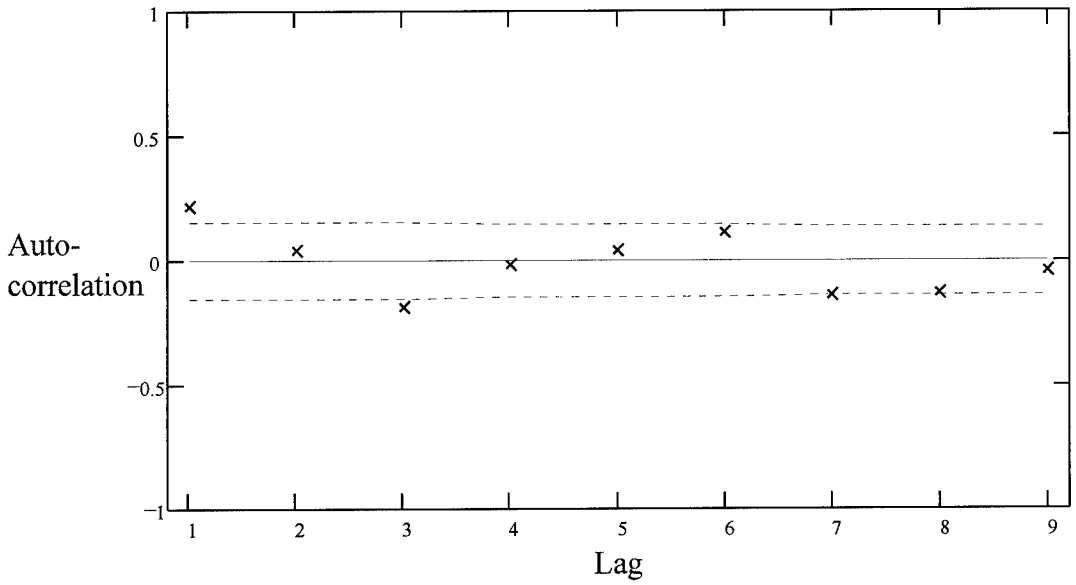


Figure 4.22 Residual Autocorrelations for Well 2548

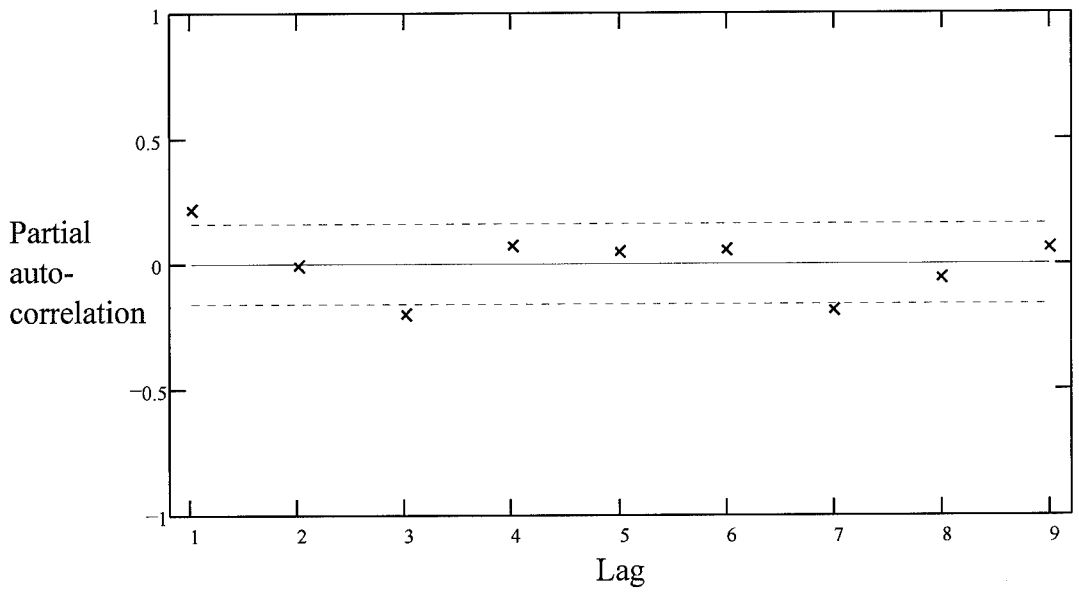


Figure 4.23 Residual Partial Autocorrelations for Well 2548

As with well 2128, Figure 4.21 suggests that variance is again greater during periods affected by the intervention (t greater than 26) than periods prior to the intervention. This is likely due to the same two reasons as with well 2128. The arsenic concentration threshold of 10.0 µg/l produced many concentration values of 10.0 µg/l, thereby reducing the variance for the pre-intervention series, and the data series reflecting a nominal level of arsenic concentration is likely to have a small variance compared to the same data with additional arsenic imposed upon it from some outside source. The autocorrelations and partial autocorrelations do not necessarily suggest that additional parameters are warranted in this case. The most parsimonious model for well 2548, then, is the AR(2) model

$$Y_t = 5.53 + 0.1104 Y_{t-1} + 0.3313 Y_{t-2} + 180.974 X_t + 283.278 X_{t-1} + 214.292 X_{t-2} + a_t \quad [24]$$

where Y_t is the original time series, and X_t is the intervention time series. The complete model has an r^2 of 0.917, accounting for 91.7 percent of the variance of the original time series with the model as estimated. The model is the same in form as that for well 2128, and close in constant term and autoregressive coefficients, as well. The impulse response weights are similar in shape, in that the greatest affect seems to be produced during the week following the intervention, with magnitude tailing off in the second week following the intervention and disappearing by the third week after the intervention.

4.2.3.3 Univariate Temporal Modeling of Well 2625

Monitor well 2625 is located approximately at coordinates (250,600) on Figure 4.3. The levels of arsenic concentration found at well 2625 are shown in Figure 4.24. The data points prior to time zero correspond to measurements made prior to the nominal pumping rate of 300 gallons per minute per pump. Again, these points are not used to perform temporal modeling. The intervention input transfer function weights are given in row 3 of Figure 4.5. The difference between two time periods corresponds to a week. Time period 1 is December 8, 1993. The last time period at t = 41 corresponds to September 13, 1994.

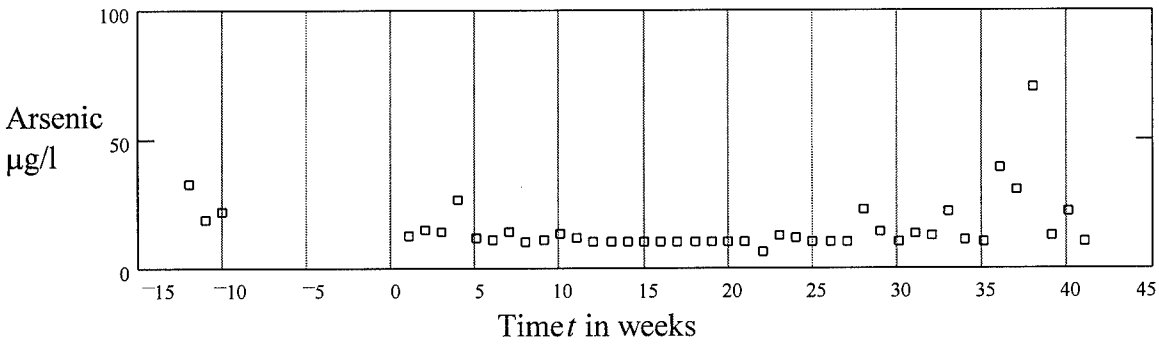


Figure 4.24 Arsenic Level Versus Time for Well 2625

The time series used in the temporal modeling of the arsenic is the same as shown above using only positive t values. This time series is shown in Figure 4.25.

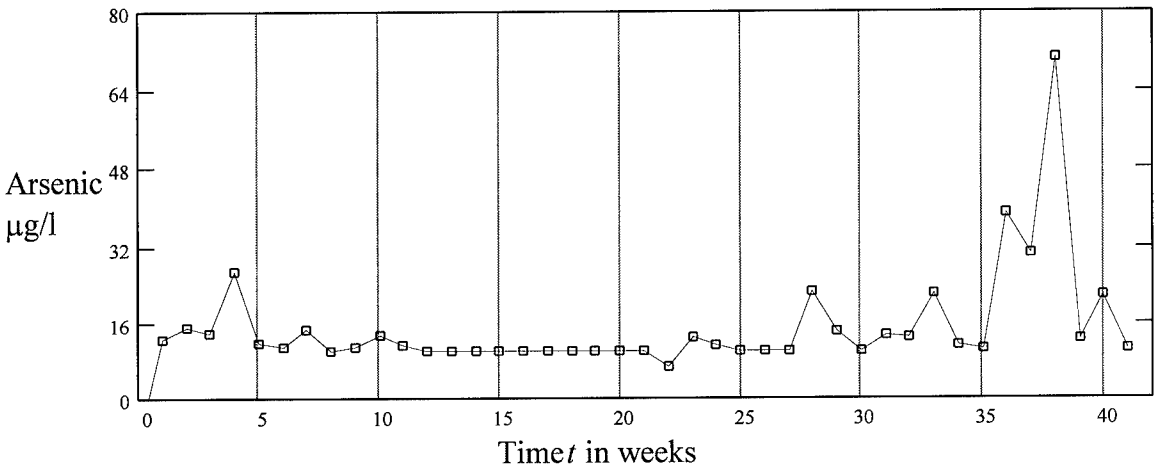


Figure 4.25 Time Series for Temporal Modeling of Well 2625

The algorithm yields the following autocorrelation and partial autocorrelation functions shown in Figures 4.26 and 4.27, respectively, for the pre-intervention time periods.

Inspection of this time series suggests that the series is approximately stationary in the mean prior to any intervention, and the autocorrelation function decaying quickly toward zero supports this conclusion, as well. Therefore, no differencing is warranted.

The first tentative model form is established next. Though the data do not clearly suggest a tentative model, homogeneity among the three modeled well sites seems promising, and a worthwhile goal. For this reason, an AR(1) model is selected, as in the initial iteration for the first two models.

The first iteration for well 2625 yields autoregressive coefficient $\phi_1 = 0.2042$ and impulse response weights $\nu_0 = 100.287$, $\nu_1 = 137.885$, $\nu_2 = 9.383$, and $\nu_3 = 88.335$. Again, since the prewhitening parameter coefficients were computed using only the non-intervention time series, it is not expected that these first obtained impulse response weights will accurately reflect the final model. At this point, subsequent iterations are performed.

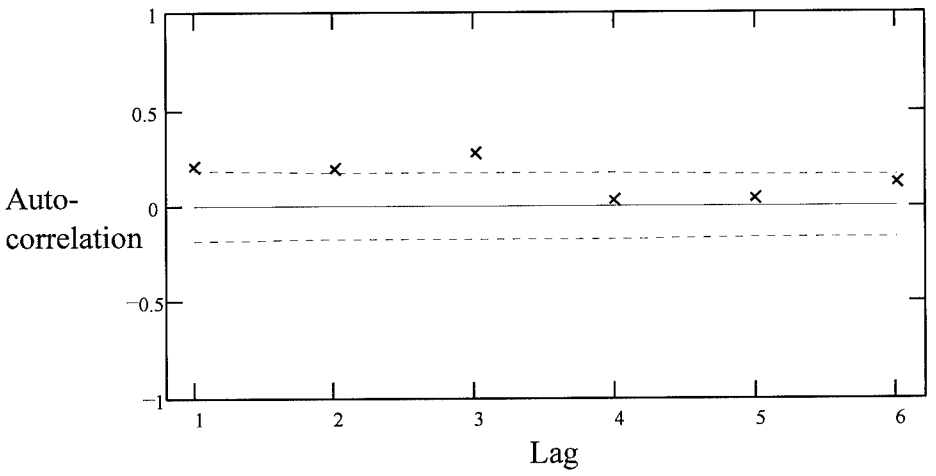


Figure 4.26 Autocorrelations for the Non-Intervention Period for Well 2625

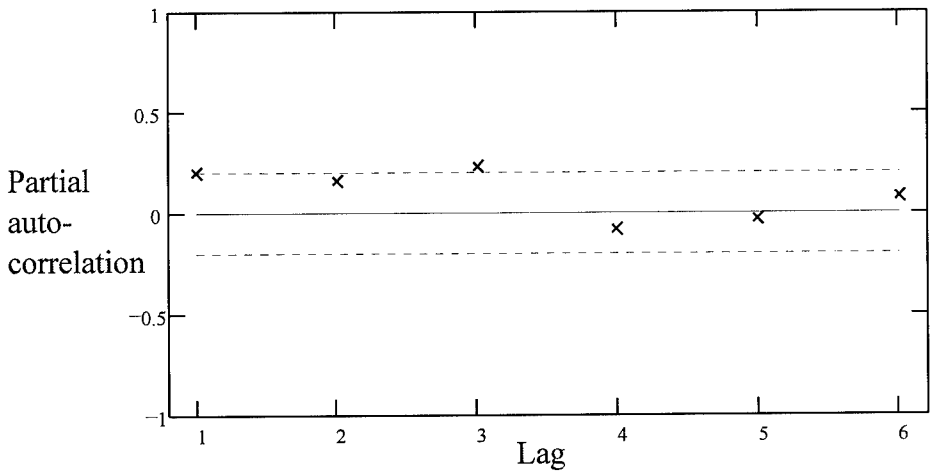


Figure 4.27 Partial Autocorrelations for the Non-Intervention Period for Well 2625

The final iteration of the temporal modeling for well 2625 displays autocorrelations and partial autocorrelations shown in Figures 4.28 and 4.29., respectively. Similar to modeling for wells 2128 and 2548, successive iterations for well 2625 suggested the inclusion of another autoregressive term.

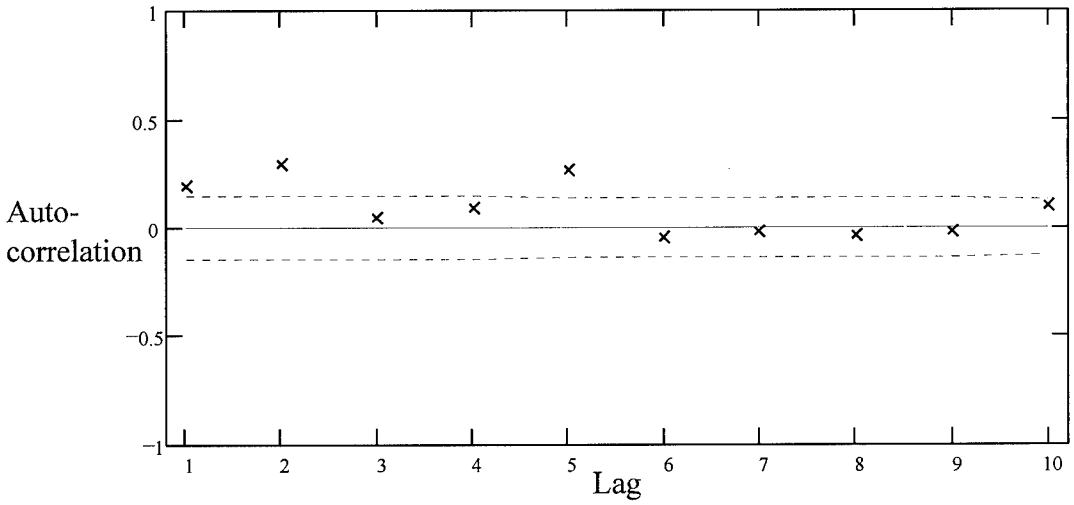


Figure 4.28 Autocorrelations for Well 2625

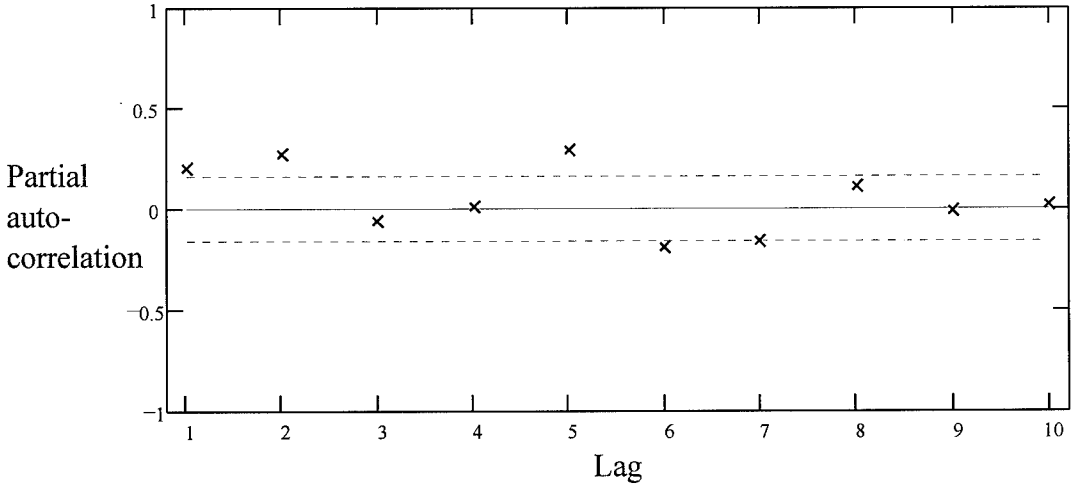


Figure 4.29 Partial Autocorrelations for Well 2625

The autocorrelation plot can again be interpreted as exponentially dampening. The partial autocorrelation plot has spikes through lag two. These interpretations suggest an AR(2) process and the final autoregressive coefficients estimated are $\phi_1 = 0.1444$ and $\phi_2 = 0.2703$. Different interpretations of the autocorrelation and partial autocorrelation plots are possible. The first two autocorrelations could be considered spikes instead of the first two partial autocorrelations, thereby suggesting a moving average process, MA(2). The algorithm was run fitting this tentative model. The MA(2) model parameter coefficients are again approximately equal to the AR(2) coefficients, but opposite in sign; the same result as the attempt with well 2128 and 2548. The levels of the transfer function impulse response weights this time were similar to those obtained with the AR(2) model. Higher order models do not produce better results in this case, and are not considered as candidate models. The AR(2) model is again suggested as the most parsimonious both for its simplicity in forecasting and since it now seems as though the entire system follows a homogeneous model form.

After convergence in the program, the final residuals are plotted along with their autocorrelations and partial autocorrelations. Figure 4.30 shows the residual plot, while Figures 4.31 and 4.32 show residual autocorrelations and residual partial autocorrelations, respectively.

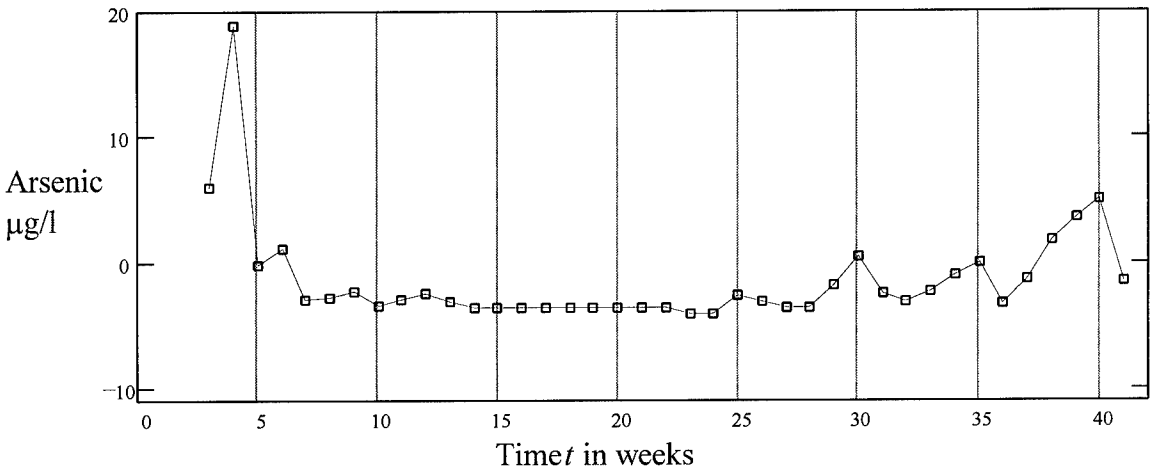


Figure 4.30 Residuals for Well 2625

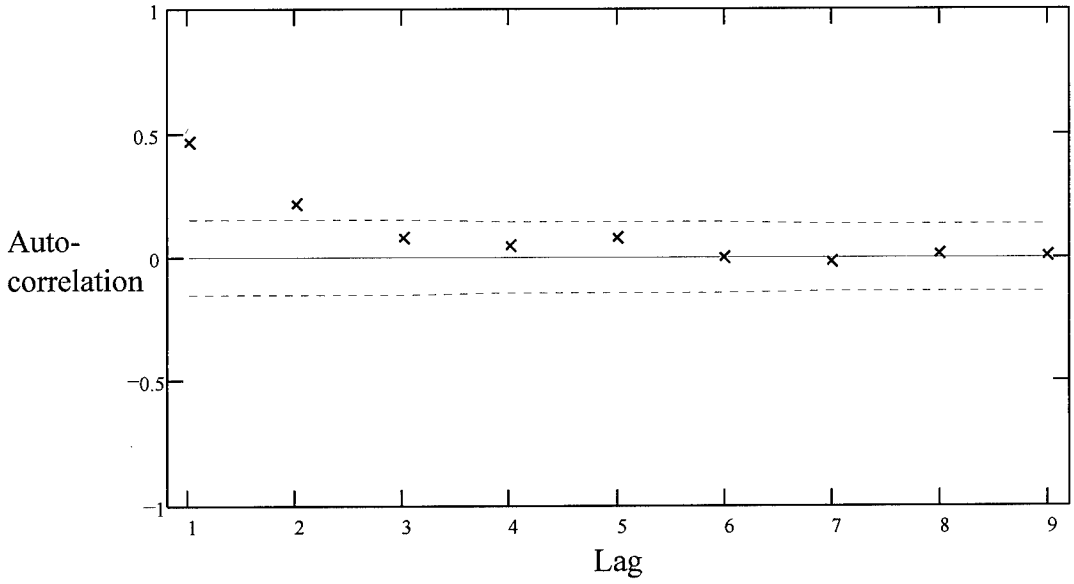


Figure 4.31 Residual Autocorrelations for Well 2625

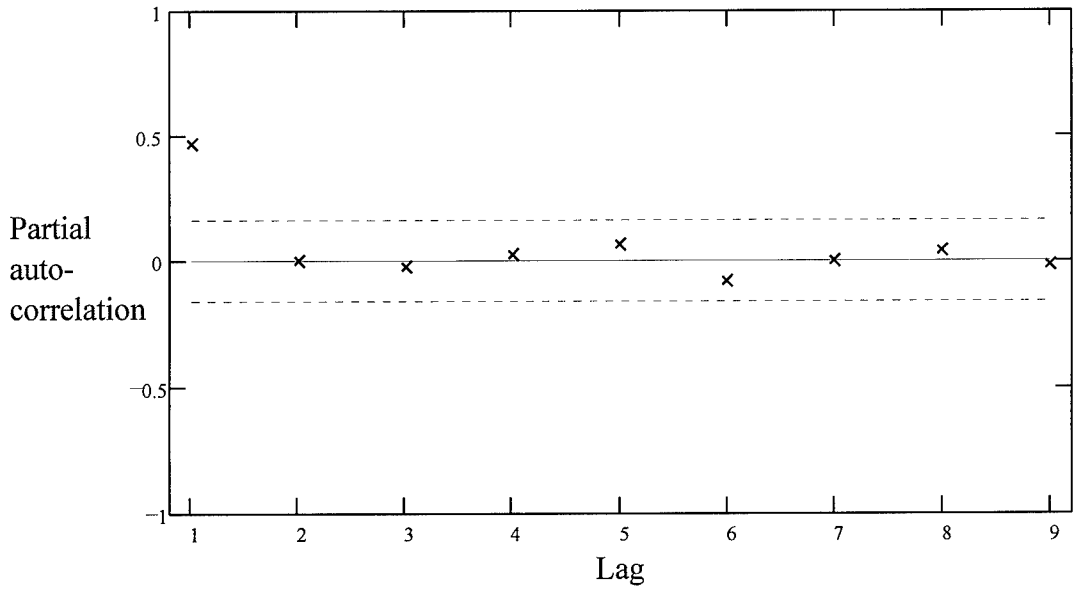


Figure 4.32 Residual Partial Autocorrelations for Well 2625

As with wells 2128 and 2548, Figure 4.30 suggests that variance is again greater during periods affected by the intervention (t greater than 26) than periods prior to the intervention, though this well also shows trouble at the beginning of the series The outliers present at the second residual is likely a left-over affect of the higher pumping rate used prior to switching to a nominal pumping level. The high values at the end of the residual series could be due in part to the same two reasons the high variances were seen during the intervention periods of wells 2128 and 2548. The arsenic concentration threshold of 10.0 µg/l produced many concentration values of 10.0 µg/l, thereby reducing the variance for the pre-intervention series, and the data series reflecting a nominal level of arsenic concentration is likely to have a small variance compared to the same data with additional arsenic imposed upon it from some outside source. Another possibility for the high values exist for well 2625, however.

The intervention input time series for this well has zero elements during the last seven periods due to its proximity to pumping well 3924 (reference Figure 4.3). Since

the number of impulse response weights is limited to four, the algorithm cannot make the intervention effect "stretch" far enough to encompass these high values, though an attempt is evidently made, as the impulse response weights do not behave as well as those for the other two wells modeled. The autocorrelations and partial autocorrelations do not suggest that additional parameters are warranted in this case, except for the first of each, but additional modeling terms do not solve the problem. The most parsimonious model for well 2625, then, is the AR(2) model

$$Y_t = 7.69 + 0.1444 Y_{t-1} + 0.2703 Y_{t-2} + 105.094 X_t + 87.414 X_{t-1} - 93.260 X_{t-2} + a_t \quad [25]$$

where Y_t is the original time series, and X_t is the intervention time series. The complete model has an r^2 of 0.626, accounting for 62.6 percent of the variance of the original time series with the model as estimated. The model is the same in form as that for wells 2128 and 2548, and close in constant term and autoregressive coefficients, as well. The impulse response weights are again similar in shape, however this time, the greatest effect seems to be produced during the week of the intervention, with magnitude tailing off in the week following the intervention and disappearing by the third week after the intervention. This slight difference could be attributable to the well's location as the western-most monitoring site, which the arsenic plume would pass first if being pulled into the groundwater flow.

4.3 Spatial-Temporal Modeling of Pollutants with Intervention Analysis

In this section, one well site of interest is modeled assuming interaction or useful correlation between sites. The algorithm used in the development of the spatial-temporal

model is presented in section 4.3.1. Data used in the application of this algorithm is shown in section 4.3.2. The actual modeling is presented in section 4.3.3.

4.3.1 Algorithm for the Spatial-Temporal Modeling of Pollutants with Intervention Analysis

This section describes the algorithm used to perform the spatial-temporal modeling of the arsenic contamination at a point of interest, to include the application of intervention analysis. The algorithm is based on the methodology discussed in sections 3.2.2 and 3.2.3. Again, the algorithm is iterative in order to develop the best possible STARMA and transfer function model.

The algorithm consists of seven distinct steps. 1) The spatially weighted time series data not including observations at periods affected by interventions is collected. 2) The autocorrelation function and the partial autocorrelation function are determined, and the time series and the autocorrelation and partial autocorrelation functions are plotted for analysis. The analysis determines if first order or spatial differencing of the data is necessary for obtaining stationarity prior to STARMA modeling. If differencing is required, the data are differenced and step 2 is repeated. If differencing is not required, the values of p, q, λ, and m to be used in modeling the stationary process are determined. 3) Autoregressive and moving average coefficients are calculated. 4) The autoregressive and moving average coefficients are assumed to represent the entire noise series (space time series minus effect of interventions) in order to calculate the impulse response weights of the interventions. 5) If this is the first time through, or if stopping criteria are not met, form a new time series by subtracting the effect of the calculated transfer function impulse response weights from the time series Return to step 2 with this new time series and repeat process until both the orders of p, q, λ, and m are the same as the previous iteration and the coefficients associated with the autoregressive parameters, the moving average parameters, and the impulse response weights are within some specified

tolerance of the previous iteration. 6) Translate the final impulse response weights to the transfer function parameters. If the model fit is adequate, the model of the form

$$Y_t = C + \sum_{i=1}^{M} \frac{\omega_i(B)B^{b_i}}{\delta_i(B)} X_{i,t} + N_t \qquad [26]$$

where N_t is some STARMA(p,q,λ,m) process is complete for the monitor well site of interest. 7) The spatial-temporal residual time series is analyzed for spatial outliers using the ICM contextual classification algorithm. If outliers are determined, further analysis is accomplished and the observations may be removed from or retained in the analysis. If outliers are removed, the entire process is repeated.

4.3.2 Data for Spatial-Temporal Modeling of Pollutants with Intervention Analysis

The raw data used in spatial-temporal modeling is the same as that used in the strict temporal modeling. Spatial-temporal modeling may assist in the search for an arsenic flow model, however, since something is known about how the groundwater flows through the aquifer. Even though the exact source of the arsenic is unknown, it is known that the use of the groundwater recovery system has introduced a west-to-east component to the prevailing north-to-south groundwater flow. Because of this, the source is generally considered to exist somewhere to the west of the monitoring wells which have seen an introduction of arsenic. To date, there has not been an increase in arsenic levels at the sites of the groundwater recovery wells.

Again, data for most of the time at six of the seven wells from which observations are being made are not exact at arsenic levels below a threshold of 10.0 µg/l, and these values again are used. Other data assumptions as described in section 4.2.2 still apply in the spatial-temporal analysis. Data from well 2900 and samples recorded from two of the

groundwater recovery system wells, wells 3924 and 3925 exhibit nearly no values above the threshold of 10 mg/l. While these wells were not modeled with the univariate ARMA algorithm, the data from them are used in the STARMA algorithm. Also, arsenic data at monitor well 2636 which used a higher threshold of 100.0 μg/l in approximately fifty percent of all observations are used in the spatial-temporal modeling by applying a non-zero spatial weight when a usable data value is available.

4.3.3 Spatial-Temporal Modeling of Pollutants South of the Groundwater Recovery System

Applying the FORTRAN code in Appendix C allows the user to proceed step-by-step through the process detailed in section 4.3.1. The process is used to model one site of interest, well 2548. Section 4.3.3.1 will detail the development of the spatial weights used in the spatial-temporal modeling. Analysis and results from the modeling of well 2548 are found in section 4.3.3.2. The analysis of residuals for delineating reasonable data from questionable data is found in section 4.3.3.3. This analysis could be used to smooth the data in order to define a cleaner data set which, in turn, could be used in determining a better STARMA model.

4.3.3.1 Spatial Weight Determination for Univariate Spatial-Temporal Modeling of Well 2548

As discussed in section 3.2.2, the spatial weights should, if possible, reflect some physical aspect of the system being modeled. Spatial weights are based on the lateral distance between the groundwater flow lines of the various monitoring wells. The natural flow at the area in question is known, and an estimation of the natural piezometric surface in feet above mean sea level is shown in Figure 4.33 [33:2-14]. The monitoring well sites are also displayed. The piezometric representation is assumed to be accurate, however, in reality, northern areas exhibit a smaller gradient than southern areas, and this is not

accounted for in the representation. Since most of the monitoring wells are located toward the south of the map, the changing slope of the piezometric surface is not expected to adversely affect the modeling. Flow lines are not shown, but flow occurs normal to the piezometric contours.

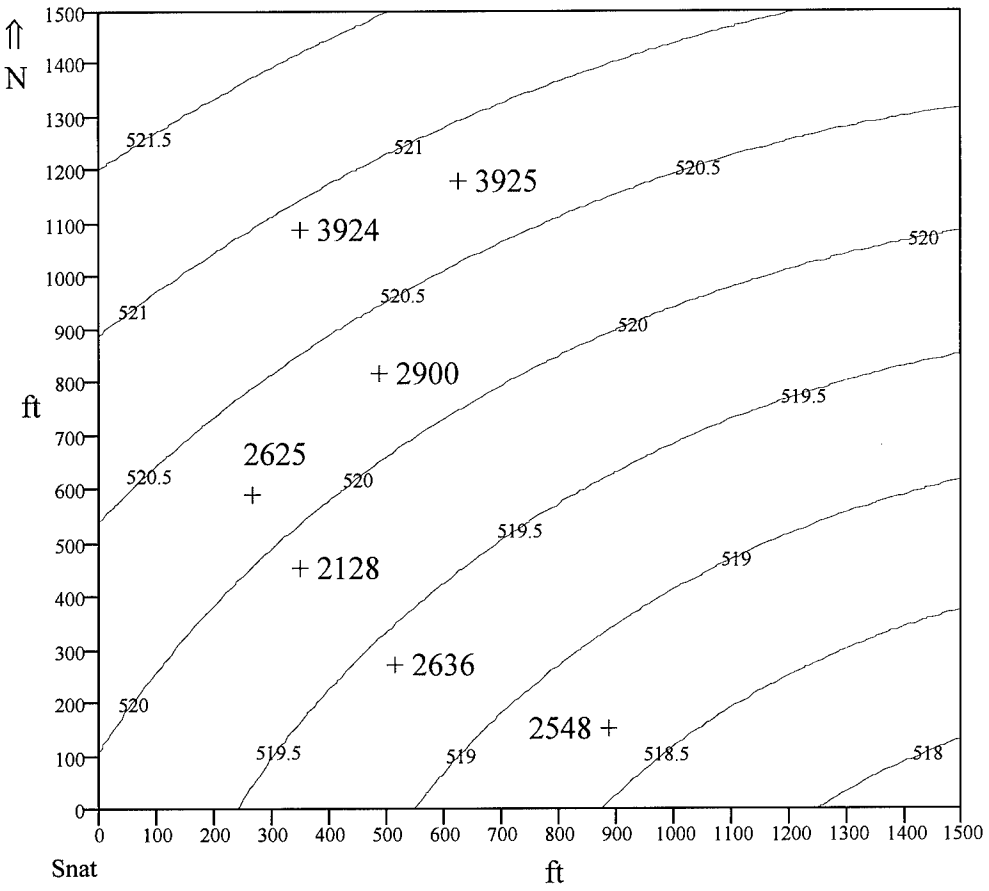


Figure 4.33 Natural Piezometric Surface at Area of Arsenic Introduction

However, this natural flow is skewed by the operation of the groundwater recovery system pumps. Since significant data are not available during periods of non-use of the groundwater recovery system, the nominal pumping scenario of 300 gallons per minute per day is used as the nominal case and is considered to have no intervention associated with it. Because of this, flow lines are constructed using information about the natural

flow and how the nominal use of the groundwater recovery system affects the natural flow. The piezometric surface including the effects of nominal groundwater recovery system usage is constructed by subtracting off the effect of the nominally pumping wells from the natural piezometric surface. Section 4.2.3 describes the process of determining the predicted drawdown and Figure 4.3 displays the predicted drawdown over the area in question. (The piezometric surface difference between contours in Figures 4.3 and 4.33 are not equal. More contours were used in Figure 4.3, so each contour represents a far less change in piezometric surface height than in Figure 4.33.)

Over-laying the effect of the nominal level of use of the groundwater recovery system on the natural piezometric surface results in Figure 4.34.

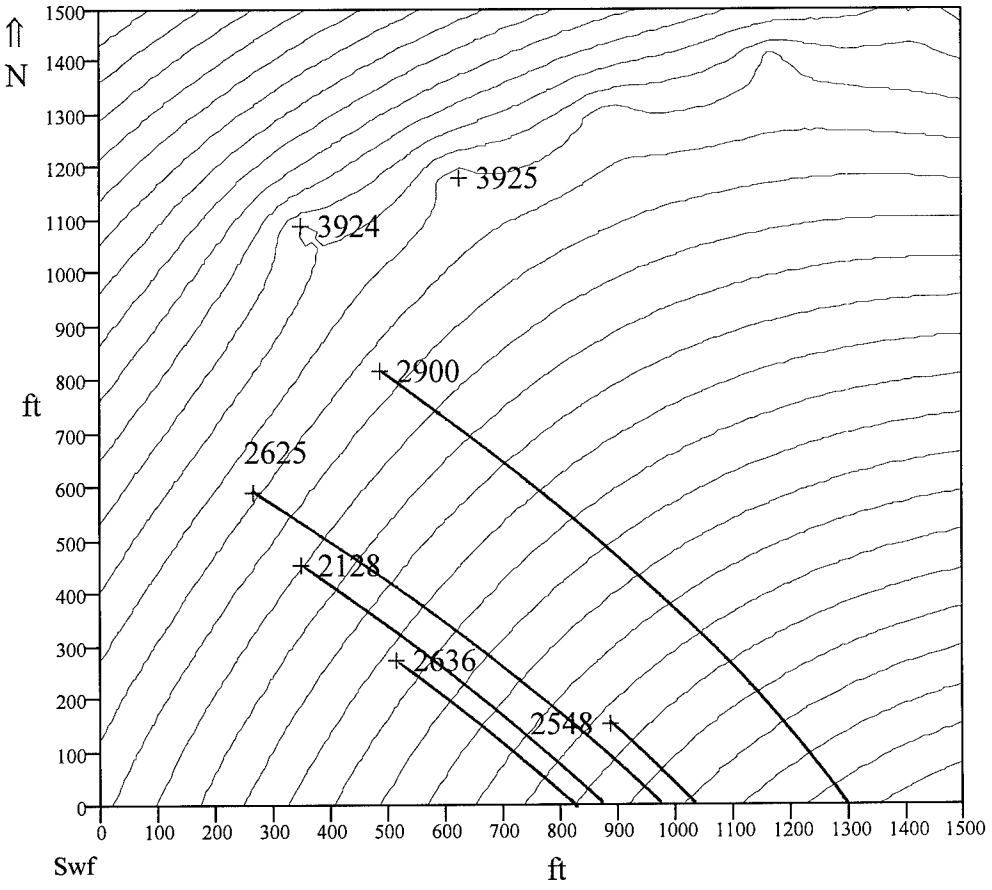


Figure 4.34 Predicted Piezometric Surface at Area of Arsenic Introduction

Though it is difficult to discern from Figures 4.33 and 4.34, flow lines normal to the lines of equal potential display larger west-to-east components, affecting the spatial weighting. The spatial weights, then, consist of the lateral differences between flow lines constructed through the various well sites of interest. Though these weights would change over time as the operation of the groundwater recovery pumps change, though this approach is not taken since the effect of the intervention is being modeled with the transfer function.

Determining which neighbors of the site of interest will be l^{th} order neighbors of the site interest is the next task, but first, the site of interest must be determined. An adjacency matrix is constructed between all points based on the methods described in section 3.2.2. The adjacency matrix is displayed in Figure 4.35.

	1	2	3	4	5	6	7
1	0	1	0	1	0	0	0
2	0	0	0	0	0	0	0
3	1	1	0	0	0	0	0
4	0	1	0	0	0	0	0
5	0	1	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

Well i (rows 1 through 7 associated numerically to well numbers) potentially exerts influence over well j (columns 1 though 7)

Figure 4.35 One Step Adjacency Matrix Based on Flow Line Proximity

This matrix describes the set of neighbors reachable in one step. The i and j wells relate to the real wells' numerical order. For instance, well 2128 is i or $j = 1$, while well 3925 is i or $j = 7$. Reachability from one site to another is determined by the smallest distance between the flow line of the first site and flow lines of downstream sites of interest.

Since groundwater is captured at the groundwater recovery system wells (wells 3924 and 3925), flow lines do not emanate from these sites, giving these two wells no neighbors (Figure 4.35, row and column 6 and 7).

The set of neighbors reachable in two steps is defined by squaring the adjacency matrix. The sum of the matrix and the squared matrix define the matrix of neighbors reachable in one or two steps. This one or two step adjacency matrix is shown in Figure 4.36. (Neighbors reachable in one or two steps would show a "2" in the matrix, but a "1" is displayed for clarity.)

	1	2	3	4	5	6	7
1	0	1	0	1	0	0	0
2	0	0	0	0	0	0	0
3	1	1	0	1	0	0	0
4	0	1	0	0	0	0	0
5	0	1	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

Well *i* (rows 1 through 7 associated numerically to well numbers) potentially exerts influence over well *j* (columns 1 though 7)

Figure 4.36 One and Two Step Adjacency Matrix Based on Flow Line Proximity

The column in Figure 4.36 with the most entries, column 2, represents the well with the greatest number of influential neighbors. This site, well 2548, is therefore selected as the site of interest.

Well 2548's first-order neighbors are selected as those which exhibit the smallest lateral distance from their flow line to well 2548. Figure 4.37 shows the lateral distance from the predicted flow line of each influential well to its one and two step neighbors. Zero entries imply that no relationship exists between the two sites.

	1	2	3	4	5	6	7
1	0	.38	0	.15	0	0	0
2	0	0	0	0	0	0	0
3	.22	.2	0	.35	0	0	0
4	0	.52	0	0	0	0	0
5	0	.67	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0

Lateral distance from flow line of well *i* to well *j*

Figure 4.37 Unscaled Spatial Weights between One and Two Step Neighbors

The site of interest is well 2548, or column two in Figure 4.37. The "closest" neighbors to this site are well 2128 (row 1) and well 2625 (row 3). These two neighbors are classified as well 2548's first order neighbors. All other wells will constitute second order neighbors.

The weights of the members of each first and second-order neighbors are created by inverting the lateral distances so heavier weights are assigned to "closer" neighbors. The weights assigned to the elements of a particular l^{th} order neighborhood are then scaled to sum to unity so the stream of data for a particular l^{th} order neighborhood becomes a weighted sum of its members and therefore remains indicative of the values of its members. The spatial weights assigned by doing this are found in Figure 4.38.

	0	1	2
1	0	.345	0
2	1	0	0
3	0	.655	0
4	0	0	.563
5	0	0	.437
6	0	0	0
7	0	0	0

*l*th order spatial weights of well *i*

Figure 4.38 Well 2548 l^{th} Order Neighbors and Spatial Weights

These spatial weights are then used in the spatial-temporal modeling of well 2548. The weighted sums of each of the zeroth, first and second order neighbors' time series are calculated, resulting in three new time series, representing each of the three neighborhoods. For instance, since the zeroth order neighborhood consists only of the well 2548 time series, the zeroth order time series also consists of this same time series. However, each of the first order time series elements is made of the weighted sum of the same element in well 2128 and well 2625's time series, in the proportion found in Figure 4.38. The second order time series is found the same way using data from the remaining wells. An exception to the way second order time series is constructed occurs when no data is available for a particular observation at well 2636, due to using a higher threshold of 100.0 μg/l in arsenic detection. The various order neighbors are then combined to form a time series three times as large as those used for temporal modeling. The first element is the first element of the zeroth order time series, the second element is the first element of the first order time series, the third element is the first element of the second order time series, the fourth element is the second element of the zeroth order time series, and so on.

4.3.3.2 Univariate Spatial-Temporal Modeling of Well 2548

The combined time series used in the spatial-temporal modeling of the arsenic is shown in Figure 4.39. For each week along the t axis, there are three values corresponding to the zeroth, first, and second order neighbors.

The high values at most of the third observations within each time period are attributable to well 2636 which, when measured with a less than 100.0 μg/l threshold, exhibited relatively high levels of arsenic both prior to and after the start of the interventions. However, most of the usable data points at well 2636 occur prior to the

intervention, making the second order neighborhood time series misleading. The second order neighborhood time series is shown in Figure 4.40.

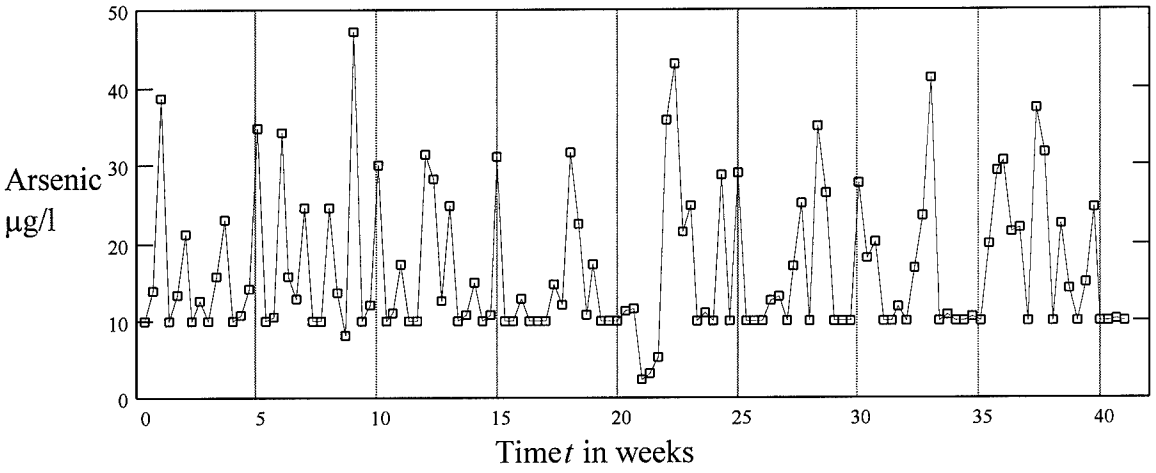


Figure 4.39 Time Series for Spatial-Temporal Modeling of Well 2548

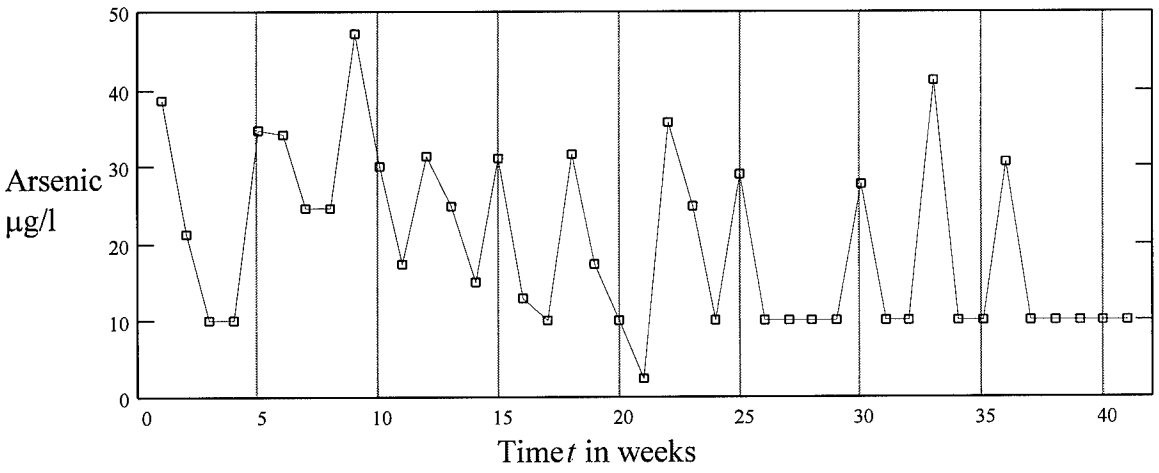


Figure 4.40 Second Order Neighborhood Time Series for Modeling Well 2548

All other time series with usable data exhibit low levels of arsenic until the intervention begins at time period 27. Such backward behavior in the second order neighbors is likely to be interpreted as negative correlations in the autocorrelation plot, thereby leading to the assumption of a false "causal" relationship, showing up as negative spatial autoregressive coefficients in the model, and leading away from determination of true causal relationships. Even if a transformation are accomplished to the well 2636 data to adjust the mean to a level similar to the rest of the wells, the lack of consistent observations during the interventions minimizes the data's significance. The autocorrelation plot for the entire time series is shown in Figure 4.41.

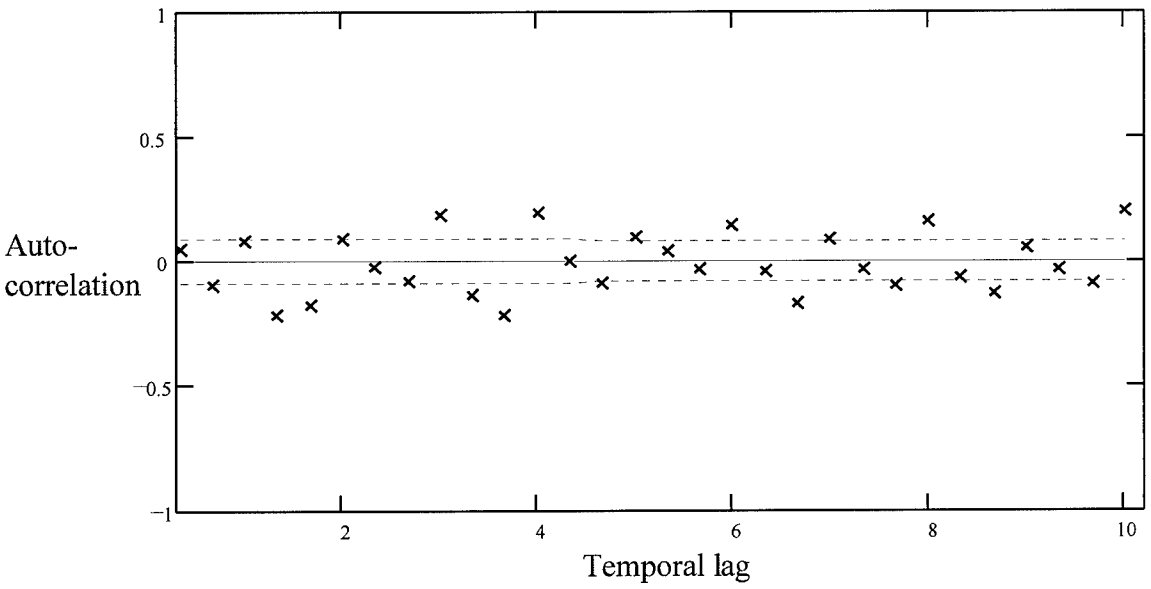


Figure 4.41 Initial Spatial-Temporal Autocorrelations for Modeling Well 2548

The first marker in the autocorrelation plot describes the first order spatial autocorrelation. The second marker and every third marker which follows (2, 5, 8, etc.) describe the second order autocorrelations at the various time lags shown. It is shown in

the figure that every second order neighbor lag is negative. Though in some instances this negative correlation may have some causal explanation, here the reason for the second order neighbor negative autocorrelations is known; it is a result of poor data. Trial runs with this time series also indicate the need for large negative second order neighbor autoregressive terms, as predicted.

Since the estimated model contradicts what is known of the system by modeling the false correlation, it is determined that data from well 2636 detracts from the modeling effort, and it is removed from the second order neighborhood. The remaining member of the second order neighborhood with a non-zero weight is well 2900. The data for this well is not dynamic, in that only three data values exist which are not at the 10.0 μg/l threshold. The conclusion here is that inclusion of this data will not improve the model since any dynamic aspect of the true data is lost due to the 10.0 μg/l detection threshold. Therefore, it is likewise determined that data from well 2900 with its low fidelity would detract from the analysis, and the data are removed as well.

Since there are no members of the second order time series left with non-zero weights, the second order time series has been effectively removed from the spatial-temporal analysis. This does not detract from a univariate spatial model, however. Removal of data known to be faulty will increase the model's validity. This is the case since in univariate spatial-temporal modeling, the string of neighboring values on which modeling is performed may produce false l^{th} order spatial correlations based on correlations which occur from the lower order neighbors to the high order neighbors. For instance, given zeroth, first and second order neighbors, correlation between the first order neighbor and the zeroth order neighbor will add to the second order spatial correlation, because the zeroth order time series at t is the second order spatial neighbor of the first order neighbor at t - 1. Likewise, correlation between the second order neighbor and the zeroth order neighbor will add to the first order spatial correlation,

because the zeroth order time series at t is a first order spatial neighbor of the second order series at t - 1. Therefore, keeping the number of candidate l^{th} order neighbors to a minimum is desirable in this univariate spatial-temporal modeling, in an attempt to reduce this misleading effect of the combined time series crossing from the last l^{th} order neighbor to the zeroth order neighbor. The analysis shown here, then, is accomplished on the combined time series consisting of two orders of neighbors, the zeroth and the first orders. On the other hand, if relatively little is known of the system being modeled, runs should be made with all information available, since such modeling can enlighten the modeler as to the nature of the system.

The interactive program for univariate STARIMA and causal transfer function model building follows the same flow as the program for univariate ARIMA and causal transfer model building, except here, the data are combined as required by the univariate STARIMA model building process. Also weighted and combined are the input transfer function time series. If a particular pair of first order neighbors which constitute the entire first order neighborhood are weighted 0.7 and 0.3, it is similarly expected the effect of the intervention on the entire first order neighborhood will consist of 70 percent of the first neighbor and 30 percent of the second.

Again in the program, the number of observations (starting at $t = 1$) in the time series occurring prior to any intervention is required input. The first estimate for autoregressive or moving average terms are based on this number of observations. Since the first 26 time periods occurred at the nominal pumping scenario of 300 gallons per minute per well, the combined pre-intervention time series will consist of 26 times the number of orders of neighbors being used in the modeling. However, if the effect of the interventions are not strong, indicating that the entire time series occurred prior to the intervention may reduce the number of iterations required to converge.

The sample autocorrelations and sample partial autocorrelations are then calculated. The time plot, and plots of the autocorrelations and partial autocorrelations are displayed. The autocorrelation plot shows the standard errors for the various lags for determining the significance of the autocorrelations. The partial autocorrelations are compared against their standard errors, as well. Actual plots from the program run are shown in Appendix D, but plots with greater fidelity are shown in Figures 4.42 through 4.44 for the combined time series, autocorrelations, and partial autocorrelations, respectively.

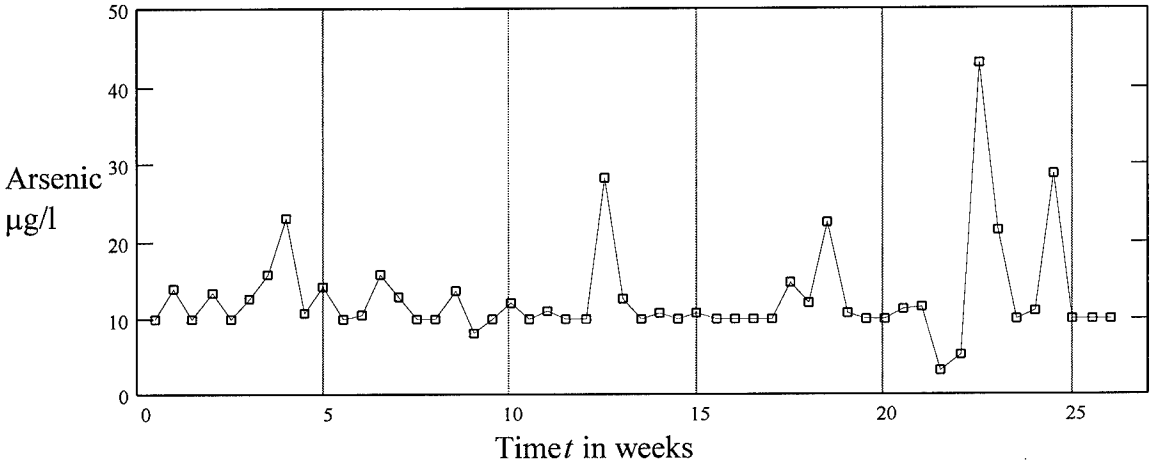


Figure 4.42 Combined Time Series for Non-Intervention Period for Spatial-Temporal Modeling of Well 2548

The user is again asked whether differencing should occur. If the user wishes to perform differencing, the differencing occurs on the time series as well as the intervention input time series. Inspection of this time series suggests that the series is approximately stationary in the mean, and the autocorrelation function decaying quickly toward zero supports this conclusion, as well. Again, no differencing is warranted.

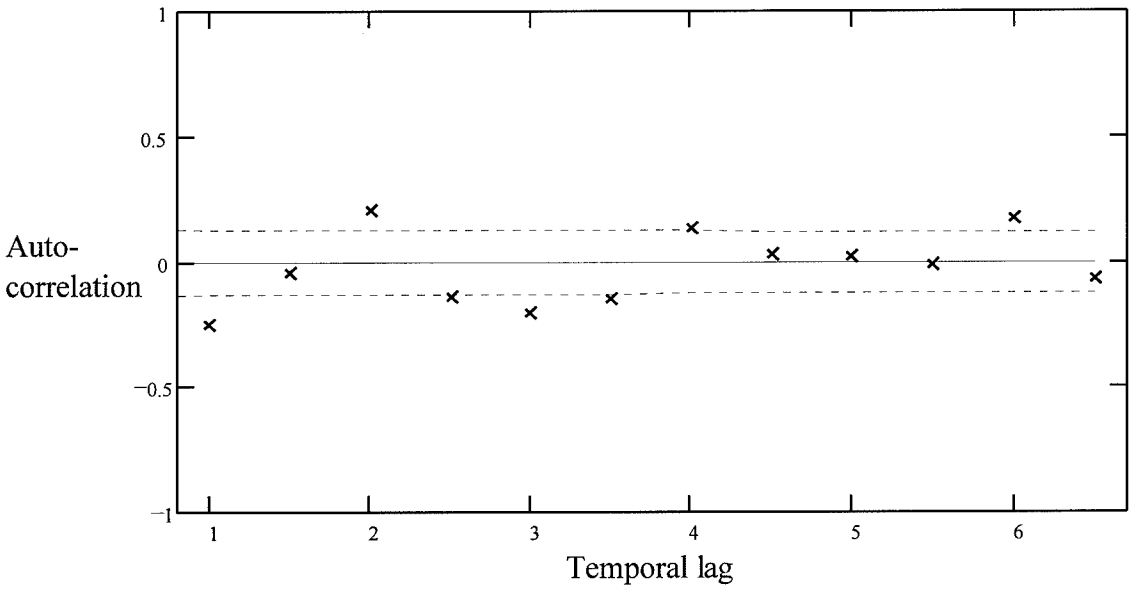


Figure 4.43 Autocorrelations for the Non-Intervention Period of Combined Time Series

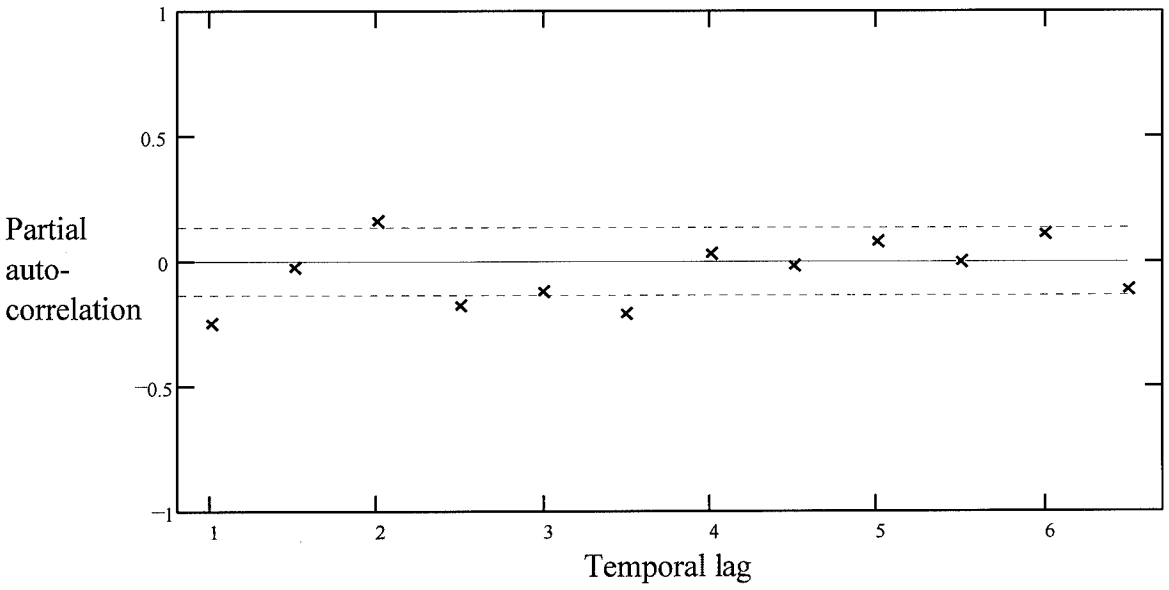


Figure 4.44 Autocorrelations for the Non-Intervention Period of Combined Time Series

Next, the first tentative model form is established. Since the estimate is not likely to reflect well the final selected model, this preliminary estimate should contain as few parameters as possible. Since both the autocorrelation and partial autocorrelation functions have one significant spike, an AR(1) or MA(1) model could be selected. The model in the spatial context would be a STARIMA$(0,0,0)^2(1,0,0)$ model form. The interpretation is one spatial autoregressive parameter and no others. The superscript two indicates that two orders of neighbors are being used in the modeling (the zeroth and the first order neighborhoods). Since neither the autocorrelation nor the partial autocorrelation functions show clear exponential decay toward zero, the choice between AR(1) and MA(1) is subjective. The AR(1) model is selected in this instance, both for parsimony reasons, as well as because the strict temporal modeling implied the probability of a homogeneous autoregressive underlying system.

Parameter coefficients for the selected model are then calculated. These estimates are then used as prewhitening parameters in the transfer function computation, which is accomplished on the entire combined time series, using the combined intervention time series as the input time series and the entire (82 observations) combined time series as the output time series.

The impulse response weights are calculated and displayed, along with the autoregressive and moving average parameter coefficients. The spatial-temporal model algorithm modeled eight impulse response weights for the combined series, which equates to the immediate response, and responses for one, two, and three lags for both zeroth and first order neighbors. Since the intervention input time series reflects the difference in drawdown at the neighbor of interest from the nominal pumping scheme to the various other pumping schemes measured in feet, assuming the drawdown calibrations are correct, the physical interpretation of the impulse response weights in the spatial-temporal model is the expected increase in the level of arsenic concentration

measured in μg/l due to a unit (one foot) increase in the drawdown at a neighborhood (made of the weighted drawdowns of the sites in the neighborhood) from the drawdown level at the nominal pumping scheme of 300 gallons per minute per pumping well. The first weight, then, suggests how much the intervention at the zeroth order neighbor immediately affects the site of interest. The second weight suggests how much the intervention at the first order neighborhood immediately affects the site of interest. The third weight suggests how the site of interest is affected by the intervention which occurred at the zeroth order neighbor one time period ago. The fourth weight suggests how the site of interest is affected by the intervention which occurred at the first order neighbor one time period ago.

As with the univariate ARMA program, the STARMA program asks whether the user would like to perform another iteration. Iterations in this research were again accomplished until prewhitening parameter coefficients converged to one within ten-thousandths on successive iterations. If the user specifies another iteration should occur, the predicted noise series which consists of the original time series minus the expected impact of the intervention based on the impulse response weight estimates is used as the time series from which the prewhitening estimates are based. The transfer function then uses the original time series with the newly estimated prewhitening parameter coefficients to determine the new impulse response weights. The process is repeated until the user is satisfied with the convergence of the parameter coefficient estimates. The number of iterations required for convergence varies with the number of parameters being estimated, but two parameters seem to converge in around nine iterations for the combined time series used in this research.

Even though autocorrelations appear relatively large at somewhat large lags, the STARIMA$(0,0,0)^2(1,0,0)$ model is fit until convergence, since ARMA models rarely are significant when many parameters are estimated. The $\phi_1 = 0.260$ implies correlation

between the zeroth and the first order neighborhood. In checking the correlations among the residuals, however, there seems to be another parameter which could be estimated. Figures 4.45 and 4.46 show the autocorrelation and partial autocorrelation plots for the combined residual series for the STARIMA$(0,0,0)^2(1,0,0)$ model.

The conclusion here is that another parameter would provide a better fit. The entire algorithm is repeated then, with the addition of another AR parameter. The model in the spatial context is a STARIMA$(1,0,0)^2(1,0,0)$ model. The interpretation is that one spatial autoregressive parameter and one autoregressive parameter are included.

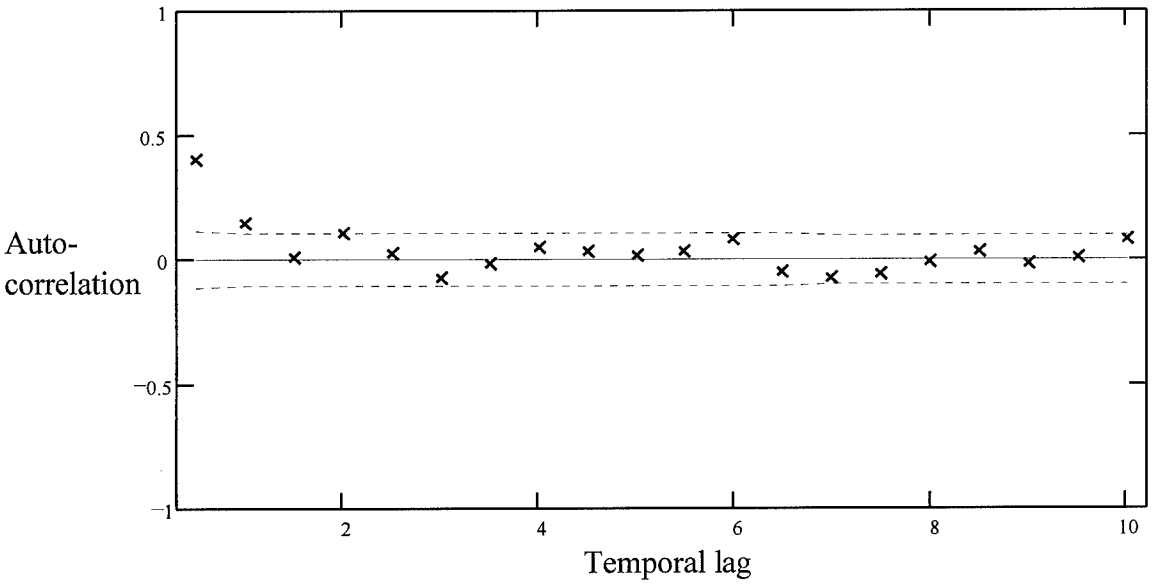


Figure 4.45 Residual Autocorrelations for STARIMA$(0,0,0)^2(1,0,0)$ Combined Time Series

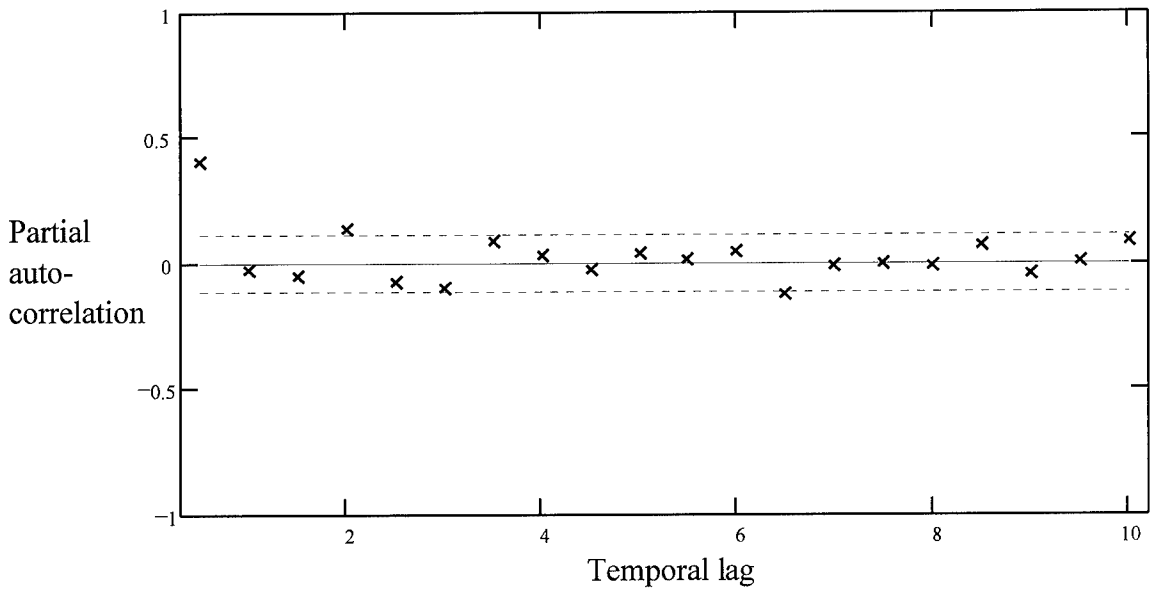


Figure 4.46 Residual Partial Autocorrelations for STARIMA(0,0,0)²(1,0,0) Combined Time Series

The final iteration of the spatial-temporal modeling for well 2548 using the STARIMA(1,0,0)²(1,0,0) model form displays autocorrelations and partial autocorrelations shown in Figures 4.47 and 4.48, respectively.

The autocorrelation plot shows a sinusoidal pattern. The partial autocorrelation plot is judged to have two spikes at lags one and two (both at temporal lag one, for the zeroth and first order neighbor). This combination of autocorrelation and partial autocorrelation plots validates the use of an STARIMA(1,0,0)²(1,0,0) process and the final autoregressive coefficients estimated are $\phi_1 = 0.3045$ (first order spatial term) and $\phi_2 = -0.1872$ (first order autoregressive term). The impulse response weights behave as expected, decreasing throughout the three weeks modeled: $\nu_0 = 140.151$, $\nu_1 = 183.441$, $\nu_2 = 128.358$, $\nu_3 = 166.648$, $\nu_4 = 107.147$, $\nu_5 = 101.213$, $\nu_6 = 65.569$, and $\nu_7 = 95.250$. Weights 0, 2, 4, and 6 represent the effect of the intervention time series of the zeroth

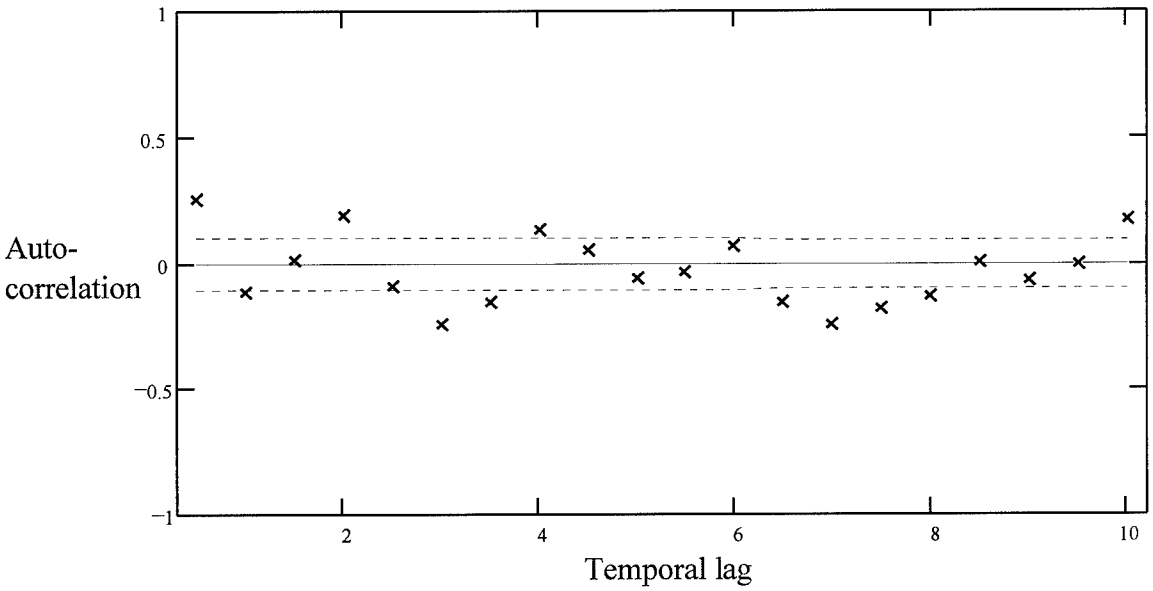


Figure 4.47 Autocorrelations for STARIMA(1,0,0)²(1,0,0) Combined Time Series

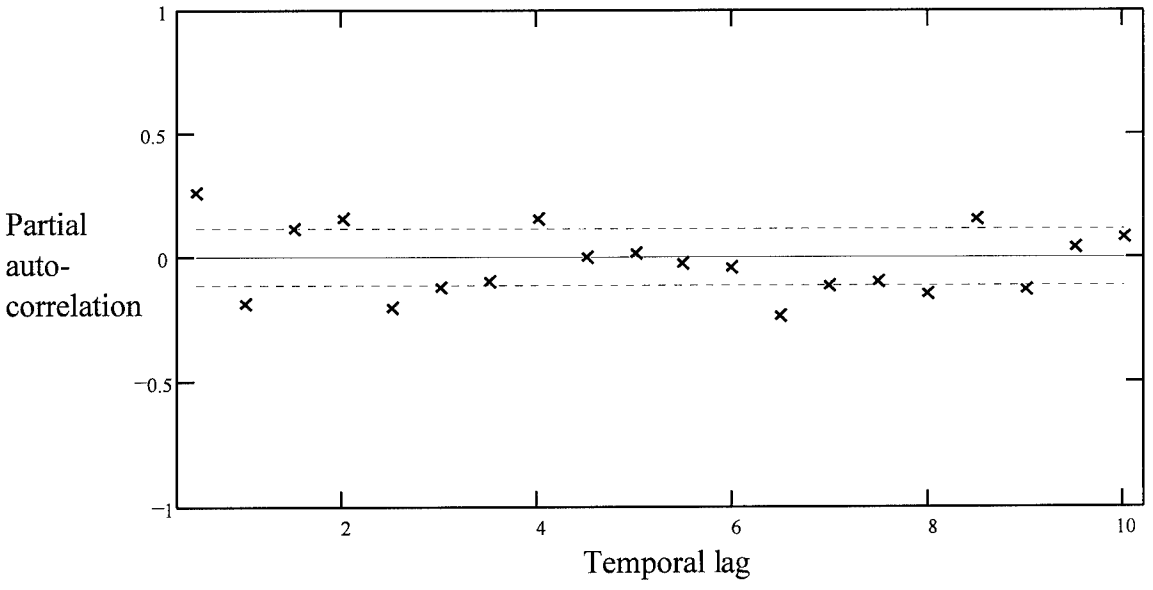


Figure 4.48 Partial Autocorrelations for STARIMA(1,0,0)²(1,0,0)

Combined Time Series

order neighbor on itself immediately and one to three time periods past (ν_2, ν_4, and ν_6, respectively), while weights 1, 3, 5, and 7 represent the effect of the intervention time series of the first order neighbor on the zeroth order site immediately and one to three time periods past (ν_3, ν_5, and ν_7, respectively). The weights obtained show a slightly stronger influence of the intervention for the first order neighbors on the site than of the intervention for the site itself, and both sets of weights decrease with time.

After convergence in the program, the final residuals are plotted along with their autocorrelations and partial autocorrelations. Figure 4.49 shows the residual plot, while Figures 4.50 and 4.51 show residual autocorrelations and residual partial autocorrelations, respectively.

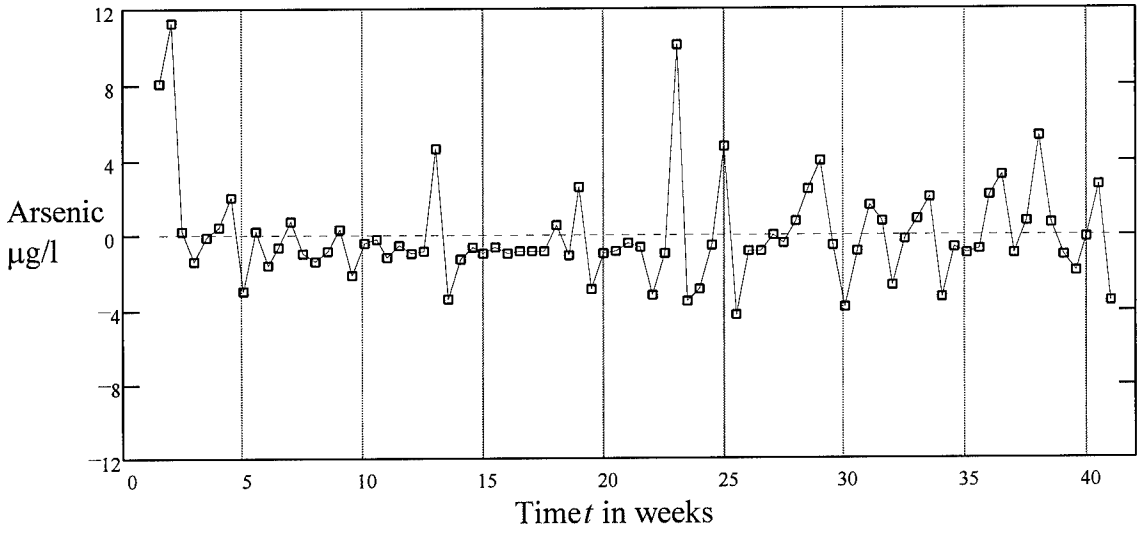


Figure 4.49 Residuals for Combined Time Series for STARIMA(1,0,0)^2(1,0,0) Model for Well 2548

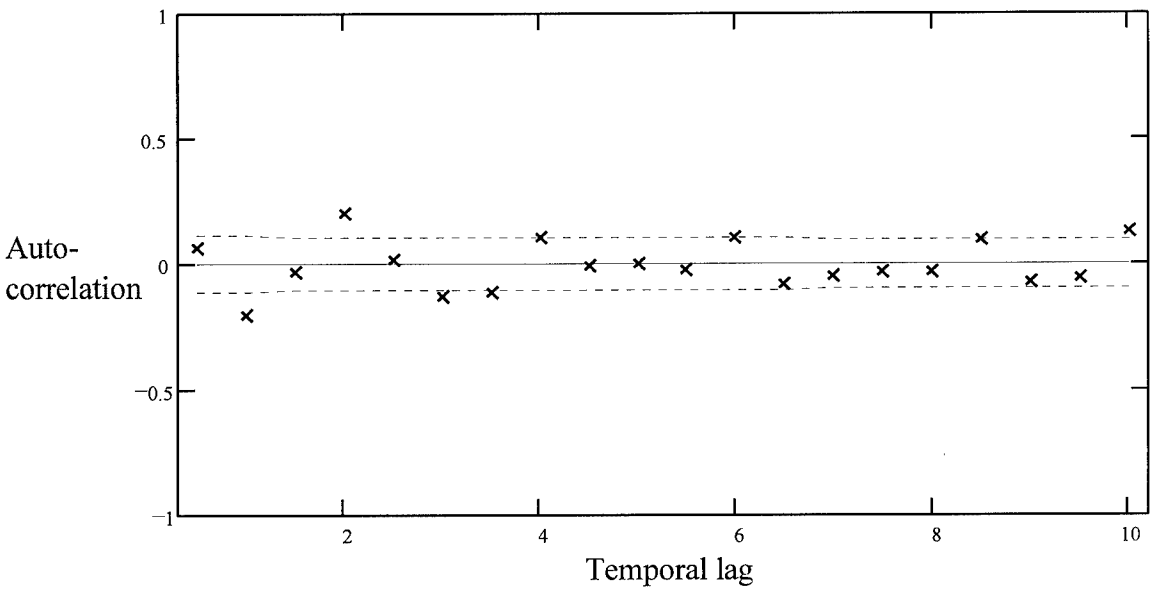


Figure 4.50 Residual Autocorrelations for STARIMA(1,0,0)²(1,0,0)

Combined Time Series

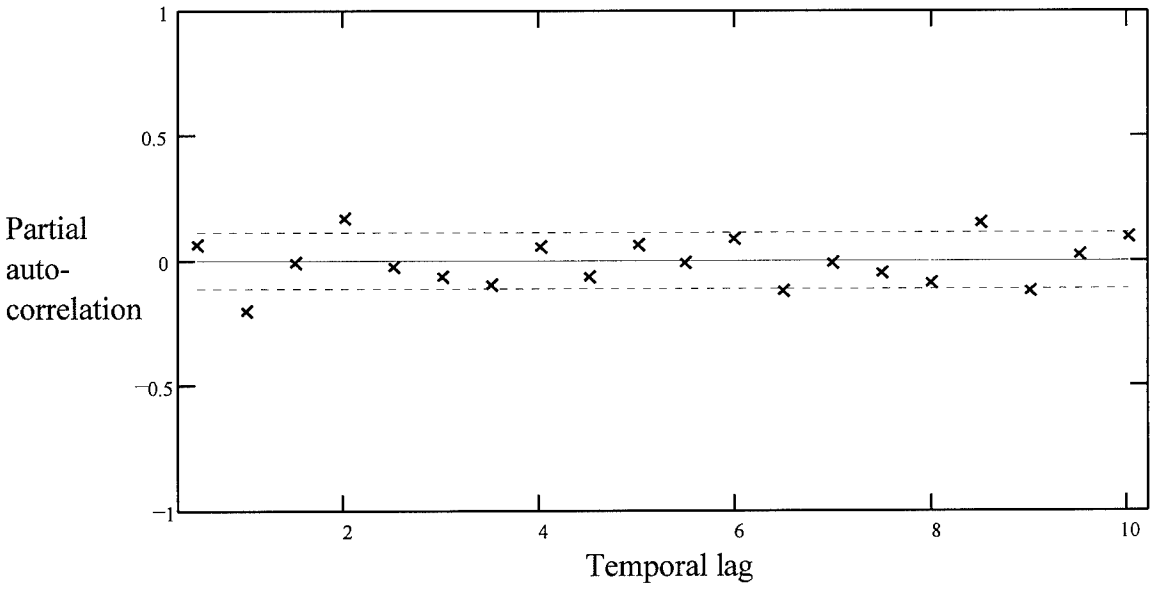


Figure 4.51 Residual Partial Autocorrelations for STARIMA(1,0,0)²(1,0,0)

Combined Time Series

Unlike the temporal modeling, Figure 4.49 suggests that variance is not much greater during periods of intervention (t greater than 26) than periods prior to the intervention. This difference is understandable since the spatial weighting of the neighbors has a smoothing effect on them. Also, the autocorrelations and partial autocorrelations for the residuals behave very well in the spatial-temporal case with no evident spikes or pattern. The most parsimonious spatial-temporal model for well 2548, therefore, is the STARIMA(1,0,0)²(1,0,0) model

$$\begin{aligned} Y_t^0 = {} & 10.52 + 0.3045 Y_t^1 - 0.1872 Y_{t-1}^0 + 140.151 X_t^0 + 183.441 X_t^1 + \\ & 128.358 X_{t-1}^0 + 166.648 X_{t-1}^1 + 107.147 X_{t-2}^0 + 101.213 X_{t-2}^1 + \\ & 65.569 X_{t-3}^0 + 95.250 X_{t-3}^1 + a_t \end{aligned} \qquad [27]$$

where Y_t^0 is the original time series of the zeroth order neighbor, Y_t^1 is the time series from the first order neighborhood, X_t^0 is the intervention time series for the zeroth order neighbor, and X_t^1 is the intervention time series for the zeroth order neighbor. The model is still relatively simple to forecast since past values of a_t are not required to be determined, as in moving average models. The r^2 for the STARIMA(1,0,0)²(1,0,0) model is 0.886, taking the combined time series into account. However, the model accounts for 95 percent of the variance of the zeroth order time series, which is the series for well 2548.

4.3.3.3 Residual Image Classification to Delineate Questionable Data for Enhancement

The objective of this image classification analysis is to enhance questionable data values so future iterations of the STARMA analysis may occur with as clean data as

possible. This could lead to the identification of a homogeneous STARMA model which would model the non-intervention level of arsenic at any of the well sites in the study equally as well. The number of usable "pixels" for use in an image classification procedure is limited in this research. Since there are only three monitoring wells which provide reasonable data in this study, it is determined that image classification, even using the ICM algorithm, would be largely subjective in nature due to the flexibility involved in deciding upon both the σ weight and β values for use in the implementation of the algorithm. However, the procedure for performing this analysis is outlined as a proposal to enhance the limited data available for this study or for use with other studies possessing more robust data sets. Figure 4.52 shows flow charts this relationship between the spatial-temporal modeling and the performance of image classification.

The data sets required for this analysis can be constructed in one of two ways. First, a univariate STARMA analysis may be accomplished for each monitor well in the area with reasonable data. The noise time series (the original series with the effect of the interventions removed) from each of these analyses constitute the data which will used in this image classification analysis. Alternatively, one univariate STARMA analysis may be accomplished on a point which is affected by many neighbors. The noise series from the zeroth order neighborhood makes up the series of data for the point of interest. Fitting the overall STARMA model will also provide estimates for the first order neighborhood, as well as for any other orders of neighborhood included in the analysis. From these estimates, noise estimates for each member of the first or subsequent neighborhood may be determined using the spatial weights (the opposite method of combining the individual time series into the neighborhoods). The noise series of each well, then constitutes the data set to be used in this image classification analysis. The latter method of obtaining the noise series data will not be as accurate as the former, since in the former, each site is modeled individually and does not rely as heavily on the

precise accuracy of the spatial weighting scheme used. However the latter method is clearly much less labor intensive.

The data set, then, includes the noise series from each site of interest. These noise time series represent the arsenic level of each well without intervention effects, or the natural pattern of arsenic contamination. Performance of the ICM algorithm as outlined in section 3.3 over the n noise series at a time t, which should be incremented from the first noise value to the last in successive iterations, should yield only one subregion, given that a homogeneous form of the STARMA model is appropriate for the region, with a mean μ and variance σ^2. In this study, μ should be somewhere between 0.0 and 10.0 μg/l, since in periods with no pumping, elevated arsenic values are not present. However, should a site be allocated to a second subregion, thereby indicating the elevated presence of arsenic, the site is assumed to possess a questionable data value. (Of course, the proper selection of β is important so as not to identify an inordinate number of questionable values.)

Occasional elevated values (above 10.0 μg/l) are expected due to random shock. However, if through successive applications of subregion allocations elevated values continue, the STARMA model used is not yet a homogeneous representation of the region's natural arsenic flow. The questionable values may then be analyzed further. Data which can be proven inaccurate may be removed from analysis. More likely, the data points will be accurate, or at least unverifiable as inaccurate, and these points may remain in the model. If this is the case, the STARMA models is refined (as well as simultaneously re-estimating the transfer function) to account for the wayward points.

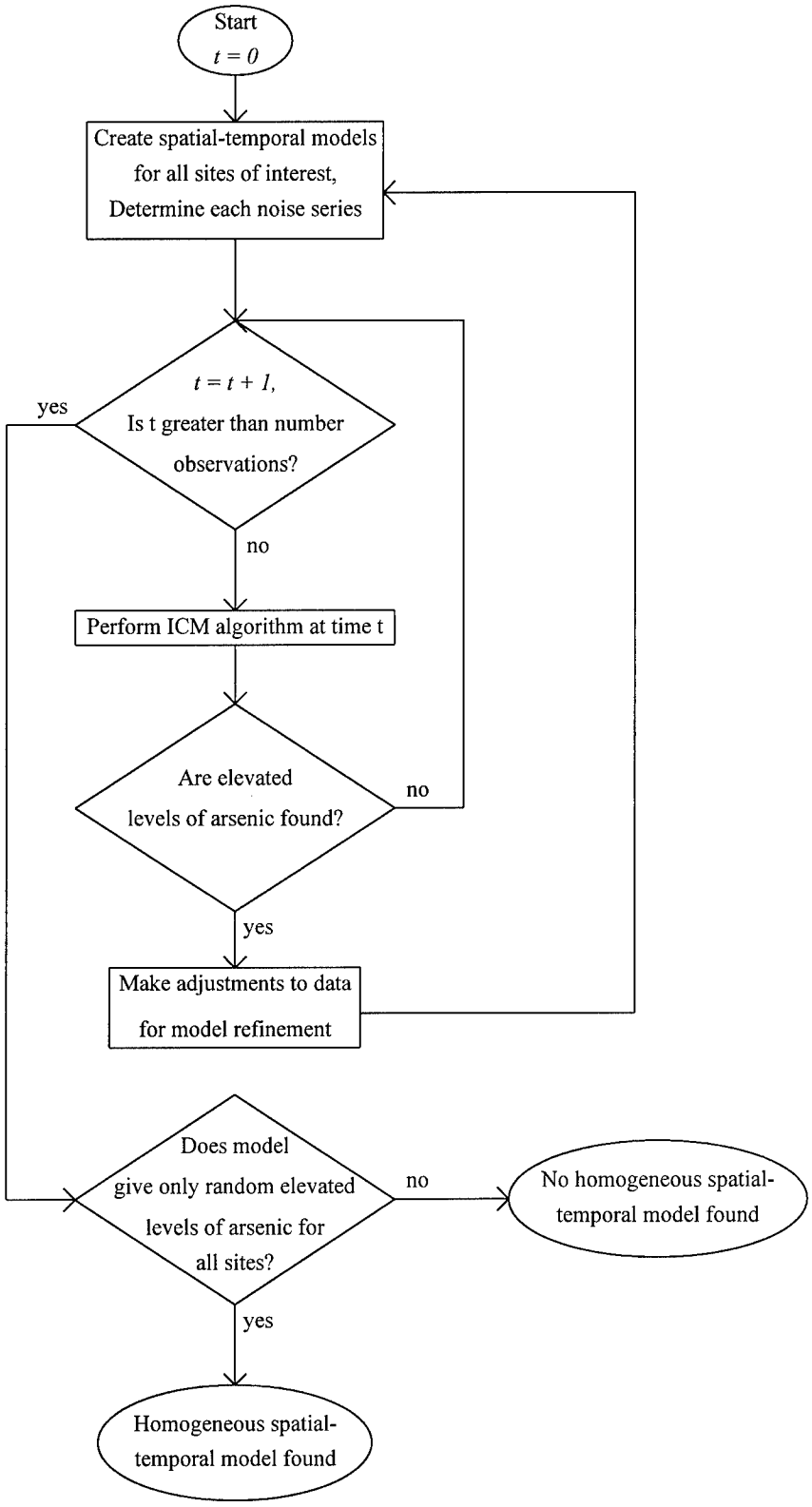


Figure 4.52 Flow Chart Integrating Spatial-Temporal Modeling and Image Classification

Starting from the time period most recently analyzed, the noise series from the refined STARMA model are placed through the image classification process. The entire image classification and STARMA/transfer function fitting procedure is applied iteratively on the noise series to further refine the STARMA model into a homogeneous model which represents the system of natural arsenic flow in the area of the groundwater recovery system. This procedure will determine, first, whether a homogeneous STARMA form can be used to represent the region. If a homogeneous form is available, the STARMA model so obtained would be considerably more validated than the univariate STARMA form resulting from this study.

V. Conclusions and Recommendations

5.1 Introduction

Methods have been presented for performing both temporal and spatial-temporal analysis of the introduction of arsenic into the groundwater at the area near the groundwater recovery system pumps. An analytic, causal method has been employed to create a spatial-temporal intervention transfer function input series to take away, as much as possible, the effect of the pumping wells on the system, leaving only a noise series representative of the steady-state level of arsenic under nominal pumping conditions. A comparison of the temporal and spatial temporal models for monitoring well 2548 is given in section 5.2. Insights to be gained from the results of this analysis are presented in section 5.3. Recommendations for model improvement are made in section 5.4.

5.2 Comparison of Temporal and Spatial-Temporal Modeling of Well 2548

The strict temporal modeling of well 2548 produced results which indicate the level of arsenic at the well is a function of some constant, the past two values of the level of arsenic at the well, and some random component. The model also takes into account predictions about the causal nature of the intervention effect.

The spatial-temporal model for well 2548 indicate that, when spatial effects are taken into account, the level of arsenic found at the well is a function of some constant, the level of arsenic present at well 2548's first order neighbors, and the past level of arsenic at well 2548, also a random component.

The general decrease in the effect of an intervention from the time of the intervention to some point three time periods later in each the temporal and the spatial-temporal models makes intuitive sense, but the spatial-temporal model's suggestion that the level of arsenic also depends on the predicted effect of the intervention at neighboring sites

seems equally as reasonable. The models' explanations of variance in the time series are comparable, though the spatial model accounts for slightly more of the variance for the site of interest than the temporal model. This is likely due to the estimation of more parameters representing the increase in arsenic at neighboring sites due to the intervention. Both the temporal and spatial-temporal models, however, provide unique insights which seem to validate the use of each of them.

5.3 Insights Gained from Temporal and Spatial-Temporal Modeling with Causal Intervention Analysis

Causally modeling the intervention of the operation of the groundwater recovery system at well sites of interest can provide some hypotheses as to the true nature of the system at the Fernald site. For instance, performing a non-causal intervention analysis using a spatial-temporal model would not likely yield good results since things such as level of intervention and the effect of the intervention on the flow lines around the sites of interest are not taken into account. Applying a causal analytic model to the intervention transfer function has allowed the forecasting of the level of arsenic due to a certain groundwater recovery system pumping scheme. For instance, if pumps near the monitoring wells sites are required to operate at a higher level, forecasting may be done to determine an expected level of arsenic due to the new level.

The inclusion of the analytic aspect of spatial weighting has also assisted in the development of a reasonable spatial relationship between sites. Had the natural flow of groundwater been used to create spatial weights, the flow lines of wells 2900 and 3924 would have passed closer to well 2548 than any other. However, the level of arsenic found at well 2548 has no correlation to the level of arsenic found at these wells. Including the effect of the nominal pumping scheme on the flow lines allows better spatial weighting, where wells 2128 and 2625 are viewed as the closest neighbors to

2548. This weighting is validated empirically, as well, since the patterns of arsenic introduction among these wells are fairly well correlated.

Obtaining a plausible spatial-temporal model with the causal intervention transfer function also suggests something about the location of the source of the arsenic. The skewing of the flow lines by the pumping of groundwater so the flow lines possess a greater eastern component combined with the construction of a model which assumes the magnitude of the eastern component is approximately correct leads to the conclusion the source of arsenic is approximately north-west of the sites along the skewed flow lines which pass through the sites having increased arsenic levels. The consistently increased level of arsenic at well 2636 during nominal pumping operations suggests the source lies upstream of its flow line during nominal operation of the groundwater recovery system, while the increased values at other nearby wells during increased operation of the pumps suggests that these flow lines are skewed more during times of increased pumping, thereby pulling the contaminated flow lines through the area of the other monitoring wells which have seen an increase in arsenic.

The values of the impulse response weights found in the temporal modeling can be used to indicate the distance of the arsenic source. The fact that the magnitude of the intervention transfer function remained consistent through each model allows implications to be made as to the actual affect of the intervention at a site. For instance, the site modeled closest to the recovery system pumps is well 2625, which showed the least affect in the increase of arsenic due to a unit increase in drawdown due to an increase in the level of pump operation. This may suggest the amount of drawdown is greater at this site than the others due to its proximity to the pumping wells, or that the source of arsenic is generally further south than this pump and it therefore requires an even greater increase in pumping to pull the flow lines with arsenic across the well. That

the values of the impulse response weights decrease from the very first value indicates the source of arsenic is closest to this well.

Impulse response weights for well 2128 show an increase from the first (immediate) weight to the second (one week after the intervention). This suggests the source of arsenic is further away from this well than from well 2625, since it seems to require more time to show its full effect. Given knowledge of the nature of dissipation of arsenic through groundwater, distance of the arsenic plume could be obtained. One possibility is that some concentration exists near the wells and that a higher concentration exists in groundwater beyond the wells. The model for well 2548 follows the same pattern as for well 2128, except here, the impulse response weight associated to two weeks following the intervention is greater than the weight for the immediate effect of the intervention. This implies that the source is even further away from well 2548 than from well 2128, which makes intuitive sense by looking at the well map and the superposition of flow lines on it (see Figure 4.34).

5.4 Recommendations for Model Improvement

The models presented in this research take an empirical view at temporal and spatial temporal modeling, using analytic filters to address the concern of interventions and spatial weighting. The quality of the analysis, therefore, relies on both the data and the analytic forms used. Improvements in the data used for this research would increase confidence in the obtained models. Should further analysis be desired, it is recommended that the concentration thresholds be reduced so each observation consists of an actual measured concentration. This seems particularly pertinent at well 2636 where increased levels of arsenic are consistently seen, even during nominal pumping periods. The high detection threshold of 100.0 μg/l used on samples from this well made the data gathered from it undesirable for use in this type of modeling.

With or without refinements in the data, it is suggested that an iterative STARMA and pattern recognition approach could be taken to develop a homogeneous (if one exists) spatial-temporal model for the region around the groundwater recovery system. Such a model could provide further insight as to the true source of the arsenic contamination, as well as increase confidence in forecasts octained with the model.

On the analytic side, improvements in the causal intervention transfer function input series could be made to even further reduce the effect of the intervention prior to the temporal or spatial-temporal modeling. The use of more accurate drawdown models in determining the magnitude of the transfer function input series is one possibility. Another is to use a combination spatial weight function which takes both the lateral distance between flow lines and the physical distance between sites into account. Also, the affect of increased groundwater flow during rain periods could be modeled into the transfer function input series, as well.

Appendix A. *FORTRAN Code for Univariate ARIMA and Causal Intervention Transfer Function Model Building*

This appendix contains the FORTRAN code for performing univariate ARMA and transfer function modeling.

```
************************************************************************
************************************************************************
* ARIMA modeling with intervention analysis
* by Samuel A. Wright
* March 1995
************************************************************************
************************************************************************
      integer i,inobsz,maxlag
      parameter(inobsz=41,maxlag=inobsz/4)
      integer obs(inobsz)
      real w2128(inobsz),w2548(inobsz),w2625(inobsz),w2636(inobsz),
     +     w2900(inobsz),inteff(inobsz),e2128(inobsz),
     +     e2548(inobsz),e2625(inobsz),e2636(inobsz),e2900(inobsz)
* diff declarations
      integer nobsz,ndiff,iper(1),iord(1),nlost
      real z(inobsz)
* acf declarations
      real zmean,acv(maxlag+1),ac(maxlag+1),seac(maxlag)
* pacf declarations
      real pac(maxlag),sepac
* sctp declarations
      character*20 xtitle,ytitle,title
      character*4 symbol
      integer icol(4)
      real range(4),zplot(inobsz,2),acplot(maxlag+1,4),pacplot(maxlag+1,
     +     4)
* nspe declarations
      integer npar,npma
      real const,par(4),pma(4),avar
* irnse declarations
      real x(inobsz),wtir(4),snoise(inobsz-3),xpw(inobsz-4),zpw(inobsz-4
     +     ),zw(inobsz)
* diff data
      data iper(1),iord(1) /1,0/
* sctp data
      data symbol /' *--'/
      data icol /1,2,2,2/
*
* get data
```

```
*
      ndiff=0
      print*,'Enter number of usable observations before interventions'
      read*,ninobsz
      nobsz=ninobsz
* determine well of interest
      print*,'Which well is being modeled?'
      print*,'Enter 1 for well 2128, 2 for well 2548, 3 for well 2625'
      read*,welloi
      open(20,file='welldata.dat',status='old')
      do 100 i=1,inobsz
        read(20,110) obs(i),w2128(i),w2548(i),w2625(i),x(i)
* zw gets time series with interventions
        if (welloi.eq.1) zw(i)=w2128(i)
        if (welloi.eq.2) zw(i)=w2548(i)
        if (welloi.eq.3) zw(i)=w2625(i)
100   continue
      close(20)
110   format(i4,4f8.3)
      do 105 i=1,nobsz
* z gets time series without interventions
        z(i)=zw(i)
105   continue
      open(23,file='effect.dat',status='old')
      do 107 i=1,inobsz
        read(23,111) e2128(i),e2548(i),e2625(i),e2636(i),e2900(i)
* x gets intervention time series
        if (welloi.eq.1) x(i)=e2128(i)
        if (welloi.eq.2) x(i)=e2548(i)
        if (welloi.eq.3) x(i)=e2625(i)
107   continue
      close(23)
111   format(5f10.6)
*
* difference data
*
* if differencing, difference noise series, time series with
* intervention, and intervention series x
*
115   if (ndiff.ne.0) then
        call diff(nobsz,z,ndiff,iper,iord,1,1,nlost,nobsz,z)
        call diff(inobsz,zw,ndiff,iper,iord,1,1,nlost,inobsz,zw)
        call diff(inobsz,x,ndiff,iper,iord,1,1,nlost,inobsz,x)
      endif
*
* get sample autocorrelations and sample partial autocorrelations
*
      mlag=nobsz/4
      call acf(nobsz,z,3,2,1,zmean,mlag,acv,ac,seac)
      call pacf(mlag,ac,pac)
      sepac=nobsz**(-.5)
* print partial autocorrelations
```

```
      print*,'partial autocorrelations: ',pac
*
* plot noise series, time series with interventions, and
* intervention series
*
* prepare plotting matrices for autocorrelations and partial
* autocorrelations
*
      do 117 i=1,nobsz
* prepare plotting matrix for noise series z
        zplot(i,1)=real(i)
        zplot(i,2)=z(i)
        print116,zplot(i,2)
117   continue
116   format(f8.3)
      do 120 i=1,mlag
        acplot(i,1)=real(i)
        acplot(i,2)=ac(i+1)
        acplot(i,3)=seac(i)
        acplot(i,4)=seac(i)*(-1.)
        pacplot(i,1)=real(i)
        pacplot(i,2)=pac(i)
        pacplot(i,3)=sepac
        pacplot(i,4)=sepac*(-1.)
        print118,acplot(i,2),acplot(i,3),acplot(i,4),
     +           pacplot(i,2),pacplot(i,3),pacplot(i,4)
120   continue
118   format(6f8.3)
      call page(-2,20)
* print time series, autocorrelations, and partial autocorrelations
      print*,'Enter when ready to view time series'
      read(*,'(a)') enter
130   title='TIME SERIES PLOT'
      xtitle='time period'
      ytitle='arsenic ug/l'
      range(1)=1
      range(2)=nobsz
      range(3)=0
      range(4)=80
      call sctp(nobsz,2,zplot,inobsz,icol,range,symbol,xtitle,ytitle,
     +          title)
      print*,'Enter when ready to view autocorrelations'
      read(*,'(a)') enter
      title='AUTOCORRELATION PLOT'
      xtitle='lag'
      ytitle='a/c'
      range(1)=1
      range(2)=mlag
      range(3)=-1
      range(4)=1
      call sctp(mlag,4,acplot,maxlag+1,icol,range,symbol,xtitle,ytitle,
     +          title)
```

```
      print*,'Enter when ready to view partial autocorrelations'
      read(*,'(a)') enter
      title='PARTIAL A/C PLOT'
      xtitle='lag'
      ytitle='partial a/c'
      call sctp(mlag,4,pacplot,maxlag+1,icol,range,symbol,xtitle,ytitle,
     +          title)
      print*,'Enter R to repeat plotting or Enter to go on'
      read(*,'(a)') enter
* query to repeat plotting or go on
      if ((enter.eq.'R').or.(enter.eq.'r')) goto 130
* query to difference time series if its mean appears non-stationary
      print*,'Enter D to difference the time series or Enter to go on'
      read(*,'(a)') enter
      if ((enter.eq.'D').or.(enter.eq.'d')) then
        ndiff=1
        iord(1)=iord(1)+1
* go back to difference all relevant time series and repeat a/c,
* partial a/c and plotting
        goto 115
      endif
*
* ARMA cofficient estimation
*
* first, query for values of p and q
*
      print*,'Enter value of p'
      read(*,'(i1)') npar
      print*,'Enter value of q'
      read(*,'(i1)') npma
      call nspe(nobsz,z,1,1,zmean,npar,npma,0.0,0,const,par,pma,avar)
*
* estimate transfer function impulse response weights
*
      print*,'Enter to estimate transfer function'
      read(*,'(a)') enter
      call irnse(inobsz,x,zw,1,npar,par,npma,pma,3,3,wtir,snoise,xpw,zpw
     +           )
      print*,'Autoregressive coefficients: ',par
      print*,'Moving Average coefficients: ',pma
      print*,'Impulse response weights:    ',wtir
      print*,'Enter Y to perform another iteration'
      read(*,'(a)') enter
      if ((enter.eq.'Y').or.(enter.eq.'y')) then
*
* repeat entire process using time series minus estimated intervention
* effect as noise series
*
* re-initialize effect of intervention
*
        do 155 i=1,inobsz
          inteff(i)=0.0
```

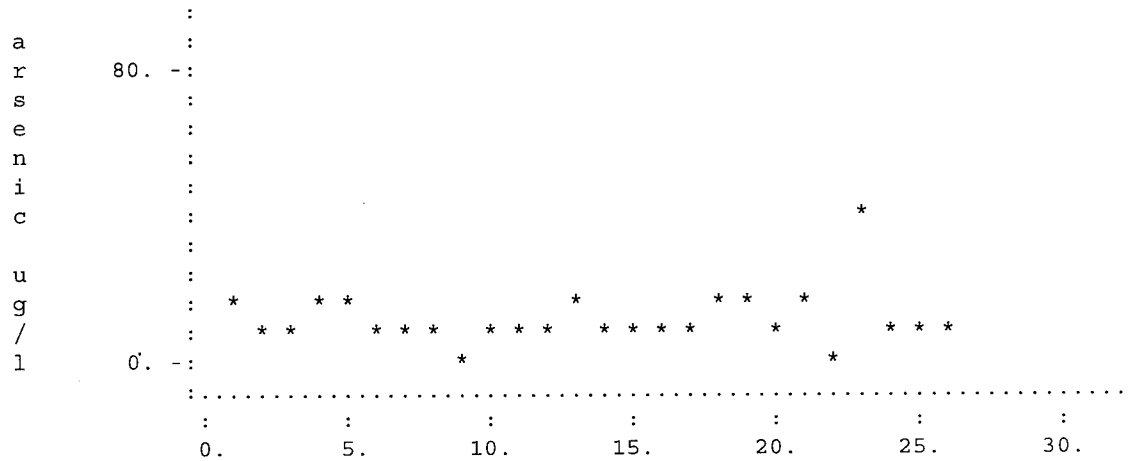
```
155     continue
        do 160 i=4,inobsz
          do 170 j=0,3
            inteff(i)=inteff(i)+x(i-j)*wtir(j)
170       continue
          z(i)=zw(i)-inteff(i)
160     continue
        nobsz=inobsz
        goto 115
      endif
*
* compute residuals and return to plot them and calculate a/c
* and partial a/c
*
      do 180 i=npar+1,inobsz
        z(i-npar)=zw(i)-zpw(i-npar)-const
180   continue
        nobsz=inobsz-npar
      goto 115
      end
```

Appendix B. *Sample Run of Univariate ARIMA Code for Well 2128*

This appendix contains the univariate ARMA code run for well 2128. The first and last iterations are complete. Other iterations have been edited, leaving the plots, autocorrelations, and parameter estimates. The obtained model is ARIMA(2,0,0). The algorithm required 12 iterations to converge.

```
Enter number of usable observations before interventions
26
Which well is being modeled?
Enter 1 for well 2128, 2 for well 2548, 3 for well 2625
1

Output from ACF/A2F

 Mean     =     12.131
 Variance =     37.706

 Lag          ACV           AC          SEAC

   0          37.706       1.00000
   1         -10.005      -0.26534       0.18531
   2           0.466       0.01235       0.18157
   3          -2.551      -0.06766       0.17775
   4          -3.057      -0.08107       0.17384
   5           3.198       0.08482       0.16984
   6          -1.573      -0.04171       0.16575
partial autocorrelations:   -0.265336   -6.24519E-02   -8.72093E-02   -0.133396
    2.22229E-02   -2.79972E-02  0.  0.  0.  0.
Enter when ready to view time series
```

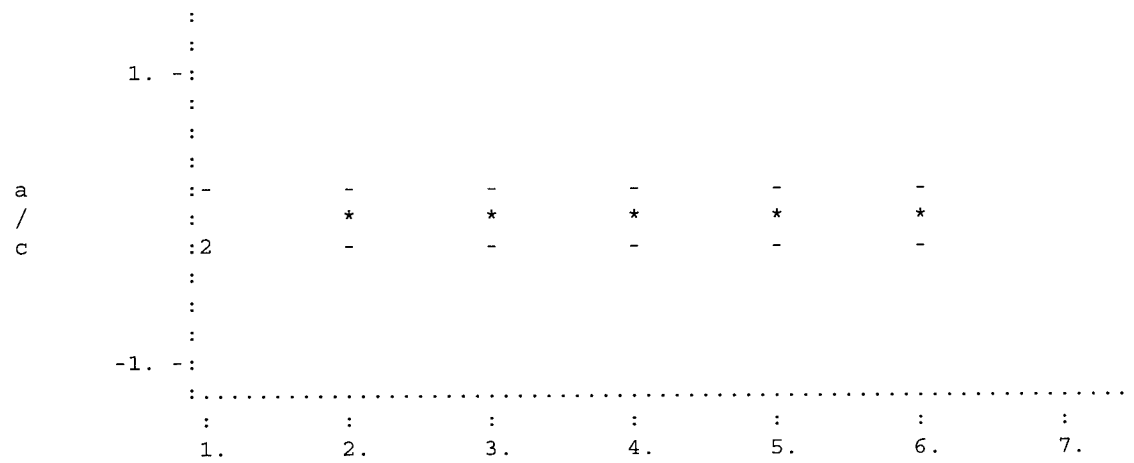
TIME SERIES PLOT



time period

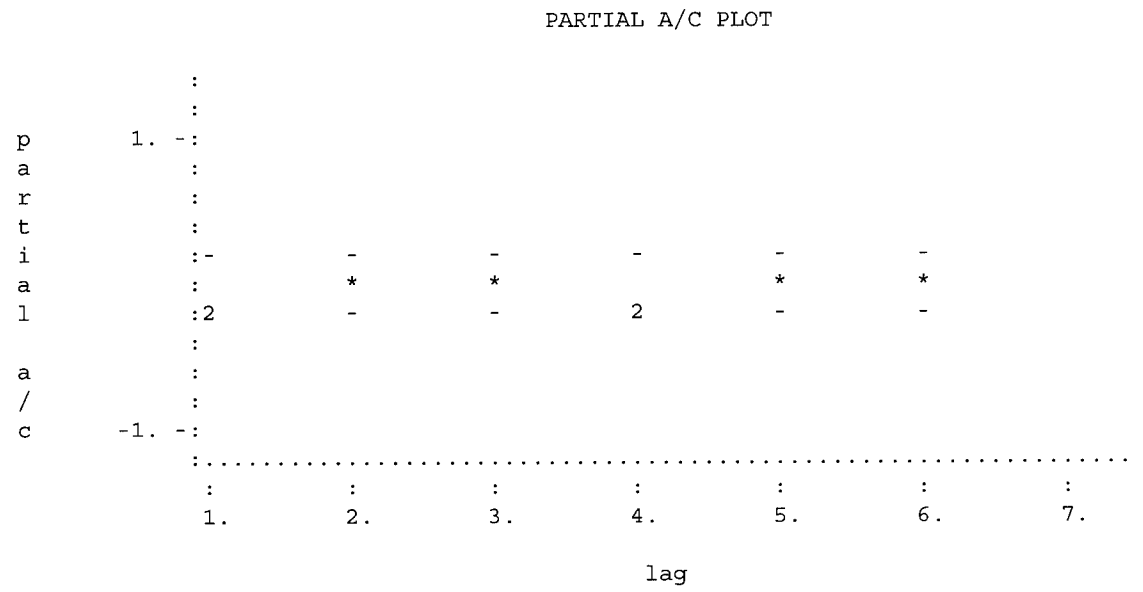
Enter when ready to view autocorrelations

AUTOCORRELATION PLOT



lag

Enter when ready to view partial autocorrelations




```
Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
1
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     12.1308
CONST =     15.3495
AVAR  =     35.0513

     PAR
    -0.2653
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function


            Output WTIR from IRNSE/I2NSE
         1            2            3            4
   677.544      669.263      522.313      297.427

                Output SNOISE from IRNSE/I2NSE
         1            2            3            4            5
   16.0000      18.3000      10.0000      10.0000      10.0000

         6            7            8            9           10
    2.9000      10.0000      10.0000      10.0000      17.1000

        11           12           13           14           15
   11.8000      11.6000      10.0000      10.0000      15.5000

        16           17           18           19           20
   12.1000      10.0000      13.9000       2.7000      37.3000
```

21	22	23	24	25
10.0000	10.0000	10.0000	0.3299	-6.0193
26	27	28	29	30
-0.6494	-47.7818	-14.9815	-27.1083	-3.4908
31	32	33	34	35
-14.5253	-9.9532	22.6629	13.4875	5.1946
36	37	38		
7.3945	23.9445	3.1923		

Output XPW from IRNSE/I2NSE

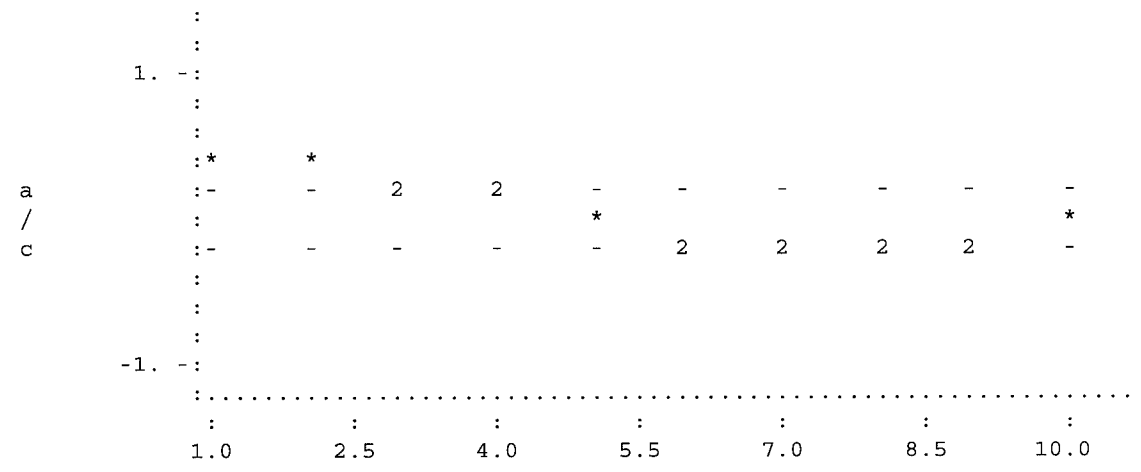
1	2	3	4	5
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
6	7	8	9	10
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
11	12	13	14	15
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
16	17	18	19	20
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
21	22	23	24	25
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
26	27	28	29	30
0.0266700	0.0337465	0.0337465	0.0337465	0.0183965
31	32	33	34	35
0.0143236	0.0143236	0.0143236	0.0075756	0.0057851
36	37	38	39	40
0.0057851	0.0057851	0.0057851	0.0057851	0.0012131

Output YPW from IRNSE/I2NSE

1	2	3	4	5
0.0058	0.0058	0.0012	22.5454	14.8557
6	7	8	9	10
12.6534	12.6534	5.5534	10.7695	12.6534
11	12	13	14	15
12.6534	19.7534	16.3373	14.7310	13.0779
16	17	18	19	20
12.6534	18.1534	16.2127	13.2106	16.5534
21	22	23	24	25
6.3882	38.0164	19.8970	12.6534	12.6534
26	27	28	29	30
21.0534	34.7822	57.1336	23.0546	35.0534
31	32	33	34	35
18.5969	28.2534	16.7926	12.6534	40.7534

```
          36            37            38            39            40
      35.5093       21.8395       21.3066       38.4403       18.9816
Autoregressive coefficients:   -0.265336  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.
Impulse response weights:        677.544    669.263    522.313    297.427
Enter Y to perform another iteration
Y

Output from ACF/A2F

 Mean     =     7.8373
 Variance =     179.85

 Lag          ACV           AC          SEAC

   0         179.846       1.00000
   1          71.187       0.39582       0.15063
   2          69.594       0.38697       0.14873
   3          31.131       0.17310       0.14681
   4          25.065       0.13937       0.14487
   5          -6.974      -0.03878       0.14290
   6         -25.581      -0.14224       0.14090
   7         -45.453      -0.25273       0.13887
   8         -22.512      -0.12517       0.13681
   9         -27.268      -0.15162       0.13473
  10         -14.219      -0.07906       0.13260
partial autocorrelations:    0.395820   0.273076   -5.93686E-02   -2.60051E-03
  -0.138588  -0.166708  -0.153224   0.114104    2.29406E-02    1.56943E-02
Enter when ready to view time series

1

                                TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                          *
             :
 u           :                                                 *    *
 g           : *   **          *      ** *       * *          *
 /           :   **   ***  ***  ****   *    ** * *             **
 l       0. -:           *                *            * **       *
             :.............................................................
              :         :         :         :         :         :         :
              0.        8.       16.       24.       32.       40.       48.

                                    time period

Enter when ready to view autocorrelations

1
```

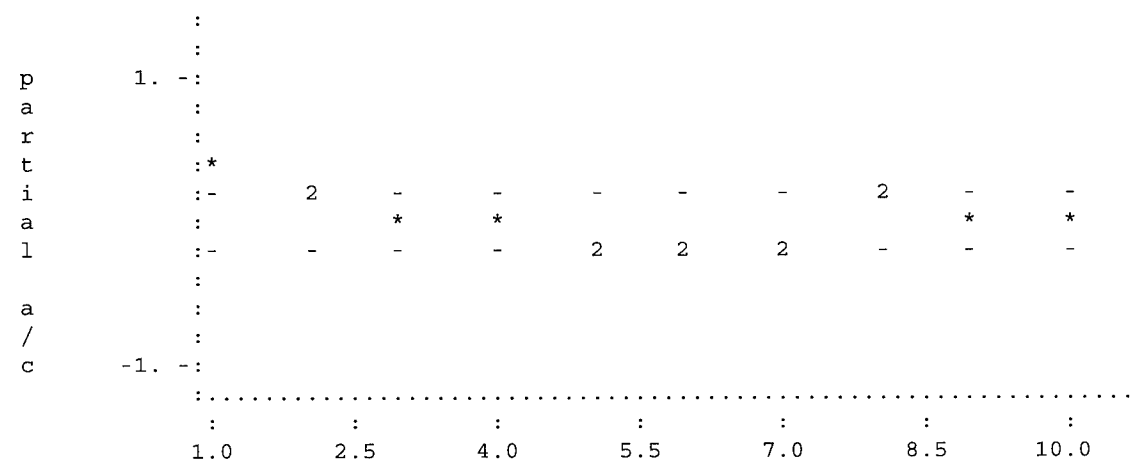
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

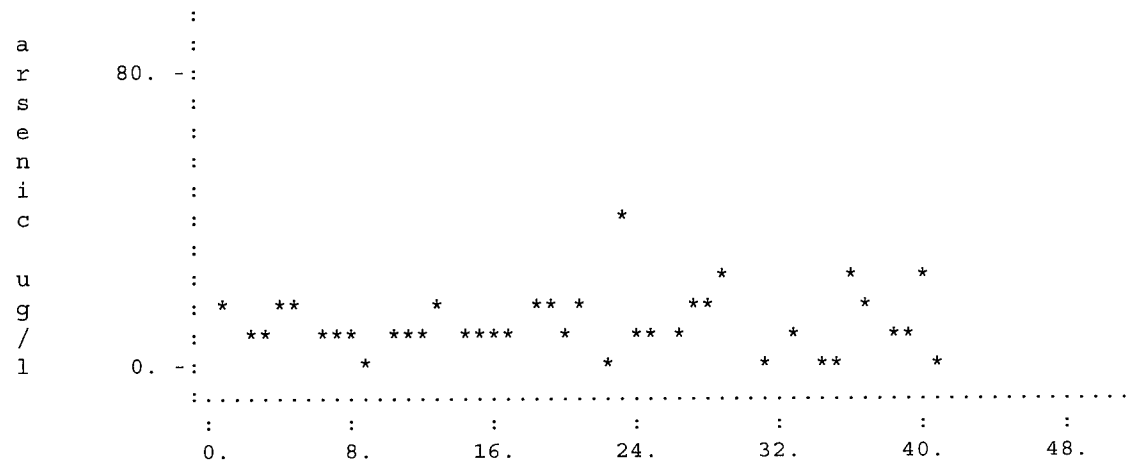
Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 7.83728
CONST = 3.44208
AVAR = 140.359

```
          PAR
            1           2
       0.2877      0.2731
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function


            Output WTIR from IRNSE/I2NSE
            1           2           3           4
      462.632     579.009     349.721    -136.389

                Output SNOISE from IRNSE/I2NSE
            1           2           3           4           5
      16.0000     18.3000     10.0000     10.0000     10.0000

            6           7           8           9          10
       2.9000     10.0000     10.0000     10.0000     17.1000

           11          12          13          14          15
      11.8000     11.6000     10.0000     10.0000     15.5000

           16          17          18          19          20
      12.1000     10.0000     13.9000      2.7000     37.3000

           21          22          23          24          25
      10.0000     10.0000     10.0000      6.0616      2.1194

           26          27          28          29          30
      12.0924    -23.4701      6.0313     -7.4809     13.4873

           31          32          33          34          35
      -4.2063     -1.0844     30.9227     20.5826      9.3623

           36          37          38
      11.5623     28.1123      6.3774

                  Output XPW from IRNSE/I2NSE
            1           2           3           4           5
  0.0000000   0.0000000   0.0000000   0.0000000   0.0000000

            6           7           8           9          10
  0.0000000   0.0000000   0.0000000   0.0000000   0.0000000

           11          12          13          14          15
  0.0000000   0.0000000   0.0000000   0.0000000   0.0000000

           16          17          18          19          20
  0.0000000   0.0000000   0.0000000   0.0000000   0.0000000

           21          22          23          24          25
  0.0000000   0.0000000   0.0000000   0.0000000   0.0266700

           26          27          28          29          30
  0.0189962   0.0117133   0.0117133  -0.0036367   0.0007800

           31          32          33          34          35
  0.0049717   0.0049717  -0.0017763   0.0001653   0.0020080
```

```
        36            37            38            39
 0.0020080     0.0020080     0.0020080    -0.0025640

                  Output YPW from IRNSE/I2NSE
         1             2             3             4             5
    0.0020       -0.0026       10.9655        0.3653        2.1254

         6             7             8             9            10
    4.3919       -2.7081        6.4348        6.3308        4.3919

        11            12            13            14            15
   11.4919        4.1490        3.5352        3.4400        3.9550

        16            17            18            19            20
    9.8919        4.9094        2.2858        7.7185       -4.0302

        21            22            23            24            25
   32.7274       -1.4697       -3.0630        4.3919       12.7919

        26            27            28            29            30
   21.8750       35.5723      -12.3213       16.0874       -2.0532

        31            32            33            34            35
   13.8750       -0.0967        0.1320       32.4919       11.7067

        36            37            38            39
   -2.6125        6.0191       24.7488       -4.4639
Autoregressive coefficients:    0.287731   0.273076  0.  0.
Moving Average coefficients:   0.  0.  0.  0.
Impulse response weights:        462.632    579.009    349.721   -136.389
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     9.8791
 Variance =     122.63

 Lag          ACV           AC          SEAC

   0         122.632       1.00000
   1          20.400       0.16635       0.15063
   2          28.308       0.23083       0.14873
   3           4.631       0.03776       0.14681
   4           8.465       0.06903       0.14487
   5          -9.974      -0.08134       0.14290
   6         -21.191      -0.17280       0.14090
   7         -34.347      -0.28008       0.13887
   8          -9.692      -0.07903       0.13681
   9         -17.736      -0.14462       0.13473
  10          -5.534      -0.04513       0.13260
partial autocorrelations:    0.166352   0.208944   -2.94361E-02    2.10963E-02
   -1.02081E-01  -0.180913  -0.220309    5.48345E-02   -2.74350E-02    4.87368E-
03
Enter when ready to view time series
```

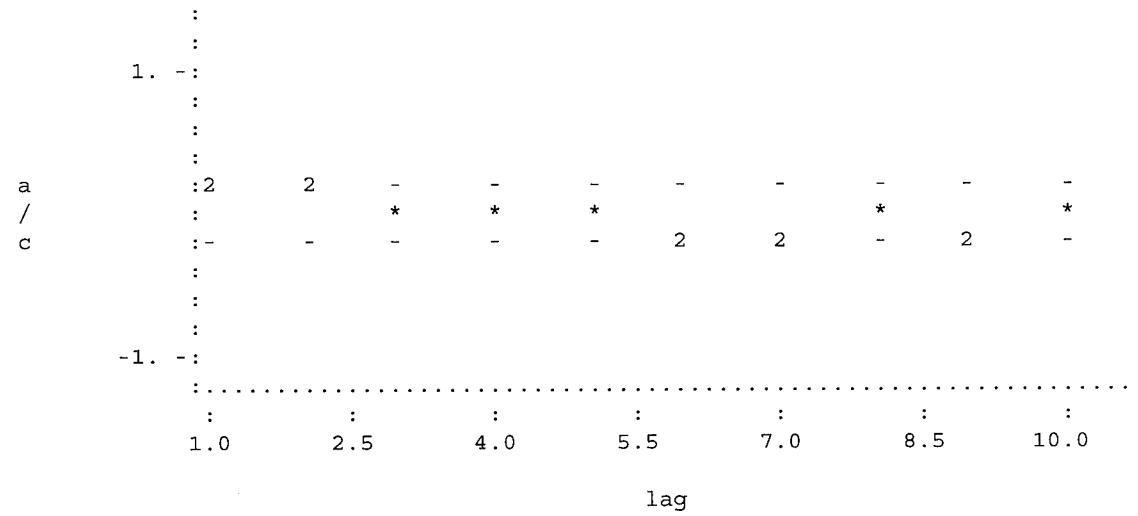
TIME SERIES PLOT



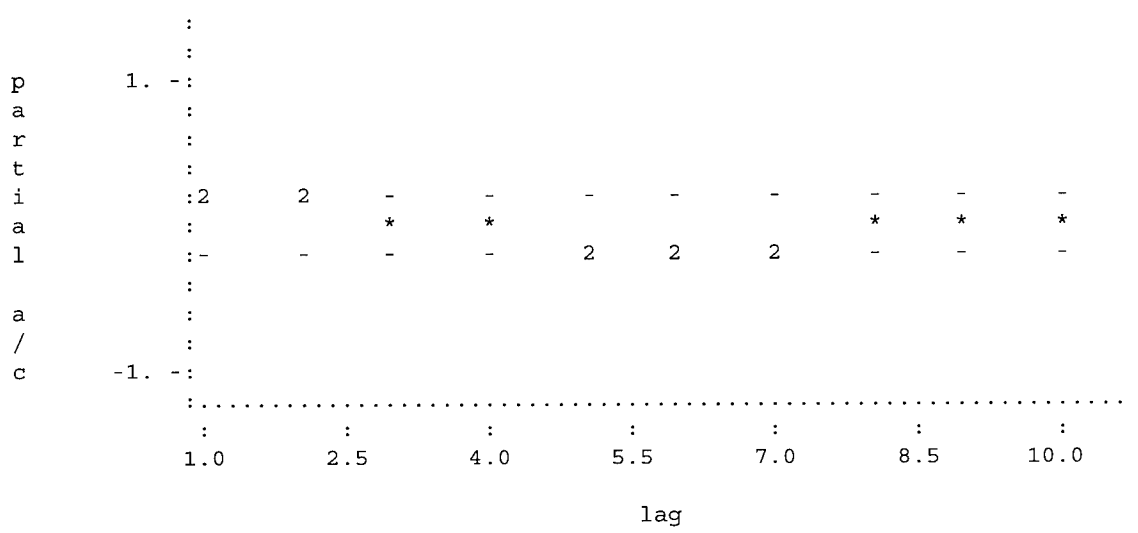
Enter when ready to view autocorrelations

1

AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.87912
CONST = 6.51491
AVAR = 114.032

PAR	
1	2
0.1316	0.2089

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.131594 0.208944 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 546.453 608.133 412.194 54.2666
Enter Y to perform another iteration
Y

Output from ACF/A2F

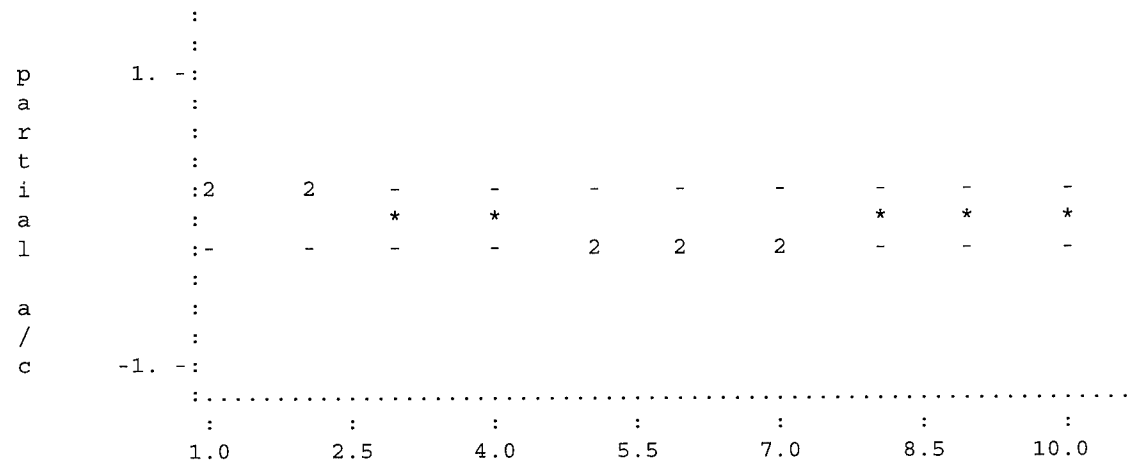
Mean = 9.1288
Variance = 140.59

Lag	ACV	AC	SEAC
0	140.592	1.00000	
1	36.289	0.25811	0.15063
2	41.260	0.29347	0.14873
3	12.792	0.09099	0.14681
4	13.608	0.09679	0.14487

```
   5         -9.627      -0.06848       0.14290
   6        -23.148      -0.16464       0.14090
   7        -38.536      -0.27410       0.13887
   8        -14.386      -0.10233       0.13681
   9        -21.081      -0.14994       0.13473
  10         -8.537      -0.06072       0.13260
partial autocorrelations:    0.258113   0.243044   -3.31190E-02    1.55090E-02
  -0.118489  -0.178455  -0.195100    8.05619E-02   -1.71583E-03    1.61135E-02
Enter when ready to view time series

1
                                    TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                              *
             :
 u           :                                                  *    *
 g           : *   **          *      ** *       ***           *
 /           :   **   ***  ***  ****   *    ** *                 **
 l       0. -:            *                *          * * **       *
             :.....................................................................
              :         :         :         :         :         :         :
              0.        8.       16.       24.       32.       40.       48.

                                        time period

Enter when ready to view autocorrelations

1
                                  AUTOCORRELATION PLOT

             :
             :
         1. -:
             :
             :
             :
 a           :2      2     -      -      -     -      -      -      -      -
 /           :              *      *      *                                  *
 c           :-      -     -      -      -     2      2      2      2      -
             :
             :
             :
        -1. -:
             :.....................................................................
              :         :         :         :         :         :         :
             1.0       2.5       4.0       5.5       7.0       8.5      10.0

                                          lag

Enter when ready to view partial autocorrelations
```

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.12876
CONST = 5.12649
AVAR = 123.474

PAR	
1	2
0.1954	0.2430

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.195380 0.243044 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 513.400 596.466 385.198 -23.5860
Enter Y to perform another iteration
y

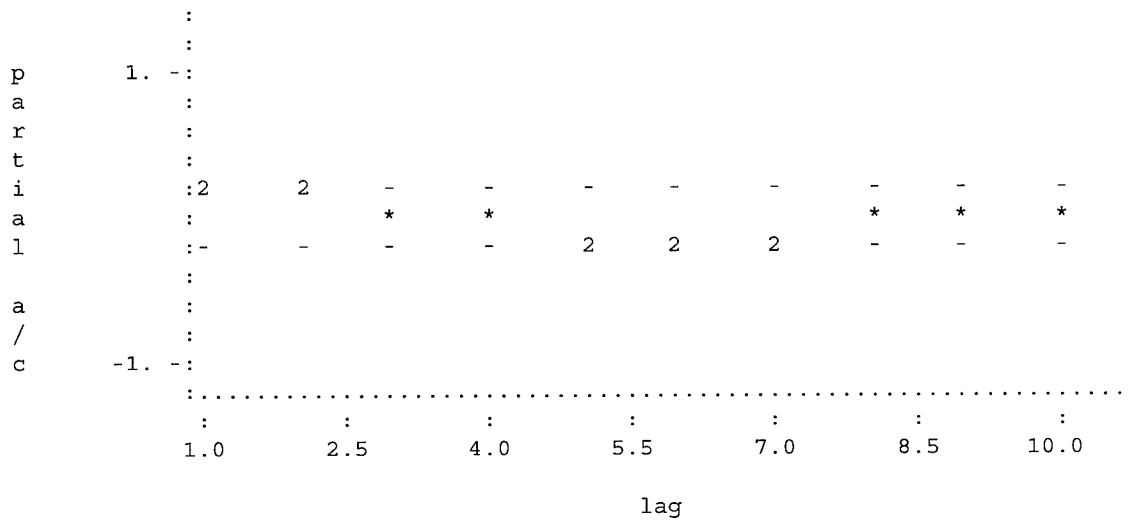
Output from ACF/A2F

Mean = 9.4352
Variance = 132.79

Lag	ACV	AC	SEAC
0	132.787	1.00000	
1	29.404	0.22144	0.15063
2	35.643	0.26842	0.14873
3	9.254	0.06969	0.14681

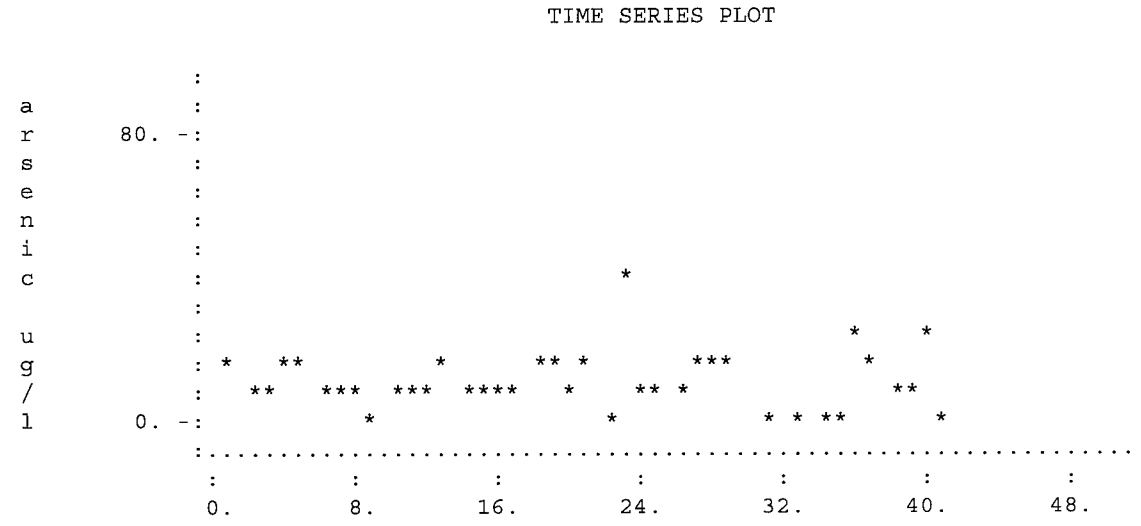
```
   4         11.396      0.08583      0.14487
   5         -9.872     -0.07435      0.14290
   6        -22.400     -0.16869      0.14090
   7        -36.841     -0.27744      0.13887
   8        -12.477     -0.09396      0.13681
   9        -19.688     -0.14826      0.13473
  10         -7.273     -0.05477      0.13260
partial autocorrelations:    0.221436   0.230703   -3.01264E-02    1.87170E-02
  -0.112215  -0.180051  -0.205505    7.05095E-02   -1.09322E-02    1.31630E-02
Enter when ready to view time series

1
                                  TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                              *
             :
 u           :                                                  *    *
 g           : *   **          *      ** *       ***          *
 /           :   **   ***  ***  ****   *    ** *              **
 l       0. -:           *                *          * * **      *
             :.....................................................................
              :         :         :         :         :         :         :
              0.        8.       16.       24.       32.       40.       48.

                                       time period

Enter when ready to view autocorrelations

1
                                AUTOCORRELATION PLOT

             :
             :
         1. -:
             :
             :
             :
 a           :2      2     -      -      -     -      -      -     -      -
 /           :              *      *      *                  *           *
 c           :-      -     -      -      -     2      2      -     2      -
             :
             :
             :
        -1. -:
             :.....................................................................
              :         :         :         :         :         :         :
             1.0       2.5       4.0       5.5       7.0       8.5      10.0

                                         lag

Enter when ready to view partial autocorrelations
```

PARTIAL A/C PLOT

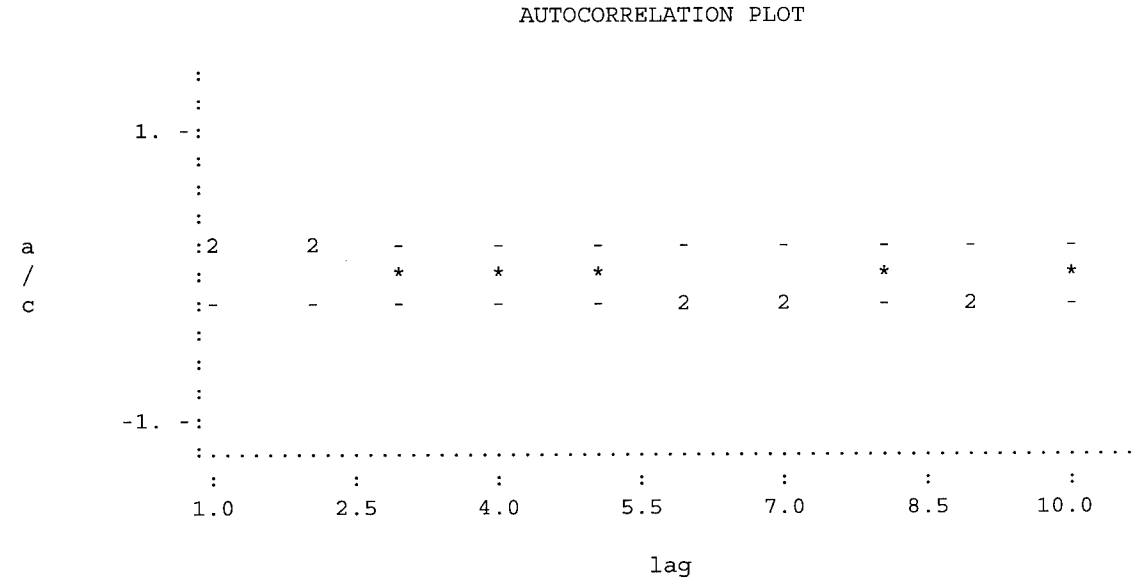



```
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     9.43523
CONST =     5.65121
AVAR  =     119.555

          PAR
         1           2
     0.1703      0.2307
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients:    0.170350   0.230703  0.  0.
Moving Average coefficients:   0.  0.  0.  0.
Impulse response weights:        526.699    601.146    395.827    7.50249
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     9.3130
 Variance =     135.83

 Lag         ACV           AC          SEAC

   0        135.828      1.00000
   1         32.087      0.23623      0.15063
   2         37.832      0.27852      0.14873
   3         10.630      0.07826      0.14681
   4         12.258      0.09024      0.14487
   5         -9.791     -0.07209      0.14290
```

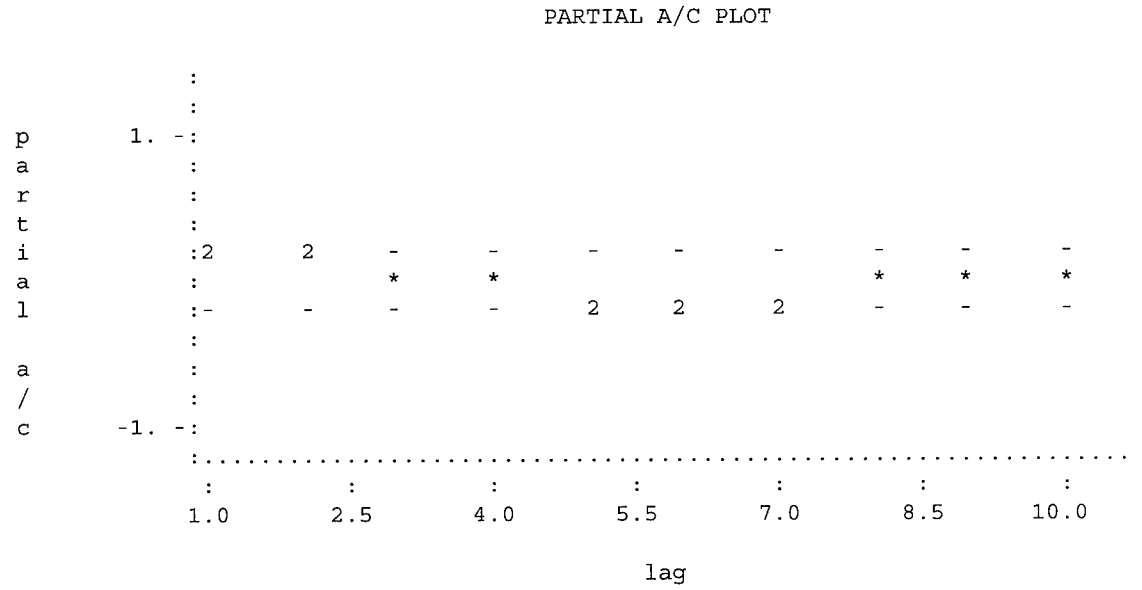
```
   6        -22.707      -0.16718       0.14090
   7        -37.521      -0.27624       0.13887
   8        -13.239      -0.09747       0.13681
   9        -20.240      -0.14901       0.13473
  10         -7.772      -0.05722       0.13260
partial autocorrelations:    0.236234   0.235882   -3.11118E-02    1.75445E-02
  -0.114782  -0.179495  -0.201356    7.46140E-02   -7.06457E-03    1.45618E-02
Enter when ready to view time series

1
```




```
Enter when ready to view autocorrelations

1
```




```
Enter when ready to view partial autocorrelations

1
```



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on
Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.31295
CONST = 5.43511
AVAR = 121.112

PAR	
1	2
0.1805	0.2359

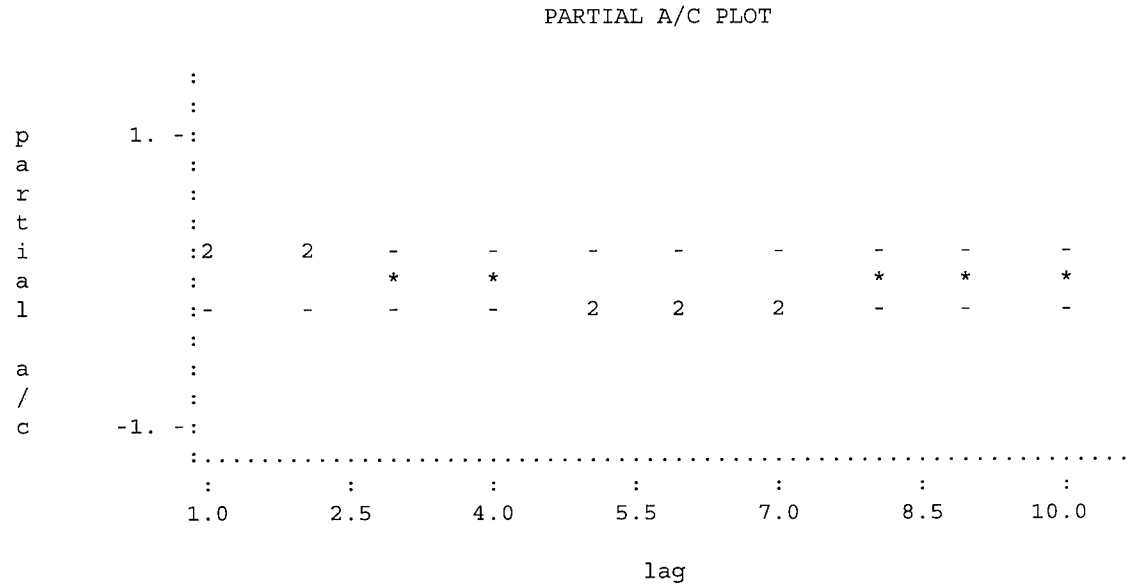
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.180511 0.235882 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 521.344 599.259 391.505 -5.05989
Enter Y to perform another iteration
y

Output from ACF/A2F

Mean = 9.3624
Variance = 134.59

Lag	ACV	AC	SEAC
0	134.587	1.00000	
1	30.992	0.23028	0.15063
2	36.939	0.27446	0.14873
3	10.068	0.07481	0.14681
4	11.906	0.08846	0.14487

```
   5         -9.827      -0.07301       0.14290
   6        -22.585      -0.16781       0.14090
   7        -37.246      -0.27675       0.13887
   8        -12.931      -0.09608       0.13681
   9        -20.016      -0.14872       0.13473
  10         -7.570      -0.05624       0.13260
partial autocorrelations:    0.230277   0.233830   -3.06784E-02    1.80377E-02
  -0.113755  -0.179733  -0.203034    7.29693E-02   -8.59645E-03    1.40341E-02
Enter when ready to view time series

1
                                  TIME SERIES PLOT

            :
 a          :
 r      80. -:
 s          :
 e          :
 n          :
 i          :
 c          :                              *
            :
 u          :                                                  *    *
 g          : *   **          *      ** *       ***           *
 /          :   **   ***  ***  ****   *    ** *                 **
 l       0. -:           *                *          * * **       *
            :.....................................................................
             :         :         :         :         :         :         :
             0.        8.       16.       24.       32.       40.       48.

                                     time period

Enter when ready to view autocorrelations

1
                                AUTOCORRELATION PLOT

            :
            :
         1. -:
            :
            :
            :
 a          :2      2     -       -       -     -       -       -     -       -
 /          :               *       *       *                       *             *
 c          :-      -     -       -       -     2       2       -     2       -
            :
            :
            :
        -1. -:
            :.....................................................................
             :         :         :         :         :         :         :
            1.0       2.5       4.0       5.5       7.0       8.5      10.0

                                        lag

Enter when ready to view partial autocorrelations
```



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.36238
CONST = 5.52135
AVAR = 120.482

PAR	
1	2
0.1764	0.2338

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.176431 0.233830 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 523.501 600.018 393.240 -5.38505E-03
Enter Y to perform another iteration
y

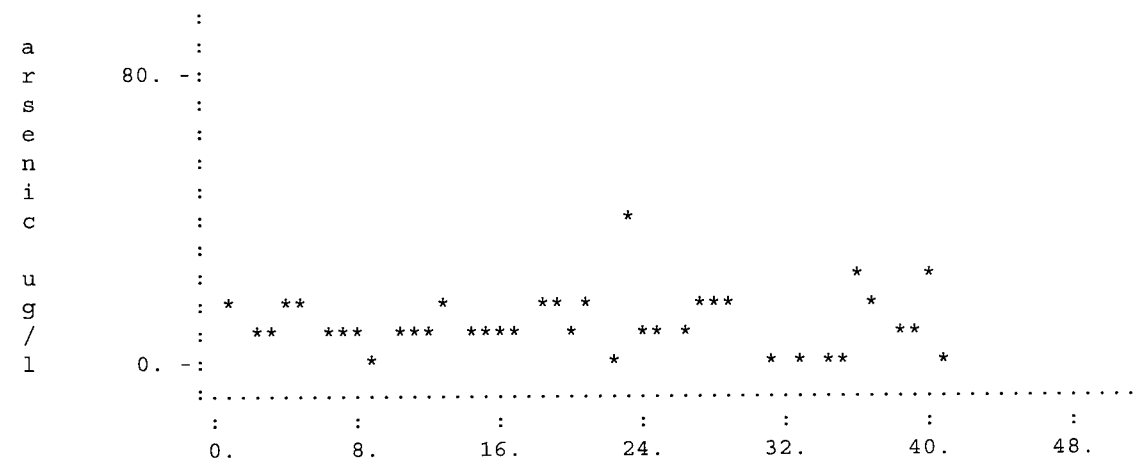
Output from ACF/A2F

Mean = 9.3425
Variance = 135.08

Lag	ACV	AC	SEAC
0	135.085	1.00000	
1	31.431	0.23268	0.15063
2	37.296	0.27610	0.14873
3	10.293	0.07620	0.14681
4	12.047	0.08918	0.14487

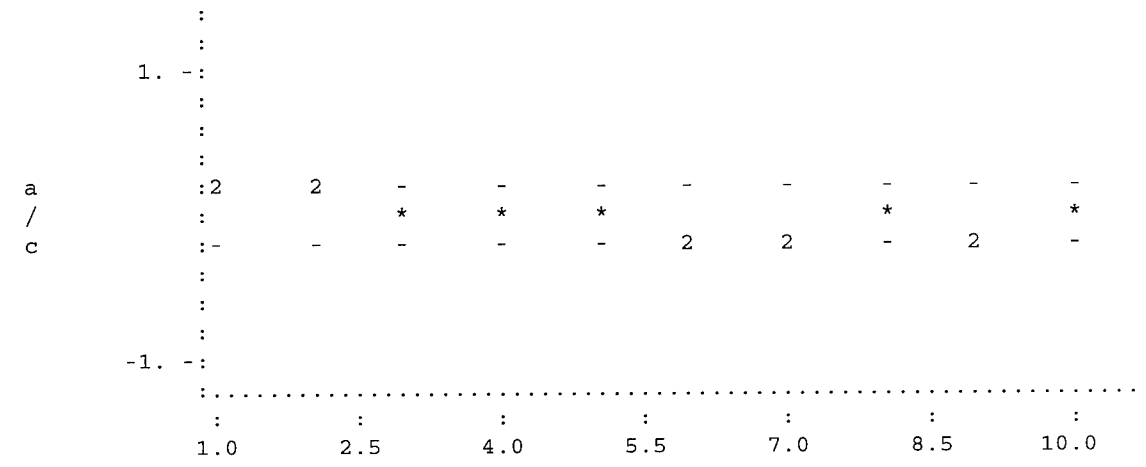
```
   5         -9.813      -0.07264       0.14290
   6        -22.634      -0.16756       0.14090
   7        -37.357      -0.27654       0.13887
   8        -13.055      -0.09664       0.13681
   9        -20.106      -0.14884       0.13473
  10         -7.651      -0.05664       0.13260
partial autocorrelations:    0.232677   0.234662   -3.08471E-02    1.78423E-02
  -0.114170  -0.179640  -0.202359    7.36332E-02   -7.97520E-03    1.42523E-02
Enter when ready to view time series
```

TIME SERIES PLOT



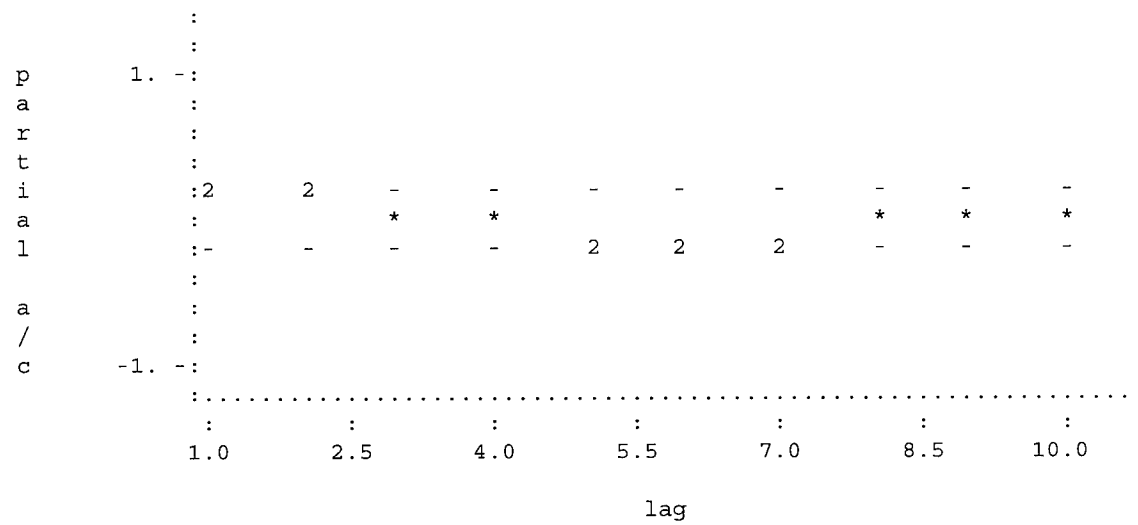

```
Enter when ready to view autocorrelations

1
```

AUTOCORRELATION PLOT




```
Enter when ready to view partial autocorrelations

1
```

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.34249
CONST = 5.48648
AVAR = 120.735

PAR	
1	2
0.1781	0.2347

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.178077 0.234662 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 522.632 599.712 392.540 -2.04250
Enter Y to perform another iteration
y

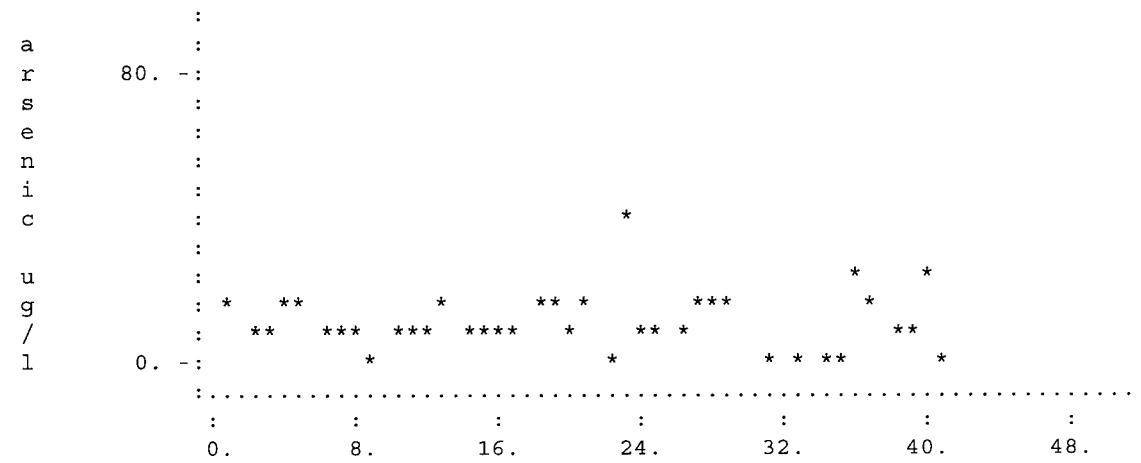
Output from ACF/A2F

Mean = 9.3505
Variance = 134.88

Lag	ACV	AC	SEAC
0	134.884	1.00000	
1	31.254	0.23171	0.15063
2	37.152	0.27544	0.14873
3	10.202	0.07564	0.14681
4	11.990	0.08889	0.14487

```
   5         -9.819      -0.07279       0.14290
   6        -22.614      -0.16766       0.14090
   7        -37.312      -0.27663       0.13887
   8        -13.005      -0.09642       0.13681
   9        -20.070      -0.14879       0.13473
  10         -7.618      -0.05648       0.13260
partial autocorrelations:    0.231711   0.234328   -3.07782E-02    1.79215E-02
  -0.114003  -0.179678  -0.202631    7.33661E-02   -8.22468E-03    1.41653E-02
Enter when ready to view time series
```

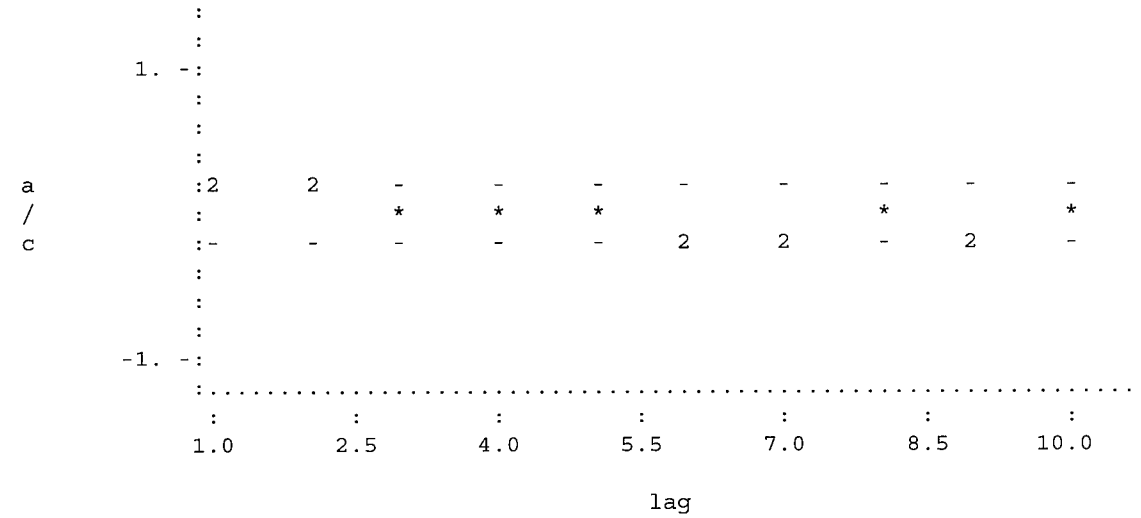
TIME SERIES PLOT



time period

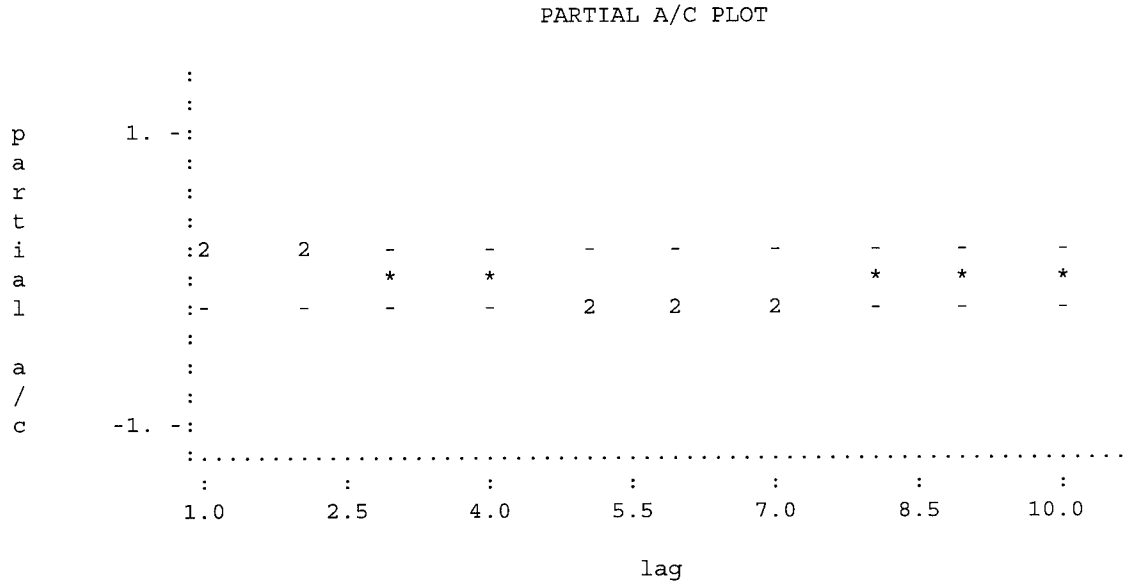
Enter when ready to view autocorrelations

AUTOCORRELATION PLOT



lag

```
Enter when ready to view partial autocorrelations
1
```



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 9.35051
CONST = 5.50051
AVAR = 120.633

PAR	
1	2
0.1774	0.2343

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.177414 0.234328 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 522.983 599.836 392.822 -1.22204
Enter Y to perform another iteration
y

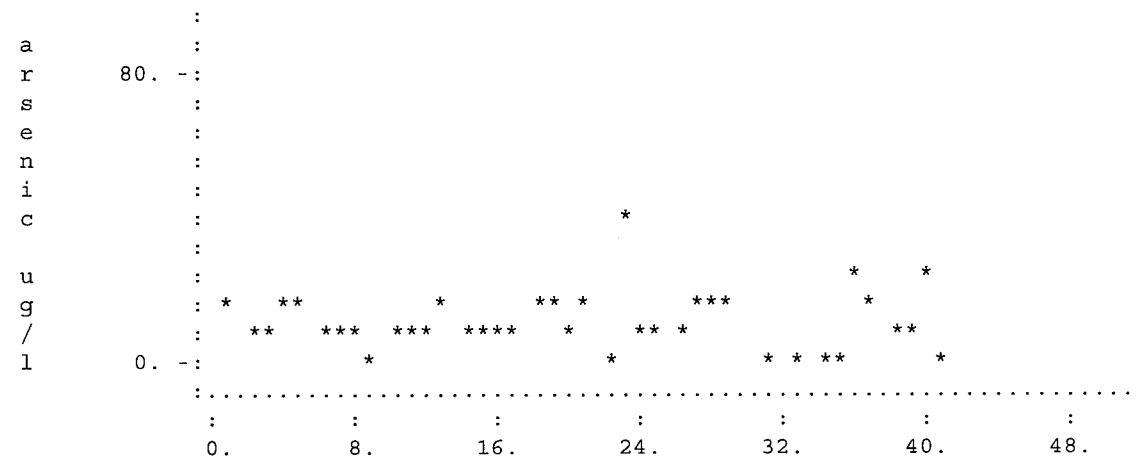
Output from ACF/A2F

Mean = 9.3473
Variance = 134.96

Lag	ACV	AC	SEAC
0	134.965	1.00000	
1	31.325	0.23210	0.15063
2	37.210	0.27570	0.14873
3	10.239	0.07586	0.14681
4	12.013	0.08901	0.14487

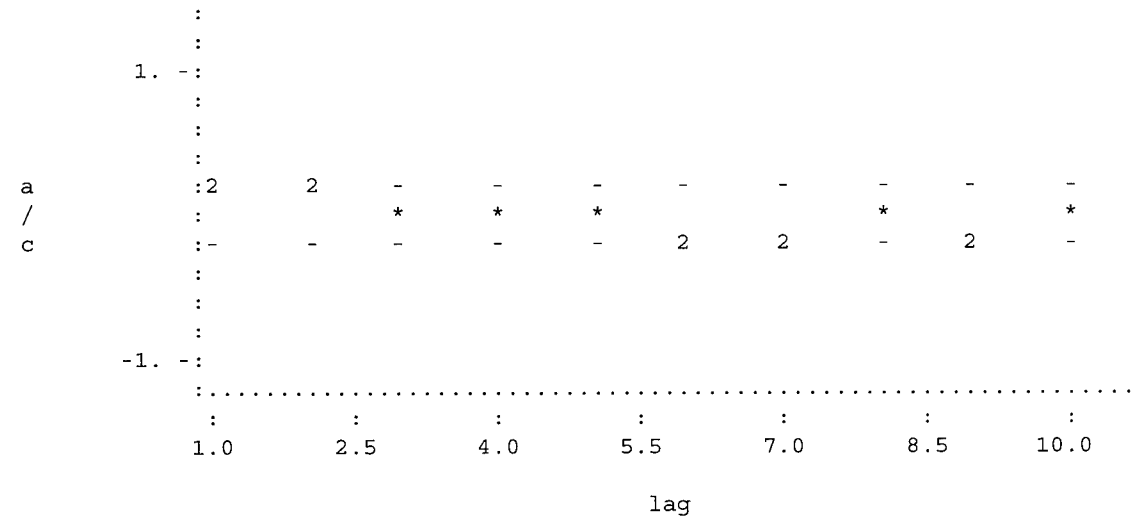
```
   5         -9.816      -0.07273       0.14290
   6        -22.622      -0.16762       0.14090
   7        -37.330      -0.27659       0.13887
   8        -13.025      -0.09651       0.13681
   9        -20.084      -0.14881       0.13473
  10         -7.631      -0.05654       0.13260
partial autocorrelations:    0.232100   0.234463   -3.08058E-02    1.78896E-02
  -0.114070  -0.179663  -0.202522    7.34738E-02   -8.12405E-03    1.42005E-02
Enter when ready to view time series
```

TIME SERIES PLOT



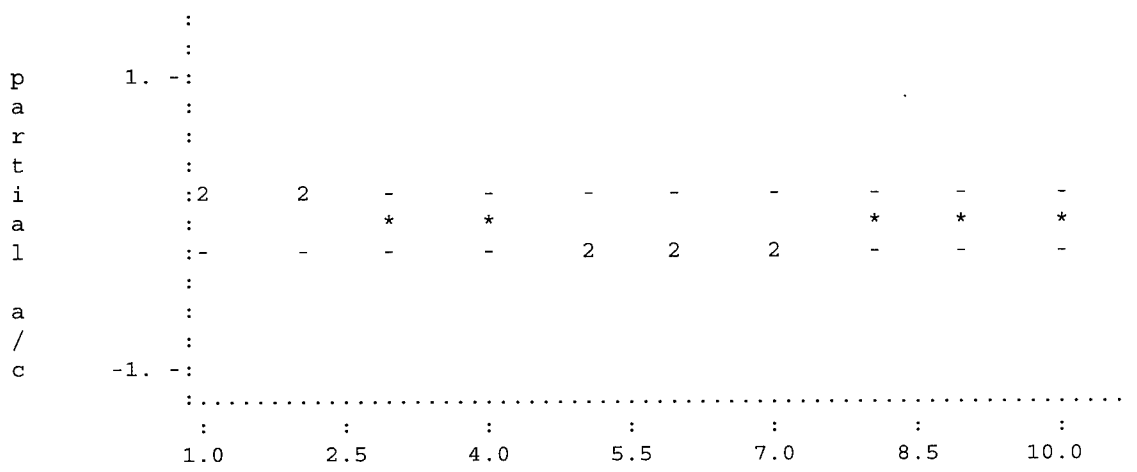

```
Enter when ready to view autocorrelations

1
```

AUTOCORRELATION PLOT



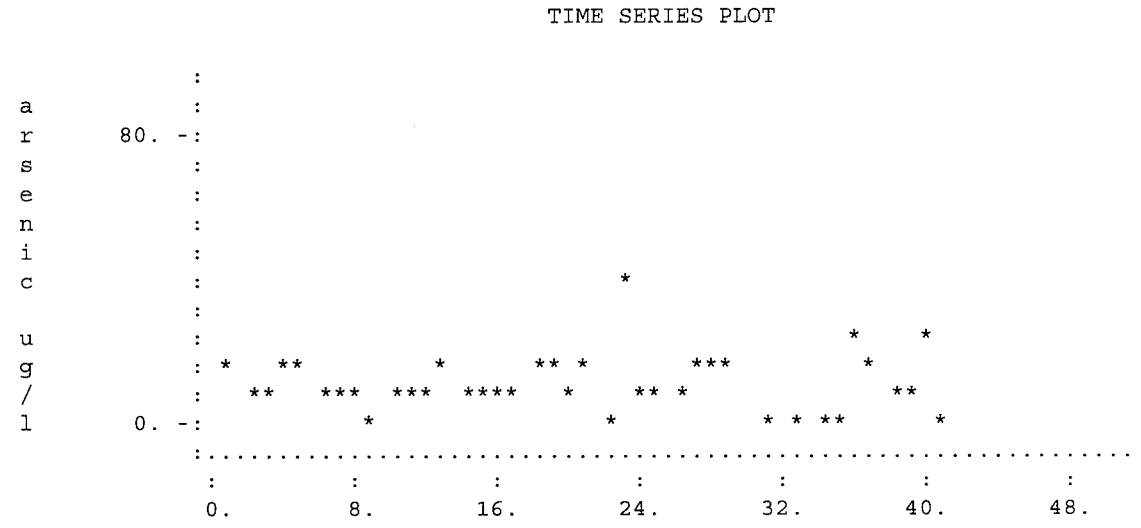

```
Enter when ready to view partial autocorrelations
```

PARTIAL A/C PLOT

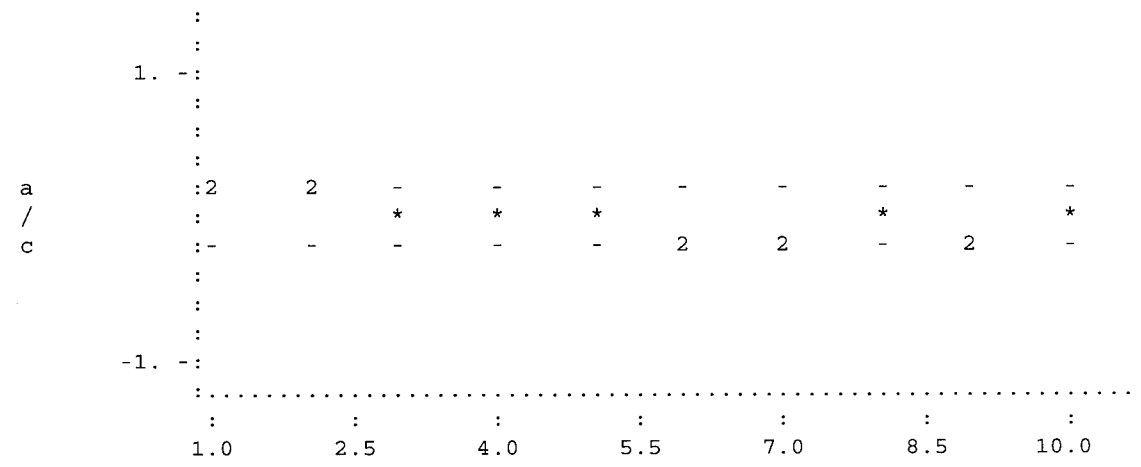



```
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     9.34728
CONST =     5.49485
AVAR  =     120.674

          PAR
         1           2
    0.1777      0.2345
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients:    0.177681   0.234463  0.  0.
Moving Average coefficients:   0.  0.  0.  0.
Impulse response weights:        522.841    599.786    392.708   -1.55277
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     9.3486
 Variance =     134.93

 Lag          ACV           AC          SEAC

   0         134.932       1.00000
   1          31.297       0.23194      0.15063
   2          37.187       0.27560      0.14873
   3          10.224       0.07577      0.14681
   4          12.004       0.08896      0.14487
   5          -9.817      -0.07276      0.14290
```

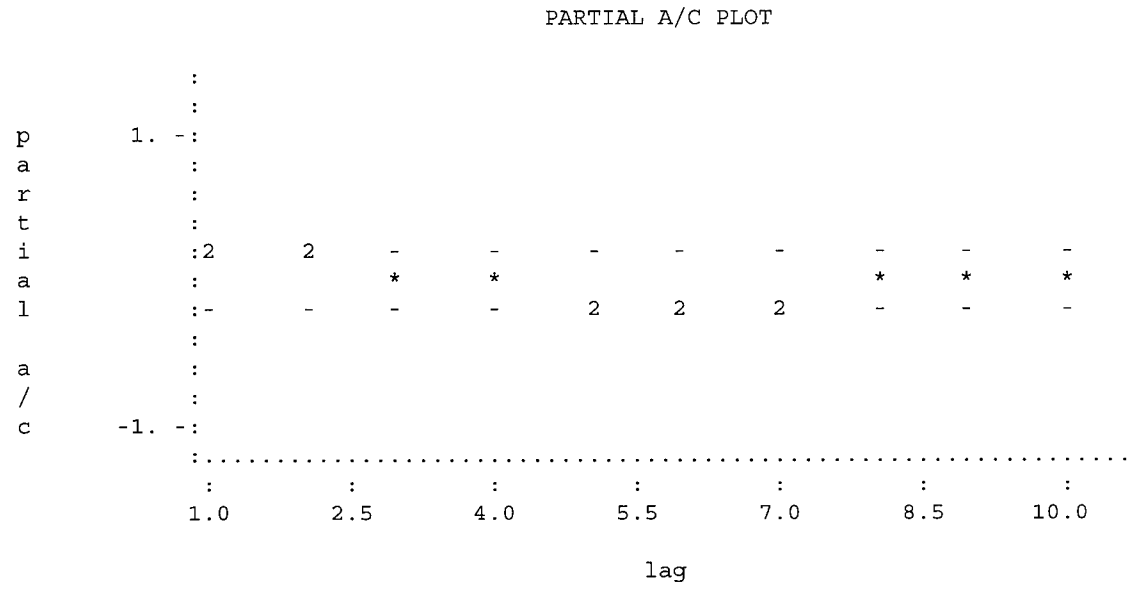
```
   6        -22.619      -0.16763       0.14090
   7        -37.323      -0.27661       0.13887
   8        -13.017      -0.09647       0.13681
   9        -20.078      -0.14880       0.13473
  10         -7.626      -0.05652       0.13260
partial autocorrelations:    0.231943   0.234409   -3.07946E-02    1.79025E-02
  -0.114043  -0.179669  -0.202566    7.34303E-02   -8.16457E-03    1.41863E-02
Enter when ready to view time series
```




```
Enter when ready to view autocorrelations

1
```




```
Enter when ready to view partial autocorrelations


1
```



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on
Enter value of p
2
0Enter value of q

Results from NSPE/N2PE

WMEAN = 9.34858
CONST = 5.49713
AVAR = 120.658

PAR	
1	2
0.1776	0.2344

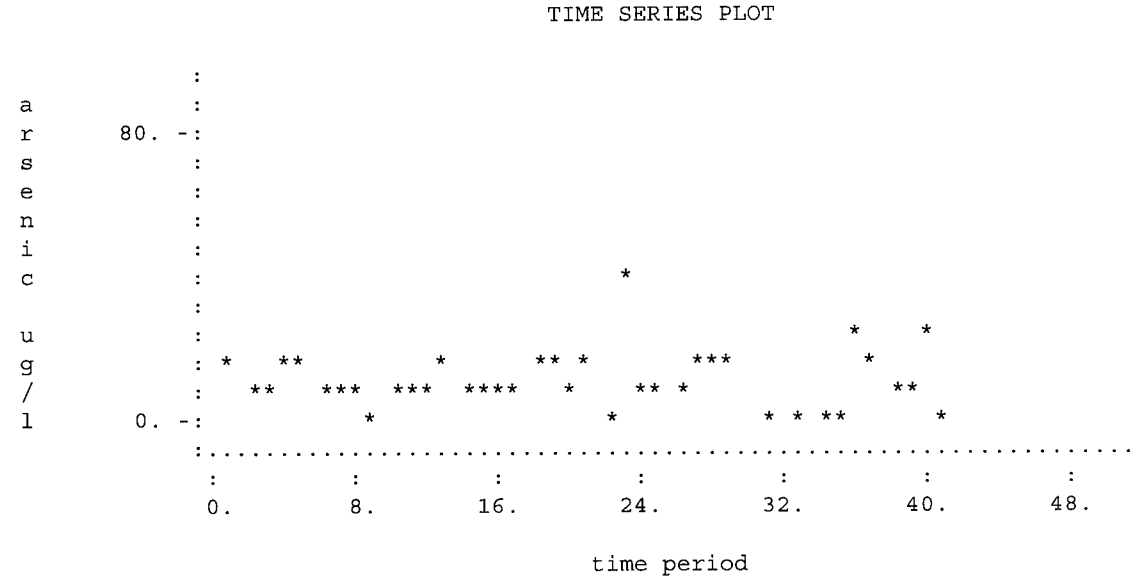
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function
Autoregressive coefficients: 0.177574 0.234409 0. 0.
Moving Average coefficients: 0. 0. 0. 0.
Impulse response weights: 522.898 599.806 392.754 -1.41939
Enter Y to perform another iteration
y

Output from ACF/A2F

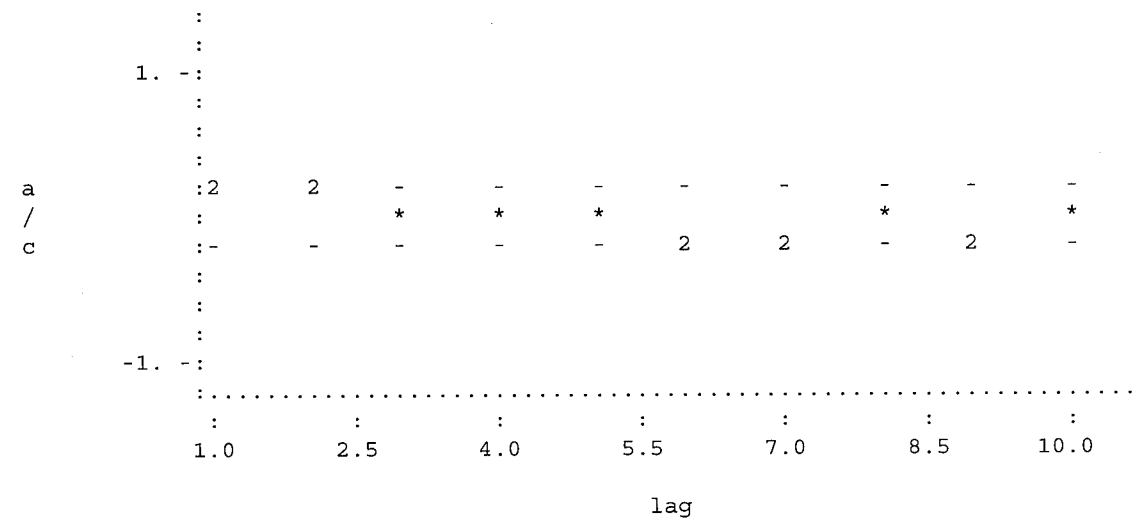
Mean = 9.3481
Variance = 134.95

Lag	ACV	AC	SEAC
0	134.945	1.00000	
1	31.308	0.23201	0.15063
2	37.196	0.27564	0.14873
3	10.230	0.07581	0.14681
4	12.008	0.08898	0.14487

```
   5         -9.817      -0.07275       0.14290
   6        -22.620      -0.16763       0.14090
   7        -37.326      -0.27660       0.13887
   8        -13.020      -0.09649       0.13681
   9        -20.081      -0.14881       0.13473
  10         -7.628      -0.05653       0.13260
partial autocorrelations:    0.232006   0.234430   -3.07991E-02    1.78974E-02
  -0.114054  -0.179666  -0.202548    7.34478E-02   -8.14836E-03    1.41921E-02
Enter when ready to view time series
```

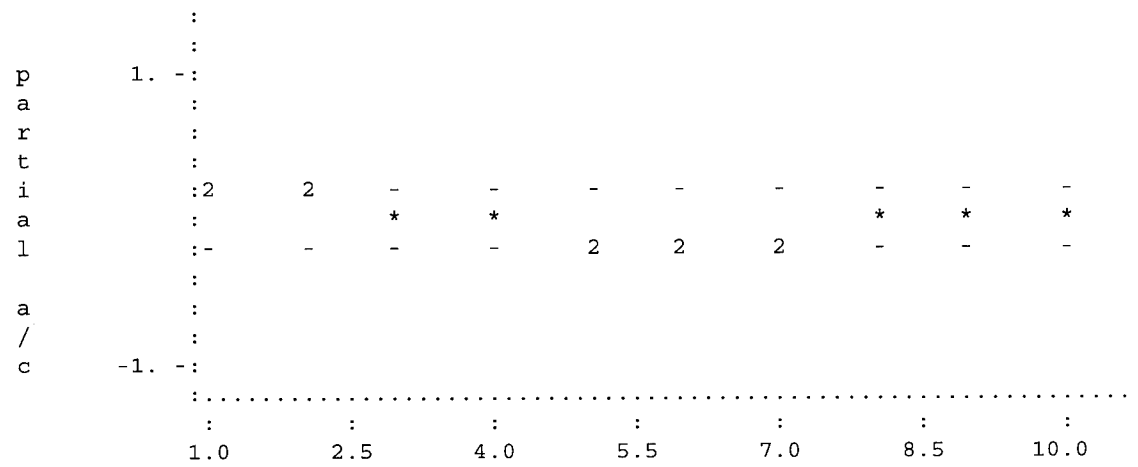



```
Enter when ready to view autocorrelations

1
```




```
Enter when ready to view partial autocorrelations

1
```

PARTIAL A/C PLOT




```
Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     9.34806
CONST =     5.49621
AVAR  =     120.664

          PAR
         1           2
    0.1776      0.2344
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function


            Output WTIR from IRNSE/I2NSE
         1           2           3           4
   522.875     599.798     392.735      -1.473

            Output SNOISE from IRNSE/I2NSE
         1           2           3           4           5
   16.0000     18.3000     10.0000     10.0000     10.0000

         6           7           8           9          10
    2.9000     10.0000     10.0000     10.0000     17.1000

        11          12          13          14          15
   11.8000     11.6000     10.0000     10.0000     15.5000

        16          17          18          19          20
   12.1000     10.0000     13.9000      2.7000     37.3000
```

21	22	23	24	25
10.0000	10.0000	10.0000	4.4549	-0.0417
26	27	28	29	30
8.7841	-30.3767	0.0495	-13.1436	8.4849
31	32	33	34	35
-7.1378	-3.6094	28.5380	18.4882	8.1783
36	37	38		
10.3783	26.9283	5.4689		

Output XPW from IRNSE/I2NSE

1	2	3	4	5
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
6	7	8	9	10
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
11	12	13	14	15
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
16	17	18	19	20
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
21	22	23	24	25
0.0000000	0.0000000	0.0000000	0.0000000	0.0266700
26	27	28	29	30
0.0219330	0.0156807	0.0156807	0.0003307	0.0030571
31	32	33	34	35
0.0066556	0.0066556	-0.0000924	0.0011062	0.0026881
36	37	38	39	
0.0026881	0.0026881	0.0026881	-0.0018839	

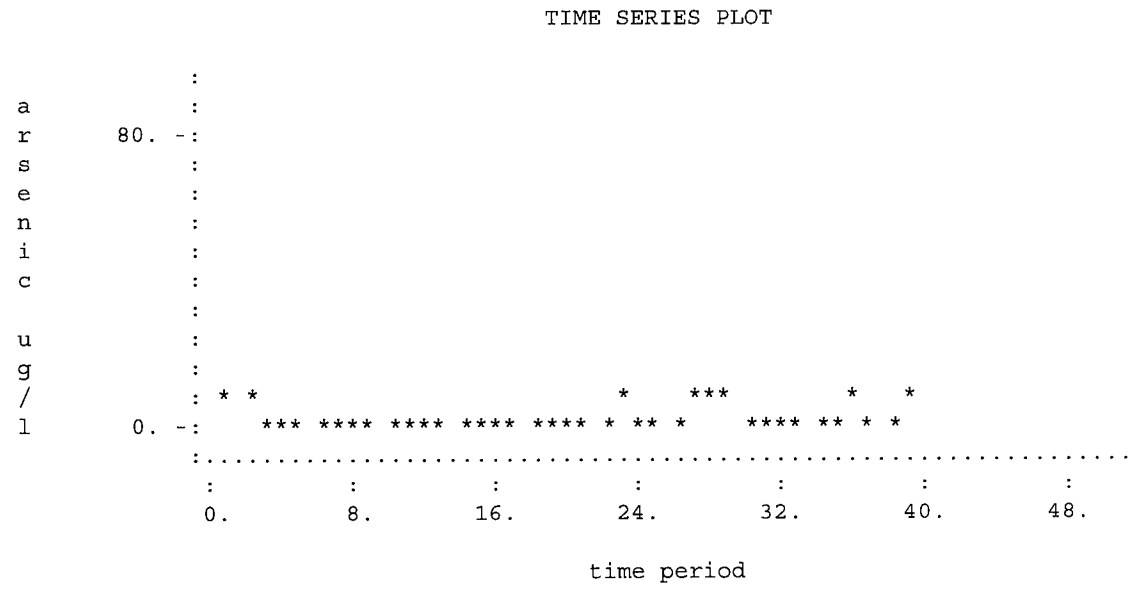
Output YPW from IRNSE/I2NSE

1	2	3	4	5
0.0027	-0.0019	13.1138	2.9987	3.9338
6	7	8	9	10
5.8795	-1.2205	7.1406	7.5440	5.8795
11	12	13	14	15
12.9795	6.4184	5.4954	5.1734	5.5044
16	17	18	19	20
11.3795	7.0026	4.2172	9.2872	-2.1132
21	22	23	24	25
33.5619	2.7419	-0.5204	5.8795	14.2795
26	27	28	29	30
24.2875	39.5757	-5.7482	19.0899	1.9009
31	32	33	34	35

```
      16.2283        3.1087        2.2224       33.9795       16.2885

          36            37            38            39
       1.6567        8.6635       27.2373       -0.0680
Autoregressive coefficients:    0.177617   0.234430  0.  0.
Moving Average coefficients:   0.  0.  0.  0.
Impulse response weights:        522.875    599.798    392.735   -1.47306
Enter Y to perform another iteration
```

The following is a plot of residuals and residual autocorrelations and partial autocorrelations.

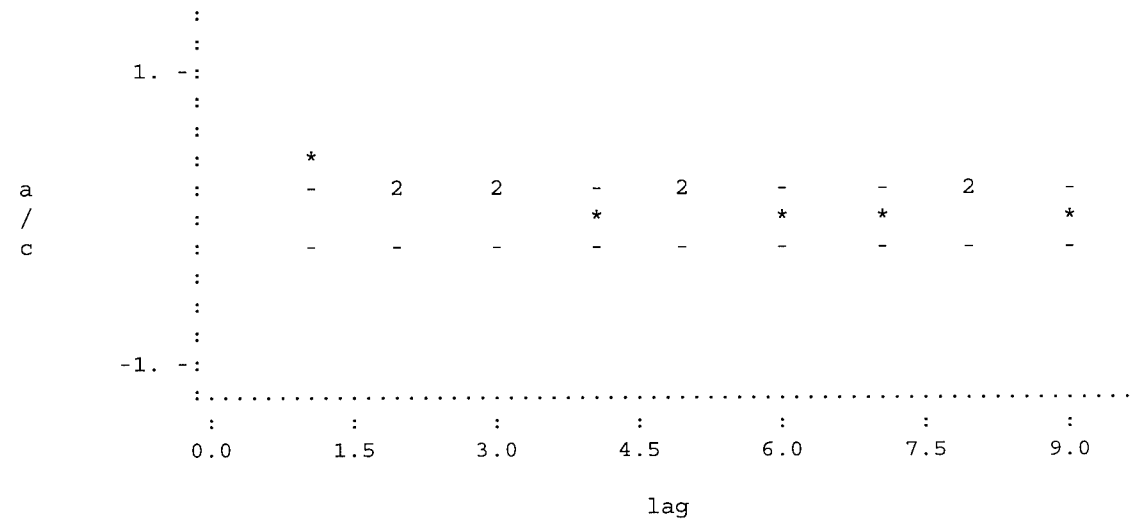
```
Output from ACF/A2F

 Mean     =     1.4122
 Variance =     11.330

 Lag          ACV           AC          SEAC

   0         11.3299      1.00000
   1          5.5524      0.49007      0.15416
   2          2.1391      0.18880      0.15212
   3          1.7246      0.15222      0.15005
   4          0.9731      0.08588      0.14795
   5          1.4342      0.12659      0.14582
   6          0.9534      0.08415      0.14366
   7          0.5943      0.05246      0.14147
   8          1.6917      0.14931      0.13924
   9          0.5375      0.04744      0.13697
partial autocorrelations:    0.490068   -6.75987E-02   0.114453   -3.61117E-02
   0.125030   -4.91324E-02    3.37960E-02   0.131059  -0.120315    1.41888E-02

Enter when ready to view time series
```



Enter when ready to view autocorrelations

1

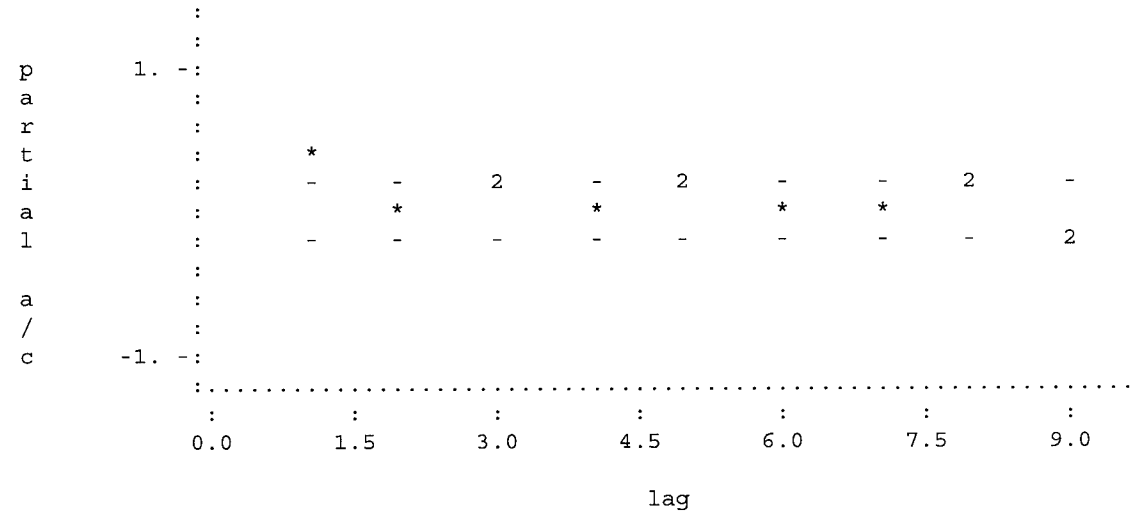
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Appendix C. *FORTRAN Code for Univariate STARIMA and Causal Intervention Transfer Function Model Building*

This appendix contains the FORTRAN code for performing univariate STARIMA and transfer function modeling.

```
************************************************************************
************************************************************************
* STARIMA modeling with intervention analysis
* by Samuel A. Wright
* March 1995
************************************************************************
************************************************************************
      integer i,inobsz,maxlag
      parameter(inobsz=41,maxlag=inobsz/4)
      integer obs(inobsz*3)
      real w2128(inobsz),w2548(inobsz),w2625(inobsz),w2636(inobsz),
     +     w2900(inobsz),w3924(inobsz),w3925(inobsz),
     +     e2128(inobsz),e2548(inobsz),e2625(inobsz),e2636(inobsz),
     +     e2900(inobsz),e3924(inobsz),e3925(inobsz),
     +     wt2128(inobsz),wt2548(inobsz),wt2625(inobsz),wt2636(inobsz),
     +     wt2900(inobsz),wt3924(inobsz),wt3925(inobsz),
     +     zero(inobsz),one(inobsz),two(inobsz),o2636(inobsz),
     +     ezero(inobsz),eone(inobsz),etwo(inobsz),
     +     plevel(5,inobsz),noint,inteff(inobsz),xt(inobsz)
* diff declarations
      integer nobsz,ndiff,iper(1),iord(1),nlost
      real z(inobsz*3),zw(inobsz*3),x(inobsz*3)
* acf declarations
      real zmean,acv(maxlag*3+1),ac(maxlag*3+1),seac(maxlag*3)
* pacf declarations
      real pac(maxlag*3),sepac
* sctp declarations
      character*20 xtitle,ytitle,title
      character*4 symbol
      integer icol(4)
      real range(4),zplot(inobsz*3,2),acplot(maxlag*3+1,4),
     +     pacplot(maxlag*3+1,4)
* nspe declarations
      integer npar,npma
      real const,par(7),pma(7),avar
* irnse declarations
      real wtir(4*3),snoise(inobsz*3-9),xpw(inobsz*3-4),zpw(inobsz*3-4)
* diff data
      data iper(1),iord(1) /1,0/
```

```
* sctp data
      data symbol /' *--'/
      data icol /1,2,2,2/
*
* get data
*
      ndiff=0
      print*,'Enter number of usable observations before interventions'
      read*,ninobsz
      nobsz=ninobsz*3
* determine well of interest
      print*,'Which well is being modeled?'
      print*,'Enter 1 for well 2128, 2 for well 2548, 3 for well 2625'
      read*,welloi
      open(20,file='welldata.dat',status='old')
      do 100 i=1,inobsz
        read(20,110) obs(i),w2128(i),w2548(i),w2625(i),w2636(i),
     +               w2900(i),w3924(i),w3925(i),xt(i),o2636(i)
100   continue
      close(20)
110   format(i4,9f8.3)
      open(23,file='effect.dat',status='old')
      do 107 i=1,inobsz
        read(23,111) e2128(i),e2548(i),e2625(i),e2636(i),e2900(i),
     +         e3924(i),e3925(i)
107   continue
      close(23)
111   format(7f10.6)
* define spatial weights
      do 30 i=1,inobsz
* first order neighbors
           wt2128(i)=0.345
           wt2625(i)=0.655
* second order neighbors
        if (o2636(i).gt.0.5) then
           wt2636(i)=0.563
           wt2900(i)=0.437
           wt3924(i)=0.000
           wt3925(i)=0.000
        else
           wt2636(i)=0.000
           wt2900(i)=1.000
           wt3924(i)=0.000
           wt3925(i)=0.000
        endif
30    continue
* apply weights to zeroth, first, and second order neighbors
      do 50 i=1,inobsz
        zero(i)=w2548(i)
        one(i)=wt2128(i)*w2128(i)+wt2625(i)*w2625(i)
        two(i)=wt2636(i)*w2636(i)+wt2900(i)*w2900(i)+
     +         wt3924(i)*w3924(i)+wt3925(i)*w3925(i)
```

```
        ezero(i)=e2548(i)
        eone(i)=wt2128(i)*e2128(i)+wt2625(i)*e2625(i)
        etwo(i)=wt2636(i)*e2636(i)+wt2900(i)*e2900(i)+
     +          wt3924(i)*e3924(i)+wt3925(i)*e3925(i)
50    continue
*
* define combined time series, zw (with intervention), and intervention
* series, x
*
      do 60 i=1,inobsz
        zw(3*i-2)=zero(i)
        zw(3*i-1)=one(i)
        zw(3*i)=two(i)
        x(i*3-2)=ezero(i)
        x(i*3-1)=eone(i)
        x(i*3)=etwo(i)
60    continue
      newobs=inobsz*3
* z gets time series without interventions (as defined by user)
      do 70 i=1,nobsz
        z(i)=zw(i)
70    continue
      read(*,'(a)') enter
*
* difference data
*
* if differencing, difference noise series, time series with
* intervention, and intervention series x
*
115   if (ndiff.ne.0) then
        call diff(nobsz,z,ndiff,iper,iord,1,1,nlost,nobsz,z)
        call diff(newobs,zw,ndiff,iper,iord,1,1,nlost,newobs,zw)
        call diff(newobs,x,ndiff,iper,iord,1,1,nlost,newobs,x)
      endif
*
* get sample autocorrelations and sample partial autocorrelations
*
      mlag=nobsz/4
      call acf(nobsz,z,3,2,1,zmean,mlag,acv,ac,seac)
      call pacf(mlag,ac,pac)
      sepac=nobsz**(-.5)
* print partial autocorrelations
      print*,'partial autocorrelations: ',pac
*
* plot noise series, time series with interventions, and
* intervention series
*
* prepare plotting matrices for autocorrelations and partial
* autocorrelations
*
      do 117 i=1,nobsz
* prepare plotting matrix for noise series z
```

```
        zplot(i,1)=real(i)
        zplot(i,2)=z(i)
        print116,zplot(i,2)
117   continue
116   format(f8.3)
      do 120 i=1,mlag
        acplot(i,1)=real(i)
        acplot(i,2)=ac(i+1)
        acplot(i,3)=seac(i)
        acplot(i,4)=seac(i)*(-1.)
        pacplot(i,1)=real(i)
        pacplot(i,2)=pac(i)
        pacplot(i,3)=sepac
        pacplot(i,4)=sepac*(-1.)
        print118,acplot(i,2),acplot(i,3),acplot(i,4),
     +           pacplot(i,2),pacplot(i,3),pacplot(i,4)
120   continue
118   format(6f8.3)
      call page(-2,20)
* print time series, autocorrelations, and partial autocorrelations
      print*,'Enter when ready to view time series'
      read(*,'(a)') enter
130   title='TIME SERIES PLOT'
      xtitle='time period'
      ytitle='arsenic ug/l'
      range(1)=1
      range(2)=nobsz
      range(3)=0
      range(4)=80
      call sctp(nobsz,2,zplot,newobs,icol,range,symbol,xtitle,ytitle,
     +          title)
      print*,'Enter when ready to view autocorrelations'
      read(*,'(a)') enter
      title='AUTOCORRELATION PLOT'
      xtitle='lag'
      ytitle='a/c'
      range(1)=1
      range(2)=mlag
      range(3)=-1
      range(4)=1
      call sctp(mlag,4,acplot,maxlag+1,icol,range,symbol,xtitle,ytitle,
     +          title)
      print*,'Enter when ready to view partial autocorrelations'
      read(*,'(a)') enter
      title='PARTIAL A/C PLOT'
      xtitle='lag'
      ytitle='partial a/c'
      call sctp(mlag,4,pacplot,maxlag+1,icol,range,symbol,xtitle,ytitle,
     +          title)
      print*,'Enter R to repeat plotting or Enter to go on'
      read(*,'(a)') enter
* query to repeat plotting or go on
```

```
      if ((enter.eq.'R').or.(enter.eq.'r')) goto 130
* query to difference time series if its mean appears non-stationary
      print*,'Enter D to difference the time series or Enter to go on'
      read(*,'(a)') enter
      if ((enter.eq.'D').or.(enter.eq.'d')) then
        ndiff=1
        iord(1)=iord(1)+1
* go back to difference all relevant time series and repeat a/c,
* partial a/c and plotting
        goto 115
      endif
*
* ARMA cofficient estimation
*
* first, query for values of p and q
*
      print*,'Enter value of p'
      read(*,'(i1)') npar
      print*,'Enter value of q'
      read(*,'(i1)') npma
      do 140 i=1,npar
        lagar(i)=i
140   continue
      do 150 i=1,npma
        lagma(i)=i
150   continue
      call nspe(nobsz,z,1,1,zmean,npar,npma,0.0,0,const,par,pma,avar)
*
* estimate transfer function impulse response weights
*
      print*,'Enter to estimate transfer function'
      read(*,'(a)') enter
      call
irnse(newobs,x,zw,1,npar,par,npma,pma,11,11,wtir,snoise,xpw,zpw
     +           )
      print*,'Autoregressive coefficients: ',par
      print*,'Moving Average coefficients: ',pma
      print*,'Impulse response weights:    ',wtir
      print*,'Enter Y to perform another iteration'
      read(*,'(a)') enter
      if ((enter.eq.'Y').or.(enter.eq.'y')) then
*
* repeat entire process using time series minus estimated intervention
* effect as noise series
*
* re-initialize effect of intervention
*
        do 155 i=1,newobs
          inteff(i)=0.0
155     continue
        do 160 i=4,newobs
          do 170 j=0,3
```

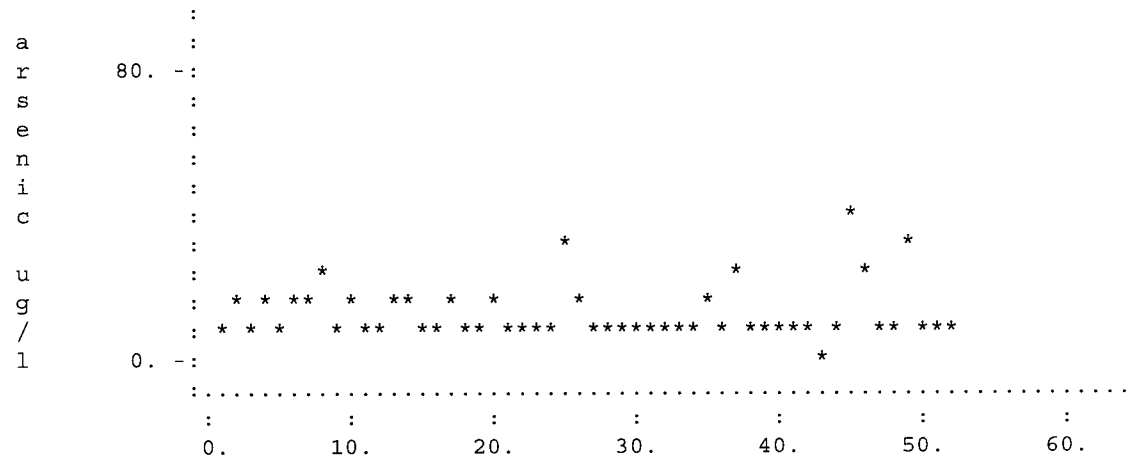
```
            inteff(i)=inteff(i)+x(i-j)*wtir(j)
170         continue
            z(i)=zw(i)-inteff(i)
160       continue
          nobsz=newobs
          goto 115
        endif
        do 180 i=npar+1,inobsz*2
          z(i-npar)=zw(i)-zpw(i-npar)-const
180     continue
          nobsz=inobsz*2-npar
        goto 115
        end
```

Appendix D. *Sample Run of Univariate STARIMA Code for Well 2548*

This appendix contains the univariate STARMA code run for well 2548. The obtained model is STARIMA$(1,0,0)^2(1,0,0)$. The process required 10 iterations to converge.

```
Enter number of usable observations before interventions
26

Output from ACF/A2F

 Mean     =     12.730
 Variance =     39.532

 Lag         ACV

   0        39.5315
   1         1.0945
   2        -9.8004
   3        -1.6167
   4         8.2635
   5        -5.3174
   6        -8.0455
   7        -5.7288
   8         5.4406
   9         1.1331
  10         0.8789
  11        -0.3735
  12         6.8972
  13        -2.4359
partial autocorrelations:     2.76875E-02  -0.248870   -2.71337E-02   0.159566
  -0.177474  -0.121685  -0.206802    3.56280E-02   -1.46741E-02    7.87291E-02
    3.89043E-03   0.111253  -0.111123  0.  0.  0.  0.  0.  0.  0.
Enter when ready to view time series
```

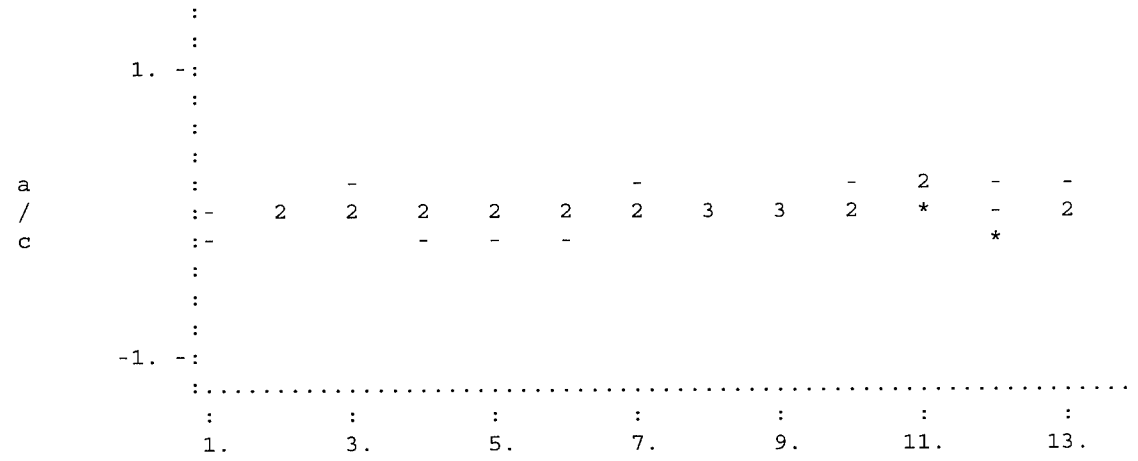
TIME SERIES PLOT



Enter when ready to view autocorrelations

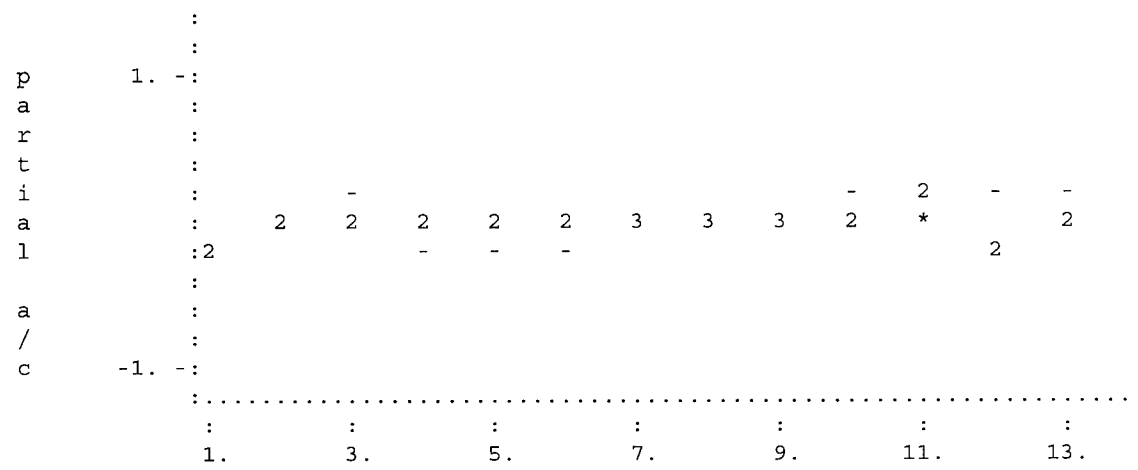
1

AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 12.7299
CONST = 15.4579
AVAR = 37.0546

PAR	
1	2
0.0346	-0.2489

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients: 3.45781E-02 -0.248870 0. 0. 0. 0. 0.
Moving Average coefficients: 0. 0. 0. 0. 0. 0. 0.
Impulse response weights: 224.289 246.915 211.241 220.038
170.627
146.453 116.568 122.562
Enter Y to perform another iteration
y

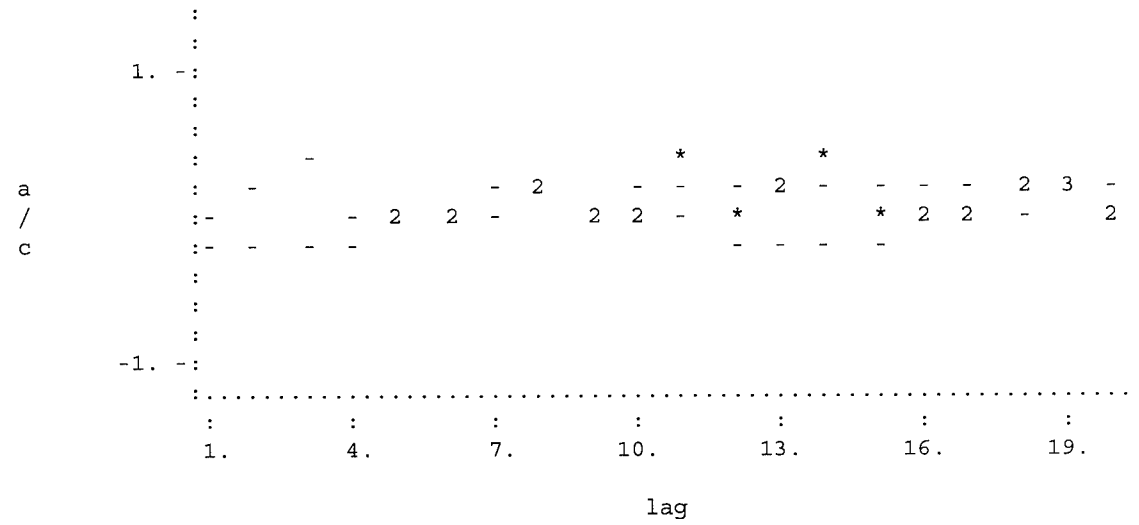
Output from ACF/A2F

Mean = 10.507
Variance = 57.085

Lag	ACV
0	57.085
1	20.568

```
   2          3.813
   3          8.785
   4         17.515
   5          2.163
   6         -5.257
   7         -1.622
   8         11.512
   9          6.260
  10         -0.442
  11          0.258
  12          4.561
  13         -6.686
  14        -12.318
  15         -9.509
  16         -7.521
  17         -0.887
  18         -3.578
  19         -0.768
  20          8.574
partial autocorrelations:    0.360315   -7.24262E-02   0.178123   0.227436
  -0.180161   -5.66786E-02   -3.22975E-02   0.199271    3.19630E-02   -2.82255E-
03
   -1.88649E-02   -5.90360E-02  -0.198720   -9.07670E-02   -3.81001E-02   -
9.61878E-02
   0.209141   -4.81550E-02    6.85013E-02   0.158700
Enter when ready to view time series

1
                                TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                          *
             :                  *               *                *
 u           :     *                    *     *
 g           : * *** * 2 * *   *     *          *  *     *   ** 2
 /           : ***  *** *****2* *2*2*****2** ******2 * * *   * *   2
 l       0. -:                             *         ***** 2**    * **
             :.............................................................
              :         :         :         :         :         :         :
              0.       15.       30.       45.       60.       75.       90.

                                    time period

Enter when ready to view autocorrelations
```

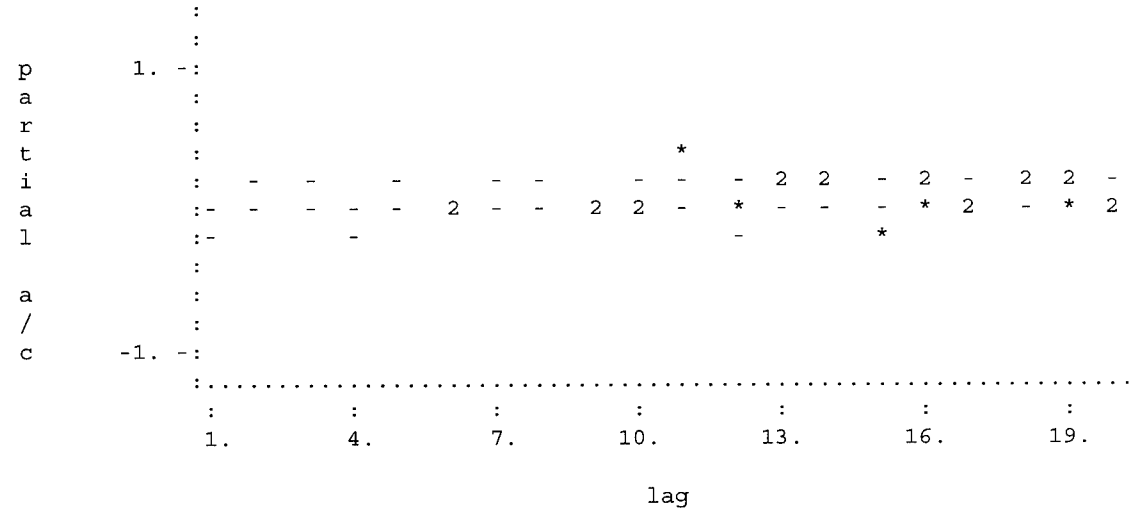
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

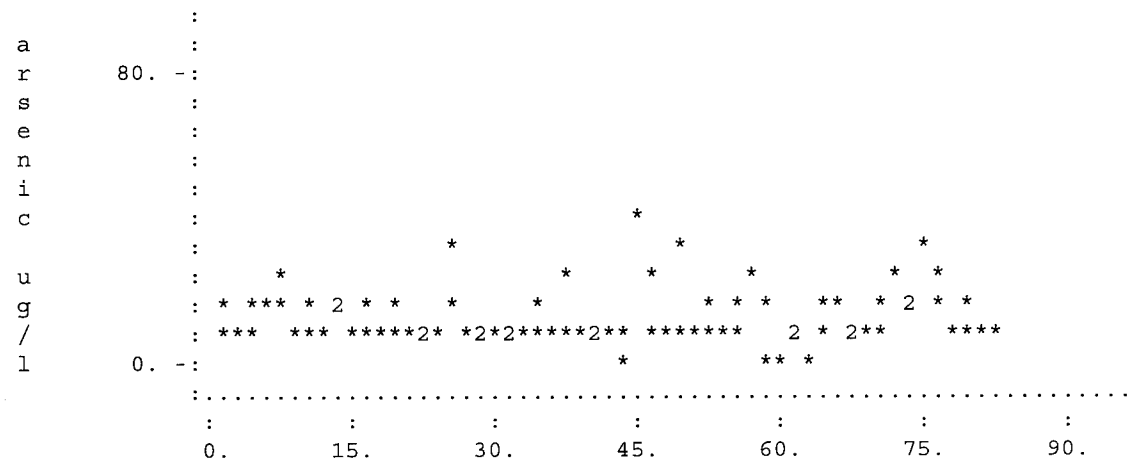
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 10.50750
CONST = 7.20830
AVAR = 49.4129

```
          PAR
          1           2
     0.3864     -0.0724
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.386412   -7.24262E-02  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        96.0169    152.192    89.8595    145.567
75.8942
    69.6388    32.6139    81.4953
Enter Y to perform another iteration
Y

Output from ACF/A2F

 Mean     =     12.613
 Variance =     46.674

 Lag         ACV

   0         46.674
   1         10.946
   2         -6.751
   3         -0.687
   4          7.702
   5         -5.353
   6        -12.556
   7         -8.408
   8          5.527
   9          2.089
  10         -2.370
  11         -1.235
  12          3.923
  13         -6.751
  14        -10.271
  15         -7.159
  16         -4.540
  17          1.808
  18         -1.883
  19          0.920
  20          9.119
partial autocorrelations:    0.234513  -0.211246    8.40821E-02   0.129090
  -0.214914  -0.145270  -0.133788   0.129215   -1.94044E-02    2.03490E-02
   -2.79954E-02   -3.33158E-02  -0.241118  -0.115527  -0.106227  -0.153519
   0.146170  -0.154551    3.04855E-02    4.85507E-02
Enter when ready to view time series
```

TIME SERIES PLOT

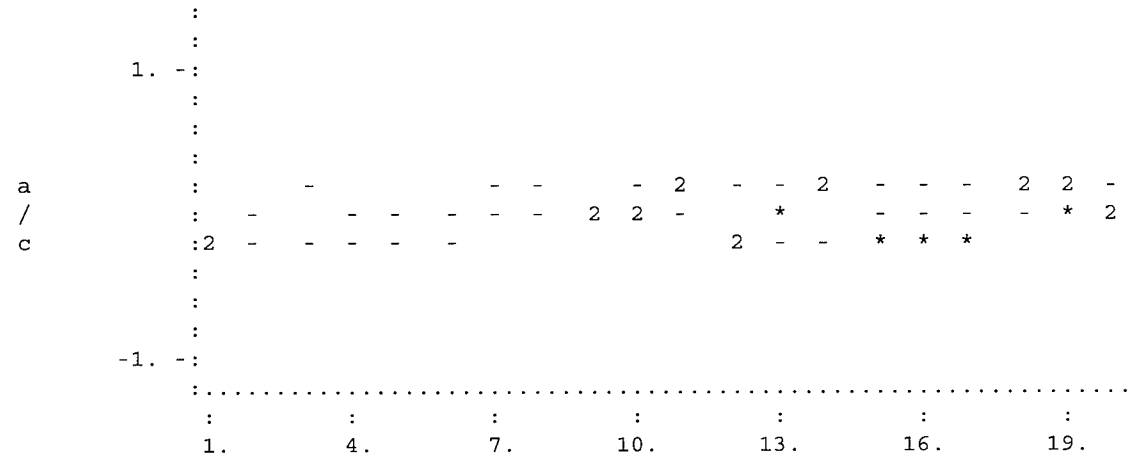


time period

Enter when ready to view autocorrelations

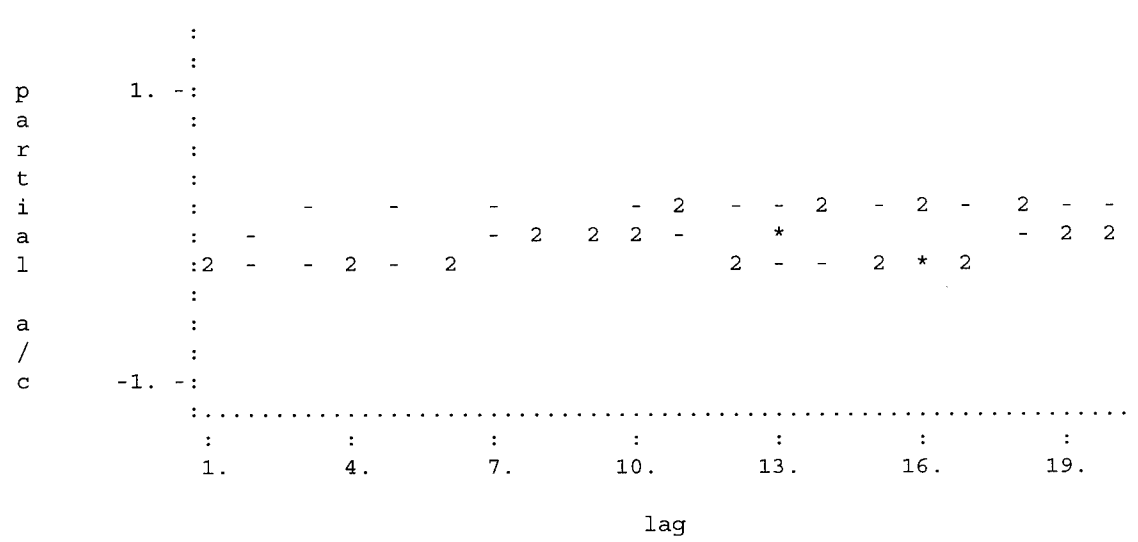
1

AUTOCORRELATION PLOT



lag

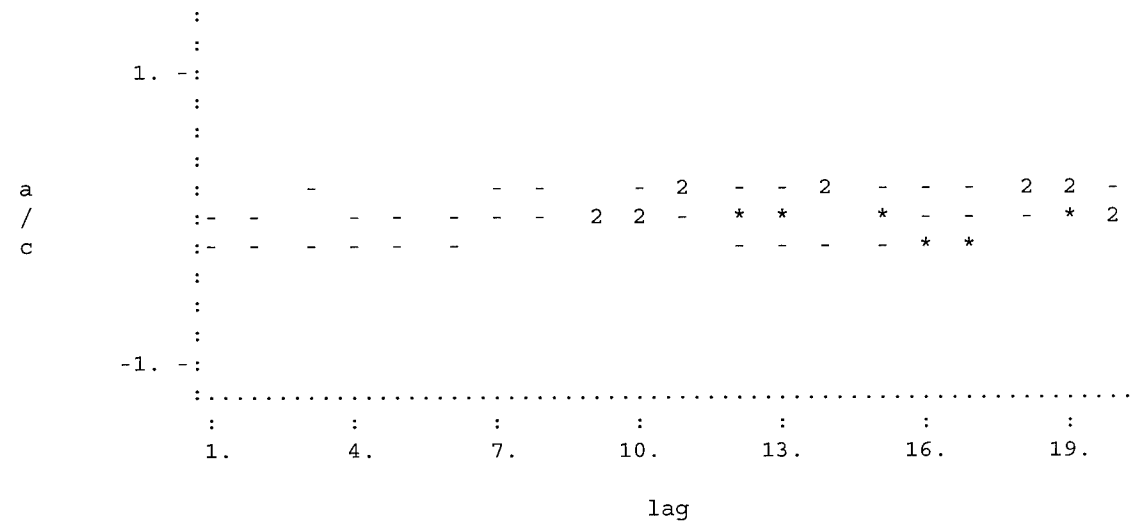
Enter when ready to view partial autocorrelations




```
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     12.6135
CONST =     11.6951
AVAR  =     42.1389

          PAR
         1           2
     0.2841     -0.2112
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.284053  -0.211246  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        149.436    190.160    136.818    171.586
113.897
    107.377    72.0915    98.2074
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     11.767
 Variance =     48.525

 Lag         ACV

   0         48.525
   1         12.830
   2         -4.653
```

```
   3          1.404
   4          9.860
   5         -3.722
   6        -11.053
   7         -6.780
   8          6.792
   9          2.838
  10         -2.568
  11         -1.458
  12          3.340
  13         -7.463
  14        -11.817
  15         -8.760
  16         -6.358
  17          0.175
  18         -3.069
  19         -0.199
  20          8.521
partial autocorrelations:    0.264400  -0.178256   0.117630   0.159747  -
0.197898
  -0.113580   -9.09876E-02   0.160784    7.38147E-03    1.77171E-02   -2.13626E-
02
   -4.26360E-02  -0.228606  -0.110848   -8.91308E-02  -0.140097   0.163511
  -0.119381    4.50877E-02    9.36938E-02
Enter when ready to view time series

1
                                  TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                              *
             :                  *              *               *
 u           :     *                  *     *      *         *  *
 g           : * *** * 2 * *   *     *          * * *    *  * 2 * *
 /           : ***  *** *****2* *2*2*****2** *******   2 2        **
 l       0. -:                              *         ** *  2**      **
             :.............................................................
              :         :         :         :         :         :         :
              0.       15.       30.       45.       60.       75.       90.

                                      time period

Enter when ready to view autocorrelations
```

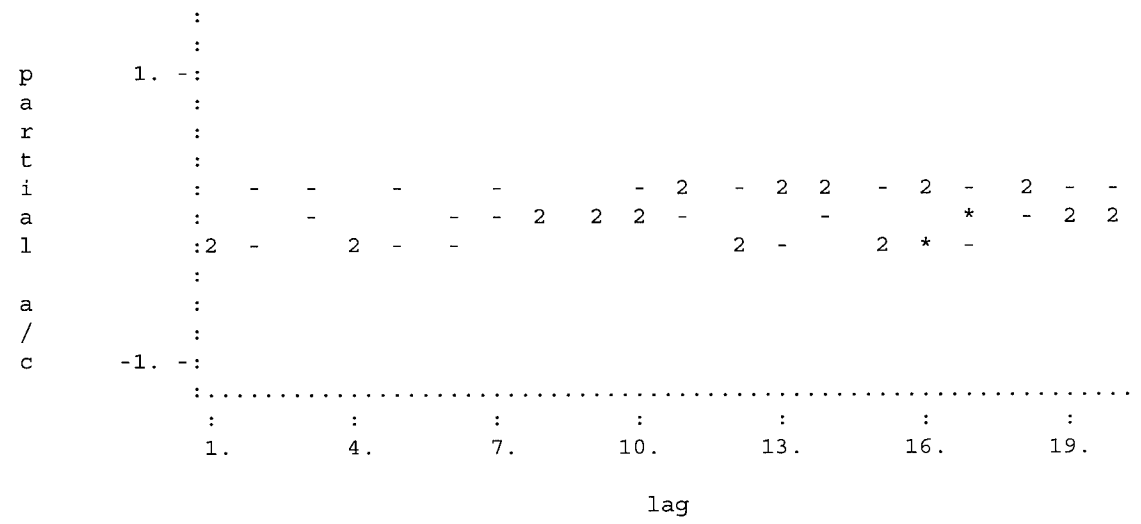
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

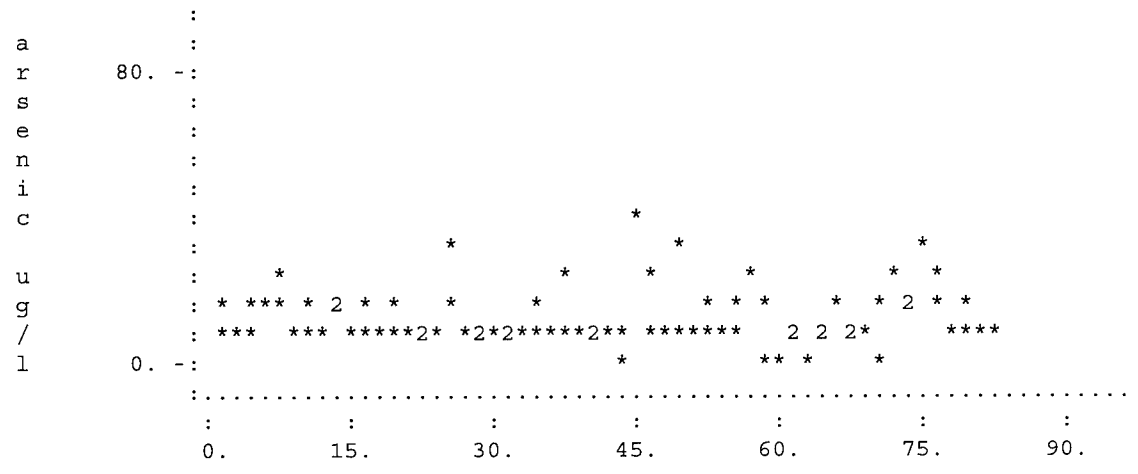
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 11.7671
CONST = 10.19883
AVAR = 43.6986

```
          PAR
          1           2
     0.3115     -0.1783
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.311531  -0.178256  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        136.788    181.022    125.330    164.909
104.7179
    98.9291    63.1647    94.1833
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     11.971
 Variance =     47.794

 Lag         ACV

   0         47.794
   1         12.134
   2         -5.421
   3          0.691
   4          9.122
   5         -4.284
   6        -11.590
   7         -7.306
   8          6.348
   9          2.544
  10         -2.640
  11         -1.506
  12          3.377
  13         -7.382
  14        -11.534
  15         -8.456
  16         -5.997
  17          0.501
  18         -2.846
  19          0.017
  20          8.619
partial autocorrelations:    0.253883  -0.190129   0.107569   0.149934  -
0.202863
  -0.123560   -1.03225E-00   0.151571    1.47544E-04    1.78610E-02   -2.37733E-
02
   -4.13353E-02  -0.233842  -0.115068   -9.79224E-02  -0.148838   0.10
Enter when ready to view time series
```

TIME SERIES PLOT

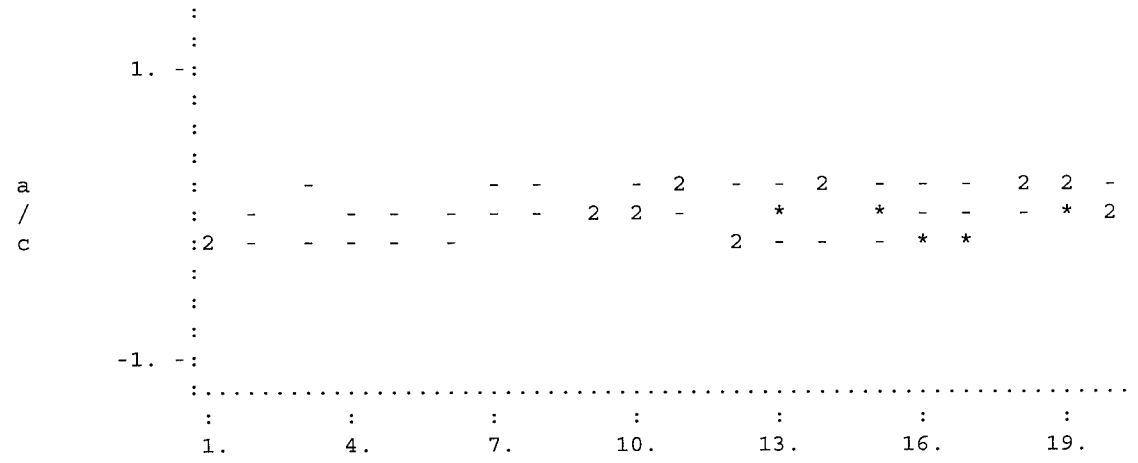


time period

Enter when ready to view autocorrelations

1

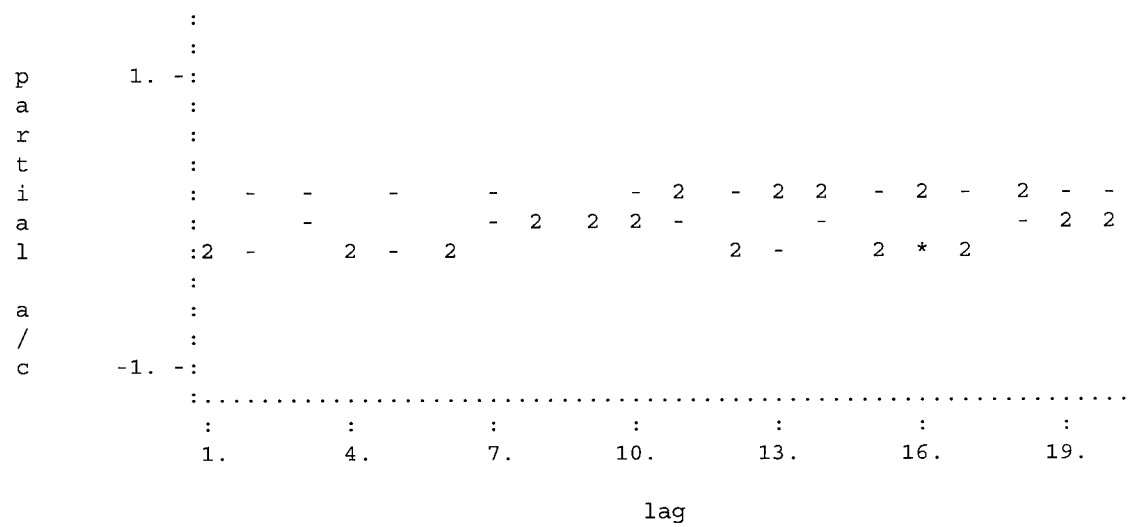
AUTOCORRELATION PLOT



lag

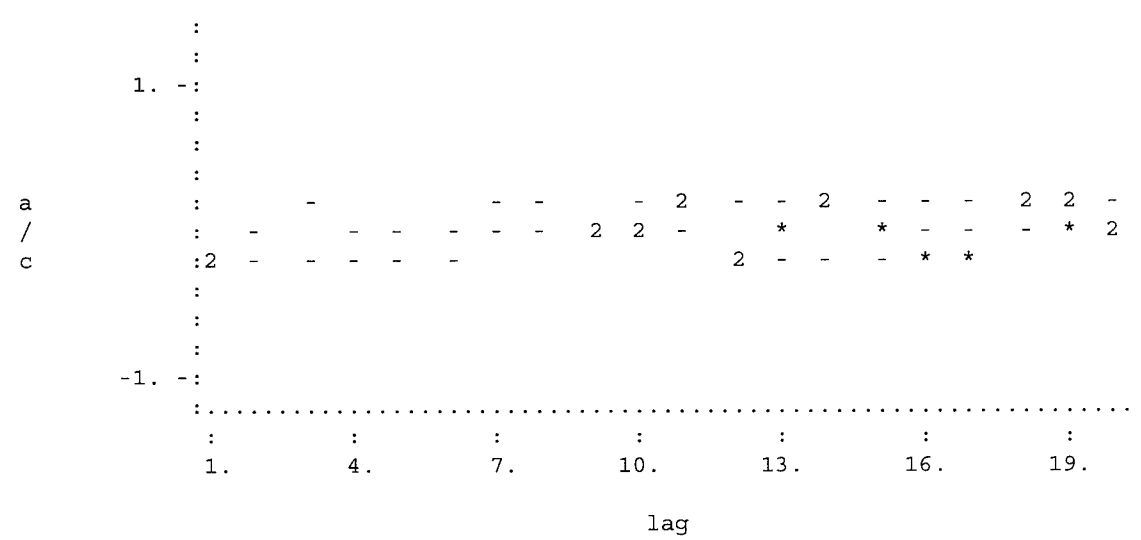
Enter when ready to view partial autocorrelations

PARTIAL A/C PLOT



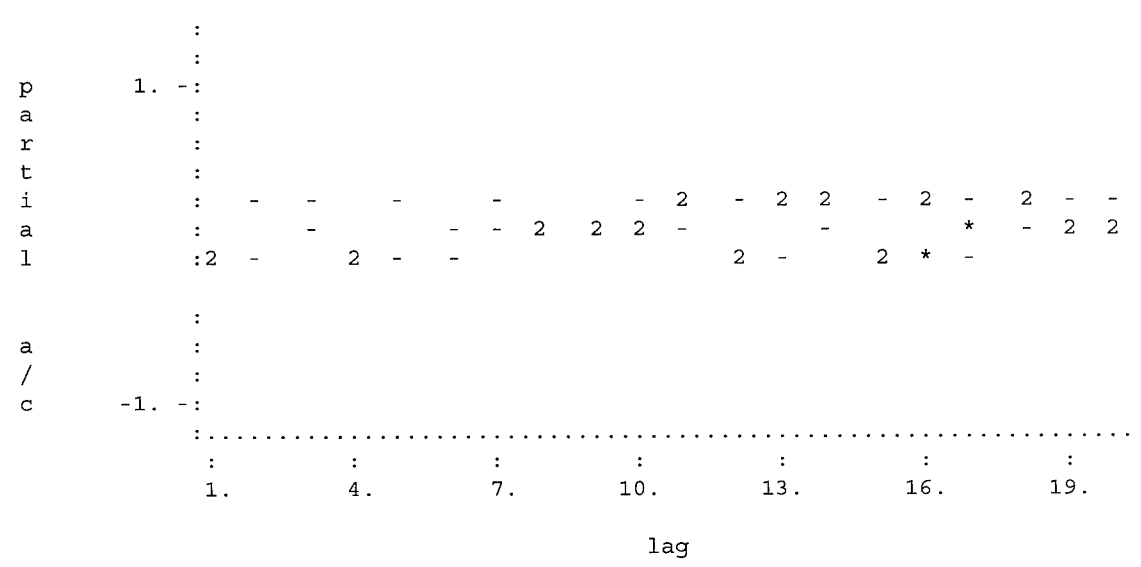

```
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on
Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN =     11.9706
CONST =     10.6296
AVAR  =     43.0967

          PAR
         1           2
    0.3022     -0.1901
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.302153  -0.190129  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        141.262    184.242    129.362    167.227
107.950
   101.9629    66.3594    95.6025
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     11.899
 Variance =     48.031

 Lag         ACV

   0         48.031
   1         12.362
   2         -5.169
   3          0.927
```

```
   4          9.366
   5         -4.098
   6        -11.414
   7         -7.131
   8          6.494
   9          2.640
  10         -2.623
  11         -1.496
  12          3.357
  13         -7.417
  14        -11.640
  15         -8.569
  16         -6.130
  17          0.381
  18         -2.929
  19         -0.063
  20          8.581
partial autocorrelations:    0.257370  -0.186195   0.111024   0.153269  -
0.201121
  -0.120127   -9.89378E-02   0.154818    2.74026E-03    1.78672E-02   -2.28957E-
02
   -4.17790E-02  -0.232043  -0.113633   -9.49321E-02  -0.145853   0.157147
  -0.128453    3.94975E-02    8.38106E-02
Enter when ready to view time series

1
                                TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                          *
             :                  *               *               *
             :     *                    *     *      *         *  *
 u           : * *** * 2 * *   *     *          * * *    *  * 2 * *
 g           : ***  *** *****2* *2*2*****2** *******   2 2 *      ****
 /           :                              *         ** *  ***
 l       0. -:.............................................................
             :         :         :         :         :         :         :
             0.       15.       30.       45.       60.       75.       90.

                                time period

Enter when ready to view autocorrelations
```



Enter when ready to view partial autocorrelations

1



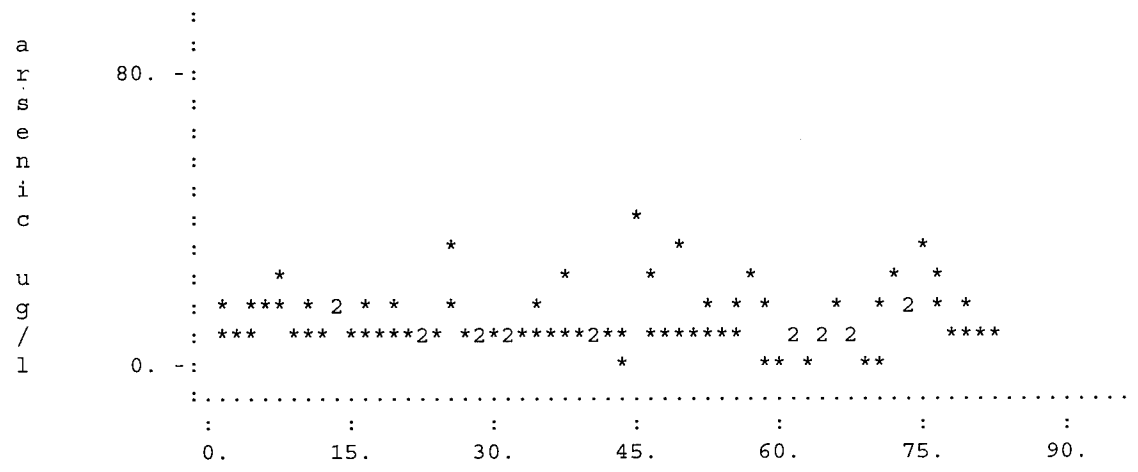
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 11.8989
CONST = 10.48178
AVAR = 43.2943

```
          PAR
          1           2
     0.3053     -0.1862
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.305291  -0.186195  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        139.780    183.174    128.023    166.455
106.878
   100.9621    65.3047    95.1319
Enter Y to perform another iteration
Y

Output from ACF/A2F

 Mean     =     11.923
 Variance =     47.950

 Lag         ACV

   0         47.950
   1         12.284
   2         -5.255
   3          0.847
   4          9.283
   5         -4.161
   6        -11.474
   7         -7.190
   8          6.444
   9          2.607
  10         -2.630
  11         -1.500
  12          3.362
  13         -7.407
  14        -11.606
  15         -8.532
  16         -6.087
  17          0.420
  18         -2.902
  19         -0.037
  20          8.593
partial autocorrelations:    0.256187  -0.187531   0.109866   0.152147  -
0.201700
  -0.121278   -1.00365E-01   0.153739    1.88388E-03    1.78702E-02   -2.31849E-
02
   -4.16316E-02  -0.232647  -0.114119   -9.59409E-02  -0.146860   0.156022
  -0.130072    3.84736E-02    8.20030E-02
Enter when ready to view time series
```

TIME SERIES PLOT

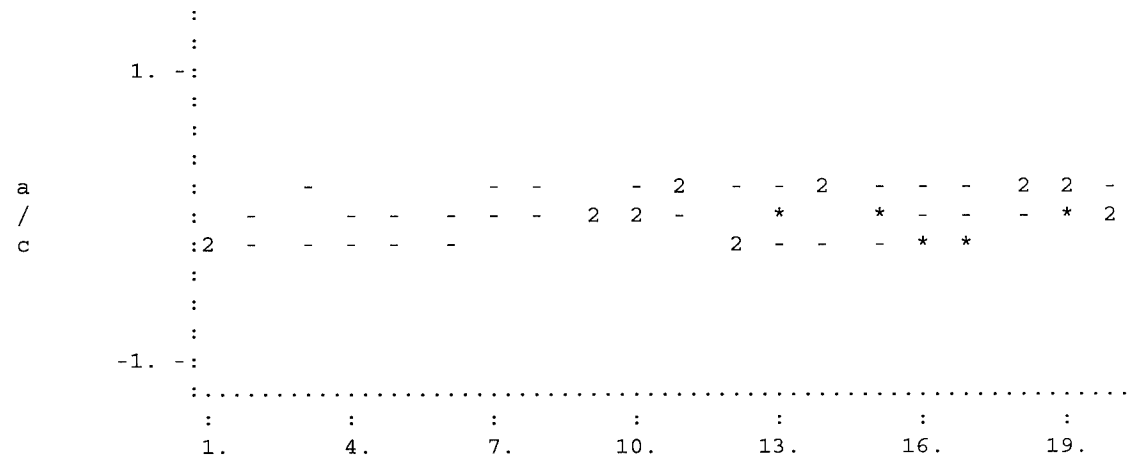


time period

Enter when ready to view autocorrelations

1

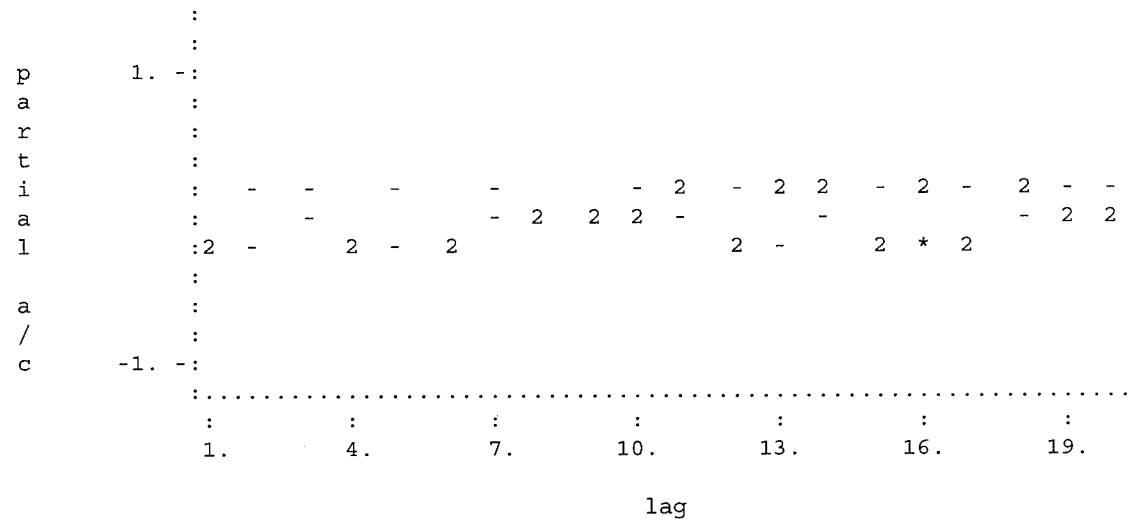
AUTOCORRELATION PLOT



lag

Enter when ready to view partial autocorrelations

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 11.9227
CONST = 10.53132
AVAR = 43.2270

PAR

1	2
0.3042	-0.1875

(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients: 0.304230 -0.187531 0. 0. 0. 0. 0.
Moving Average coefficients: 0. 0. 0. 0. 0. 0. 0.
Impulse response weights: 140.283 183.536 128.477 166.717
107.242
101.3023 65.6631 95.2915
Enter Y to perform another iteration
y

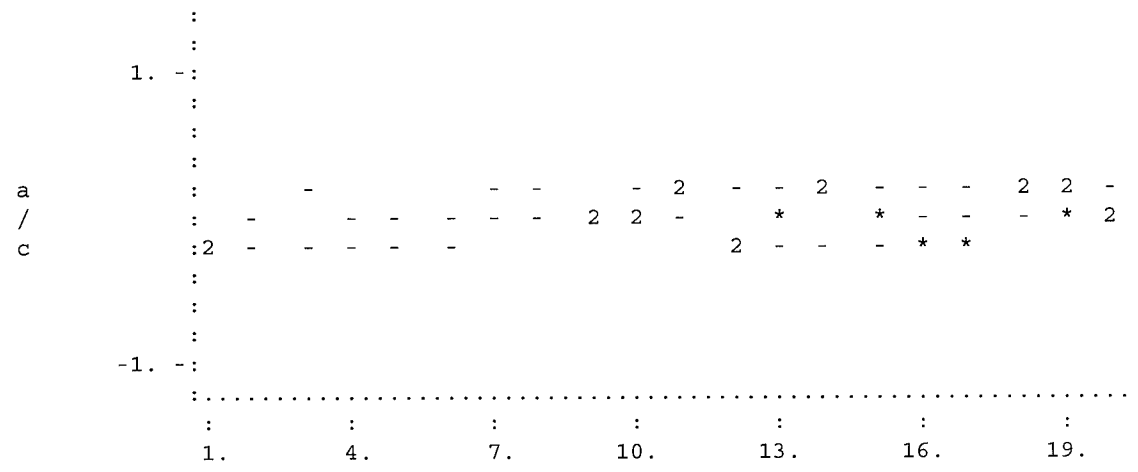
Output from ACF/A2F

Mean = 11.915
Variance = 47.977

Lag	ACV
0	47.977
1	12.310

```
   2          -5.226
   3           0.874
   4           9.311
   5          -4.140
   6         -11.454
   7          -7.170
   8           6.461
   9           2.618
  10          -2.628
  11          -1.499
  12           3.360
  13          -7.410
  14         -11.618
  15          -8.545
  16          -6.101
  17           0.407
  18          -2.911
  19          -0.046
  20           8.589
partial autocorrelations:    0.256586  -0.187081   0.110258   0.152526  -
0.201504
  -0.120888   -9.98813E-02   0.154105    2.17483E-03    1.78699E-02   -2.30866E-
02
   -4.16814E-02  -0.232443  -0.113955   -9.56003E-02  -0.146520   0.156403
  -0.129524    3.88210E-02    8.26166E-02
Enter when ready to view time series

1
                                  TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                          *
             :                  *               *                 *
             :     *                    *     *      *          *  *
 u           : * *** * 2 * *   *     *           * * *    *  * 2 * *
 g           : ***  *** *****2* *2*2*****2** *******   2 2 2      ****
 /           :                               *         ** *   **
 l       0. -:.............................................................
             :         :         :         :         :         :         :
             0.       15.       30.       45.       60.       75.       90.

                                      time period

Enter when ready to view autocorrelations
```

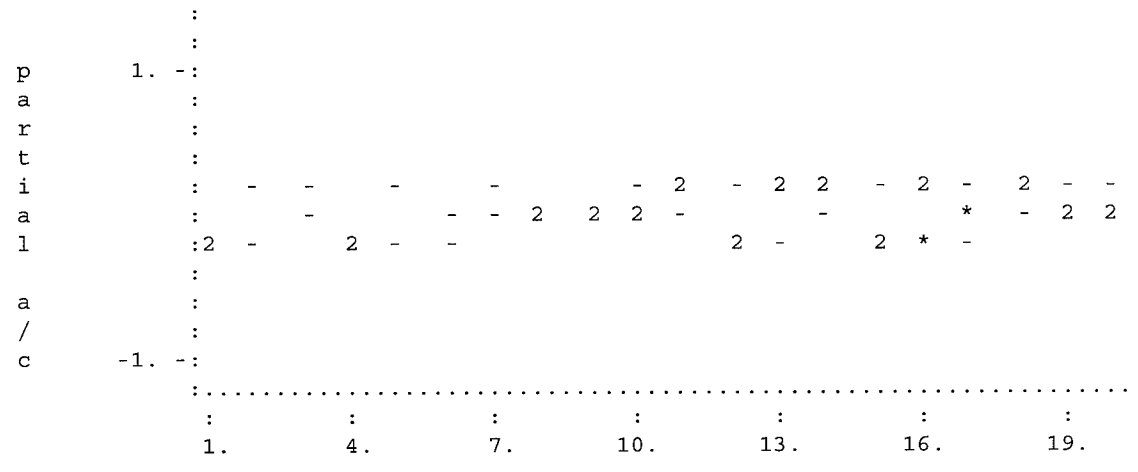
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on

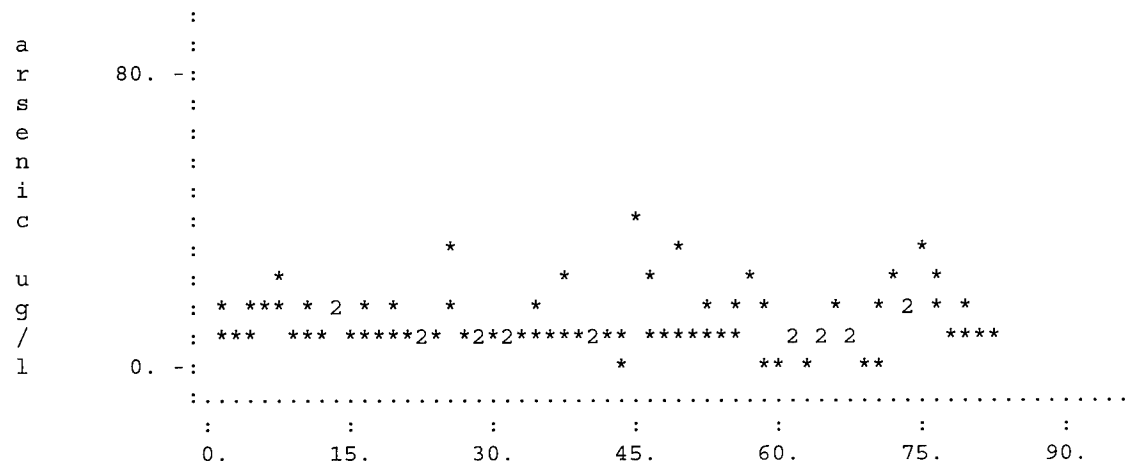
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 11.9146
CONST = 10.51456
AVAR = 43.2496

```
          PAR
          1           2
     0.3046     -0.1871
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.304588  -0.187081  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        140.114    183.414    128.324    166.629
107.119
   101.1878    65.5424    95.2377
Enter Y to perform another iteration
Y

Output from ACF/A2F

 Mean      =     11.917
 Variance =     47.968

 Lag         ACV

   0         47.968
   1         12.301
   2         -5.236
   3          0.865
   4          9.302
   5         -4.147
   6        -11.460
   7         -7.177
   8          6.455
   9          2.614
  10         -2.628
  11         -1.500
  12          3.361
  13         -7.409
  14        -11.614
  15         -8.540
  16         -6.096
  17          0.412
  18         -2.908
  19         -0.043
  20          8.590
partial autocorrelations:    0.256451  -0.187233   0.110126   0.152398  -
0.201570
  -0.121020   -1.00044E-01   0.153981    2.07687E-03    1.78700E-02   -2.31197E-
02
   -4.16646E-02  -0.232512  -0.114010   -9.57152E-02  -0.146634   0.156274
  -0.129708    3.87039E-02    8.24098E-02
Enter when ready to view time series
```

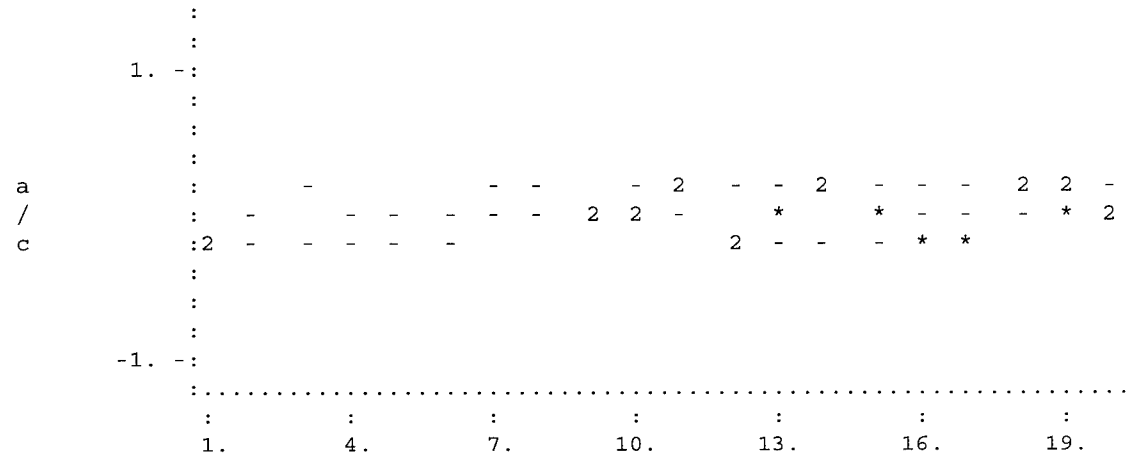
TIME SERIES PLOT



time period

Enter when ready to view autocorrelations
1

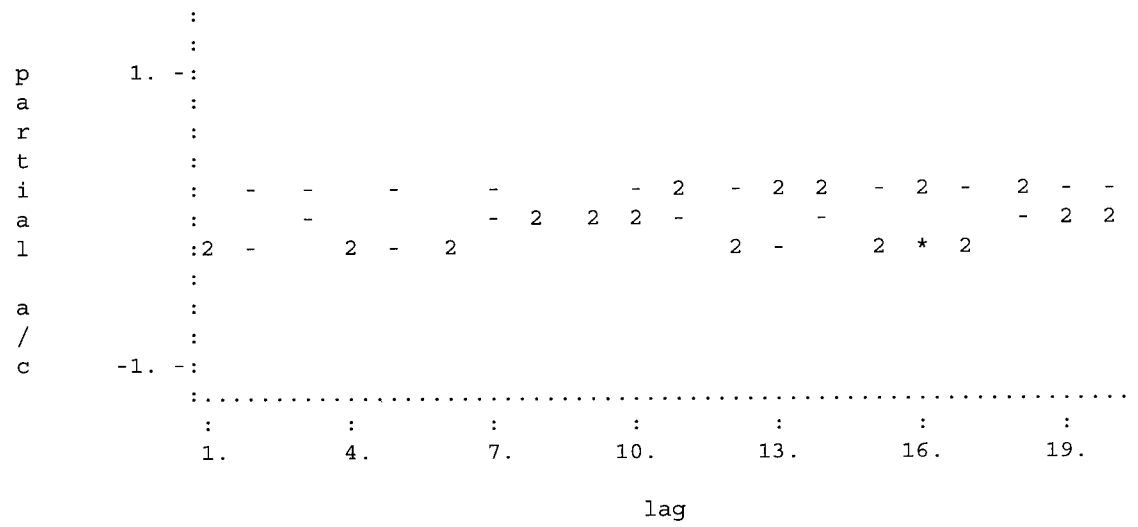
AUTOCORRELATION PLOT



lag

Enter when ready to view partial autocorrelations

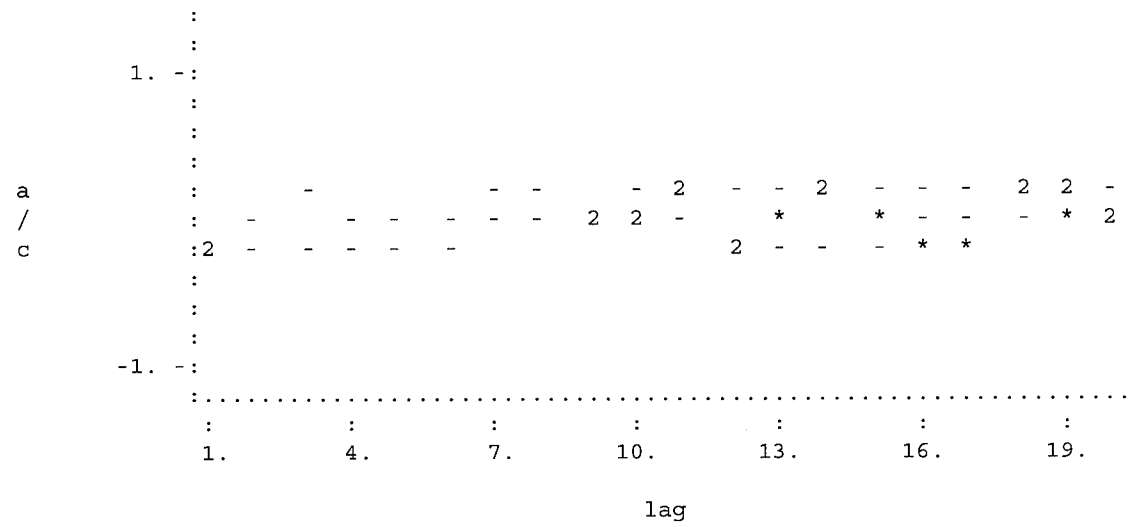
PARTIAL A/C PLOT




```
Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
0Enter value of q

Results from NSPE/N2PE

WMEAN =     11.9173
CONST =    10.52021
AVAR  =     43.2420

          PAR
         1           2
     0.3045     -0.1872
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.304467  -0.187233  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        140.171    183.455    128.376    166.659
107.161
   101.2265    65.5832    95.2559
Enter Y to perform another iteration
y

Output from ACF/A2F

 Mean     =     11.916
 Variance =     47.971

 Lag         ACV

   0         47.971
   1         12.304
   2         -5.232
```

```
   3          0.868
   4          9.305
   5         -4.145
   6        -11.458
   7         -7.174
   8          6.457
   9          2.615
  10         -2.628
  11         -1.499
  12          3.361
  13         -7.409
  14        -11.615
  15         -8.542
  16         -6.098
  17          0.410
  18         -2.909
  19         -0.044
  20          8.590
partial autocorrelations:    0.256496  -0.187181   0.110170   0.152441  -
0.201548
  -0.120975   -9.99893E-02   0.154023    2.11000E-03    1.78700E-02   -2.31085E-
02
   -4.16704E-02  -0.232488  -0.113991   -9.56764E-02  -0.146596   0.156318
  -0.129646    3.87435E-02    8.24797E-02
Enter when ready to view time series

1
                                TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :                              *
             :                   *             *               *
 u           :     *                   *     *      *         *  *
 g           : * *** * 2 * *   *     *          * * *    *  * 2 * *
 /           : ***  *** *****2* *2*2*****2** *******   2 2 2      ****
 l       0. -:                             *         ** *   **
             :.............................................................
              :         :         :         :         :         :         :
              0.       15.       30.       45.       60.       75.       90.

                                time period

Enter when ready to view autocorrelations
```

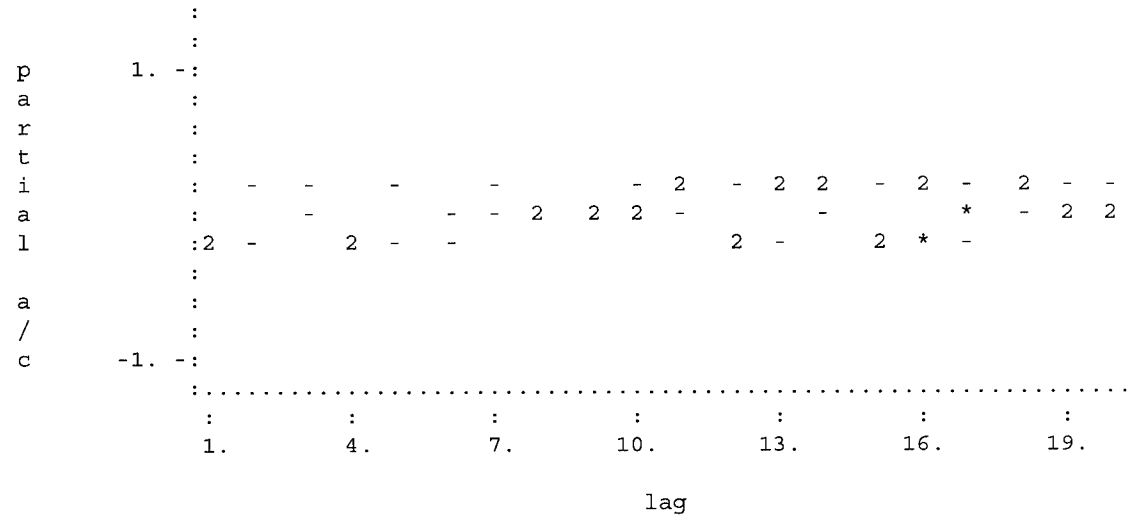
AUTOCORRELATION PLOT



Enter when ready to view partial autocorrelations

1

PARTIAL A/C PLOT



Enter R to repeat plotting or Enter to go on
Enter D to difference the time series or Enter to go on

Enter value of p
2
Enter value of q
0
Results from NSPE/N2PE

WMEAN = 11.9164
CONST = 10.51830
AVAR = 43.2445

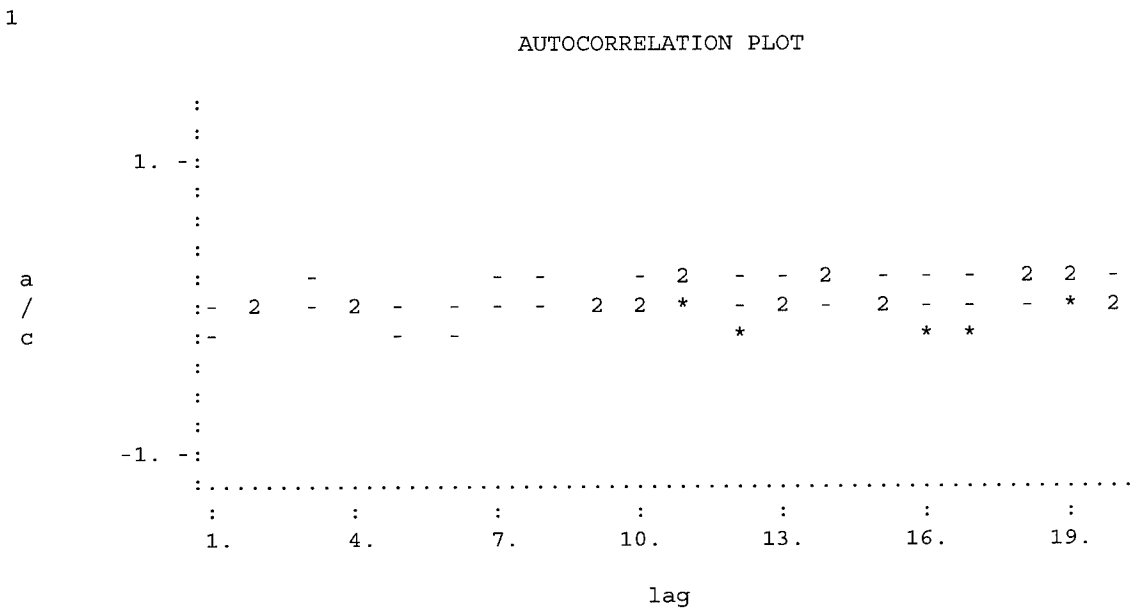
```
          PAR
          1           2
     0.3045     -0.1872
(PMA is not printed since NPMA = 0.)
Enter to estimate transfer function

Autoregressive coefficients:    0.304508  -0.187182  0.  0.  0.  0.  0.
Moving Average coefficients:   0.  0.  0.  0.  0.  0.  0.
Impulse response weights:        140.151    183.441    128.358    166.648
107.147
   101.2134    65.5694    95.2498
Enter Y to perform another iteration
n
```

Following are residuals and residual autocorrelations and partial autocorrelations.

```
Output from ACF/A2F

Mean      =     -8.5200
Variance =      7.3934

Lag          ACV

  0         7.39335
  1         0.45934
  2        -1.48230
  3        -0.25045
  4         1.48444
  5         0.10270
  6        -0.93155
  7        -0.81472
  8         0.76928
  9        -0.07581
 10        -0.01593
 11        -0.21004
 12         0.76169
 13        -0.62026
 14        -0.35070
 15        -0.21725
 16        -0.26915
 17         0.68552
 18        -0.51388
 19        -0.39054
 20         0.98681
partial autocorrelations:      6.21284E-02  -0.205142   -6.41192E-03   0.170448
   -2.29231E-02   -6.23641E-02   -9.50357E-02    5.85157E-02   -6.27093E-02
    6.14620E-02   -8.97615E-03    8.57209E-02  -0.118849   -9.43215E-03   -
4.48017E-02
   -8.93516E-02   0.153036  -0.117390    2.73601E-02    1.00112E-01
Enter when ready to view time series
                                      TIME SERIES PLOT

             :
 a           :
 r      80. -:
 s           :
 e           :
 n           :
 i           :
 c           :
             :
 u           :
 g           :
 /           :
 l       0. -: 2*2*2*2*2*2*2*2*2*2*2*2*2*2*2*2*2*2*2*2 2***2*2***2*2*2*2*2*2**
            :.....................................................................
             :         :         :         :         :         :         :
             0.       15.       30.       45.       60.       75.       90.

                                      time period
```

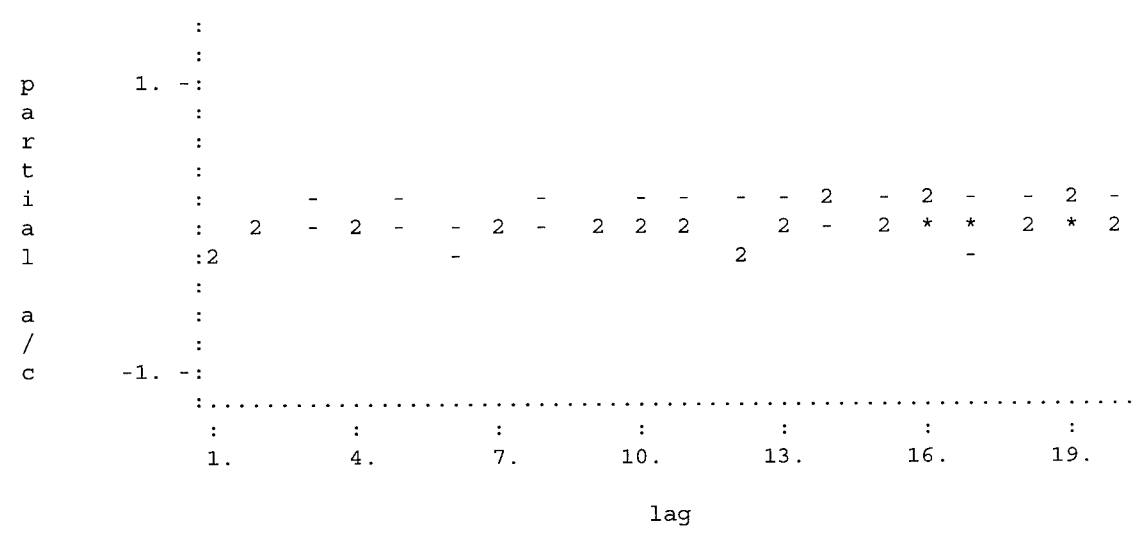
Enter when ready to view autocorrelations

1



Enter when ready to view partial autocorrelations

1



Bibliography

1. Amrine, John A. *Spectral and Spatial Pattern Recognition in Digital Imagery*. MS Thesis, AFIT/GSO/ENS/92D-01. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1992. (AD-A258910).

2. Aoki, Masanao. *State Space Modeling of Time Series*. Berlin: Springer-Verlag, 1987.

3. Benabdallah, Salah and Jeff R. Wright. "Multiple Subregion Allocation Models," *Journal of Urban Planning and Development, 118*: 24-40 (March 1992).

4. Black, William R. "Network Autocorrelation in Transport Network and Flow Systems," *Geographical Analysis, 24*: 207-222 (July 1992).

5. Box, George E. P. and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control* (Revised Edition). San Fransisco: Holden-Day, Inc., 1976.

6. Box, George E. P. and G. C. Tiao. "Intervention Analysis with Applications to Ecomomic and Environmental Problems," *Journal of of the American Statistical Association, 70*: 70-79 (March 1975).

7. Chan, Yupo. *Facility Location and Land Use: Multicriteria Decision Making Procedures*. Boyd and Fraser, draft of September 1994.

8. Environmental Protection Department, Fernald Environmental Restoration Management Corporation. *1993 Site Environmental Report*. Fernald Environmental Restoration Management Corporation, 1994.

9. Gallagher, Mark Austin. *Identification of the Initial Transient in Discrete-Event Simulation Output Using the Kalman Filter*. PhD Dissertation. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1992. (AD-A256611).

10. Getis, Arthur. "Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach," *Environment and Planning A, 23*: 1269-1277 (1991).

11. Getis, Arthur and J. K. Ord. "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis, 24*: 189-206 (July 1992).

12. Gonzalez, Rafeal C. and Richard C. Woods. *Digital Image Processing*. Reading, MA: Addison-Wesley Publishing Company, 1992.

13. Greene, Kelly A. *Causal Univariate Spatial-Temporal Autoregressive Moving Averages (STARMA) Modelling of Target Information to Generate Tasking of a World-wide Sensor System.* MS Thesis, AFIT/GOR/ENS/92M-12. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1992. (AD-A148109).

14. Harvey, Andrew C. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press, 1989.

15. Harvey, Andrew C. and R. G. Pierse. "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association, 79*: 125:131 (March 1984).

16. IMSL. "User's Manual: FORTRAN Subroutines for Statistical Analysis (STAT/Library)." Texas: IMSL, 1988.

17. Kashyap, Rangasami L. and A. Ramachandra Rao. *Dynamic Stochastic Models from Empirical Data.* New York: Academic Press, 1976.

18. Kaufman, Leonard and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis.* New York: John Wiley & Sons, Inc., 1990.

19. Klein, Ruben and S. James Press. "Adaptive Bayesian Classification of Spatial Data," *Journal of the American Statistical Association, 87*: 844-851 (September 1992).

20. Kohn, Robert and Craig F. Ansley. "Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data," *Journal the American Statistical Association, 81*: 751-761 (September 1986).

21. Lang, Laura. "Water's New World," *Civil Engineering, 62*: 48-50 (June 1992).

22. Lin, D. S. and others. "Comparisons of Remotely Sensed and Model-Simulated Soil Moisture over a Heterogeneous Watershed," *Remote Sensing of Environment, 48*: 159-171 (May 1994).

23. Makridakis, Spyros and others. *Forecasting: Methods and Applications* (Second Edition). New York: John Wiley & Sons, Inc., 1983.

24. McGee, Donald W., Jr. *The Application of Statistical Kriging to Improve Satellite Imagery Resolution.* MS Thesis, AFIT/GSO/ENS/91D-13. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1991. (AD-A244014).

25. McLachlan, Geoffrey J. *Discriminant Analysis and Statistical Pattern Recognition.* New York: John Wiley & Sons, Inc., 1992.

26. Pankratz, Alan. *Forecasting with Dynamic Regression Models.* New York: John Wiley & Sons, Inc., 1991.

27. Pfeifer, Phillip E. and Stuart Jay Deutsch. "A Three Stage Iterative Procedure for Space-Time Modeling," *Technometrics, 22*: 35-47 (February 1980).

28. Porter, Charles H. *Comparison of Batch Means and Independent Replications Technique to Applications of the Kalman Filter for Simulation Output Analysis.* MS Thesis, AFIT/GOR/ENS/90M-15. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1990. (AD-A220656).

29. Reed, Thomas Gerald. *Binary Programming Models of Spatial Pattern Recognition: Applications in Remote Sensing Image Analysis.* MS Thesis, AFIT/GSO/ENS/91D-15. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1991. (AD-A243797).

30. Rochon, James. "ARMA Covariance Structures eith Time Heteroscedasticity for Repeated Measures Experiments," *Journal of the American Statistical Association, 87*: 777-784 (September 1992).

31. Stoffer, David S. and Kent D. Wall. "Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter," *Journal of the American Statistical Association, 86*: 1024-1033 (December 1991).

32. Tiao, G. C. and G. E. P. Box. "Modeling Multiple Time Series with Applications," *Journal of the American Statistical Association, 76*: 802-816 (December 1981).

33. Tobler, W. "An Alternative Formulation for Spatial-Interaction Modeling," *Environment and Planning A, 15*: 693-703 (1992).

34. U.S. Department of Energy, Fernald Office. *Groundwater Modeling Evaluation Report and Improvement Plan*, Fairfield, Ohio: Parsons, 1993.

35. U.S. Department of Energy, Fernald Office. *South Plume Groundwater Recovery System Pump Test Work Plan*, Fairfield, Ohio: Parsons, 1993.

36. Viessman, Warren, Jr. and others. *Introduction to Hydrology* (Second Edition). New York: Harper & Row, Publishers, Inc., 1977.

37. Walsh, T. A. and T. E. Burk. "Calibration of Satellite Classifications of Land Area," *Remote Sensing of Environment, 46*: 281-290 (December 1993).

Vita

Captain Samuel A. Wright was born on 20 March 1967 in Columbus, Ohio. He graduated form Whitehall-Yearling High School in Whitehall, Ohio in 1985 and attended the United States Air Force Academy, graduating with a Bachelor of Science in Astronautical Engineering in May 1989. Upon graduation, he received a regular commission into the United States Air Force. In July 1989, he married the former Victoria Lynn Huffman, also of Whitehall, Ohio. Captain Wright's first tour of duty was with Air Force Space Command. There, he served for four years with the 1st Space Operations Squadron. Captain Wright held duties as a Planner/Analyst for launch and early orbit, as well as on-orbit, operations of Global Positioning System satellites. He then became a staff evaluator in the Standardization/Evaluation Branch, also upgrading to the Mission Commander position. After serving as Chief, GPS Standardization/ Evaluation Branch, Captain Wright went on to head the GPS Engineering Branch, responsible for satellite contingency operations, as well as satellite disposals. Captain Wright attended Squadron Officers' School before entering the School of Engineering, Air Force Institute of Technology in August 1993. Samantha Lynn was born to Sam and Vicki on July 19, 1990, followed by Sammy on November 15, 1991. Both were born at the Air Force Academy Hospital.

Permanent address: 4144 Wright Park
Whitehall, Ohio 43213

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1995	Master's Thesis

4. TITLE AND SUBTITLE
Spatial Time-Series:
Pollution Pattern Recognition under Irregular Interventions

5. FUNDING NUMBERS

6. AUTHOR(S)
Samuel A. Wright, Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GOR/ENS/95M-19

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Fernald Environmental Restoration Management Corporation
P.O. Box 398704
Cincinnati, OH 45239-8704

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Fernald Environmental Restoration Management Corporation (FERMCO) has noted the introduction of arsenic contamination to groundwater around the area of the groundwater recovery system, which captures uranium contaminated groundwater. The introduction of arsenic occurs during high levels of pumping and is particularly sensitive to the western two of the five pumps. Auto-Regressive Moving Average (ARMA) and Spatial-Temporal ARMA (STARMA) empirical analyses are used to model the level of arsenic contamination found through time. The intervention of varied levels of pumping is modeled with a transfer function using analytic techniques to create a causal intervention transfer function input series to give physical meaning to the impulse response weights found. Spatial weights employed in the STARMA modeling are also created using analytic, causal methods. Results suggest a physical interpretation of the relationship between a) a particular level of pumping and b) the amount of arsenic to be found at the site in the temporal case, while including c) the effect of the pumping on a site of interest's neighbors in the spatial-temporal case. Models presented may be employed in the forecasting of arsenic levels at monitoring well sites due to a given change in the operation level of the groundwater recovery system.

14. SUBJECT TERMS
ARMA, STARMA, transfer function

15. NUMBER OF PAGES
187

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL