AN AUTOREGRESSION MODEL FOR A PAIRED WATERSHED COMPARISON

D. Meek

D. Dinnes

D. Jaynes

C. Cambardella

T. Colvin

See next page for additional authors

Author Information

D. Meek, D. Dinnes, D. Jaynes, C. Cambardella, T. Colvin, J. Hatfield, and D. Karlen

AN AUTOREGRESSION MODEL FOR A PAIRED WATERSHED COMPARISON

D. Meek, D. Dinnes, D. Jaynes, C. Cambardella, T. Colvin, J. Hatfield, and D. Karlen
USDA-ARS-MWA National Soil Tilth Laboratory
2150 Pammel Dr., Ames, IA 50011-4420 USA

Abstract

Analysis of water quality data from a paired watershed design is needed to determine if a best fertilizer management practice reduces a specific water quality variable compared to a conventional fertilizer management practice. This study examines an existing recommended method of analysis for paired watershed designs, simple analysis of covariance (ANCOVA) on time aggregated data, then offers two autoregression analyses (AR) as alternatives. The first approach models the sequence of paired differences and estimates its 95% confidence band. The second approach develops individual watershed AR models then examines the joint 95% confidence interval about the predicted difference. A reliability analysis on the water quality data reveals that the data for the controlled watershed, i.e., the covariate, has a sizable measurement error, a factor that is not considered in the usual ANCOVA model. The AR methods avoid the measurement error and other inherent problems with the published recommended method. Graphically both AR analyses are similar and reveal three distinct trend phases: a period of continued similarity; a period of transition; and a period of sustained change. The model for the sequence of paired differences is the easier one of the two AR methods to use and interpret because its trend model of splined linear segments readily defines each response phase. Hence, we recommend it over the given alternatives. It offers water resources researchers an effective and readily adoptable analysis option.

1. Introduction

Agriculture's nitrogen (N) fertilizer use and its impact on water quality are at the forefront of environmental concerns for Americans from the farm to the city and, consequently, to the U.S. Congress and state and federal regulatory agencies. The Walnut Creek Nitrogen Initiative addresses the issue of nonpoint source nitrate contamination of surface waters from subsurface drained fields at the watershed scale. Current conventional corn nitrogen fertilizer management in the Midwest is fall application of anhydrous ammonia. This practice was adopted because labor is often more available, soils are less likely to be compacted, fertilizer supply is generally abundant, and economic returns are generally favorable. Environmentally, however, fall fertilization has many risks: the N fertilizer can leach, it can move off-site via runoff, and it can denitrify. All of these processes can occur during periods when the crops are not present as well as during the growing season. Previous research in a subsurface drained agricultural watershed near Ames, IA (Walnut Creek) found that nitrate-N concentrations were at or above the EPA drinking water maximum contamination level (MCL) of 10 ppm in over 60% of weekly observations. This excessive level was observed during a baseline monitoring period that lasted from April 1990 through the end of 1996 (Jaynes et al., 1999). The field research objective of the Walnut Creek Nitrogen Initiative is to quantify changes of nitrate content in subsurface flow as a result of implementing intensive N fertilizer management at the watershed scale. Of course, the pertinent statistical issue was to select an appropriate method of analysis.

In the following section, we discuss problems with a recommended published ANCOVA method. Hence, the purpose of this work is to recommend a sound alternative method of analysis on a water quality variable from a paired watershed study. The method should demonstrate any significant changes that occur in the paired water quality observations over time and, preferably, the analysis should use the time interval of the observations.

2. Methods

2.1 The Watershed Study

Three subbasins in the Walnut Creek Watershed area near Ames, IA., were chosen to conduct the N management study because of the availability of long-term baseline monitoring data (Fig. 1). The soils, geography, and farming practices are typical of the Des Moines Lobe. The basins are all artificially drained with a network of subsurface tiles feeding into a central line that has a single outlet for the entire subbasin. The operative measurement of interest is weekly nitrate-nitrogen concentration, [NO_3-N] , in the subsurface drainage, manually obtained via grab-samples. A paired watershed approach is used to compare the water quality of subbasin 220 under a N fertilizer best management practice, BMP, to that from subbasins 210 and 230, which are both under conventional N management practices. Only comparisons of subbasin 220 to subbasin 210 are presented here since multiple baseline analyses determined that subbasin 220 was more hydrologically similar to subbasin 210 than 230. These analyses considered N concentration, subsurface flow, and estimated subsurface load along with some geographic properties of the subbasins. Further details are available in the conference proceedings by Dinnes et al. (2000).

Starting in the Fall of 1996 the Late-Spring Soil Nitrate Test (LSNT) method, a BMP advocated by Iowa State University (Blackmer et al., 1997) to manage N fertilizer for corn production, was implemented on the 220 subbasin. With this system, N fertilizer is applied in split applications. The first application is a nominal rate ($\approx$50 kg/ha) shortly before, or at, planting and the second soon after late-spring soil sampling with the rate determined by soil test results. Plot-scale studies using the LSNT or pre-sidedress nitrate test (PSNT) (Magdoff et al., 1984; Fox et al., 1989; Magdoff et al., 1990) practices have generally shown a positive impact on measured or potential nitrate leaching. While designed to provide an optimum N fertilizer rate at an appropriate time for maximum N use efficiency by the crop, the impact on water quality and risk to the farmer for adopting this approach has been poorly quantified at the whole-field scale and, until now, had not been attempted at the watershed scale. When the subsurface drains were flowing between 1/1/97 and 8/27/99, weekly grab-samples were collected at each basin's subsurface drainage outlet and chemically analyzed for [NO_3-N]. The procedures and methods are described in Hatfield et al. (1999).

2.2 Analysis

Analysis of covariance methods have been recommended for examining paired watershed data by Grabow et al. (1998; 1999a, and 1999b) who demonstrated them by using detailed examples with real data. The common existence of serial correlation in such monitoring data or in their models' residual sequences has also been recognized and either time aggregation over hydrological events or autroregressive ANCOVA models are then the suggested appropriate methods. There are also some problems using this approach. Elementary texts (e.g., Steel and Torrie, 1980, p. 406) list the necessary assumptions for proper use of ANCOVA; the very first assumption is that the covariate is fixed, measured without error, and independent of treatments. Based on the availability of additional data, the measurement error can be estimated in our study. The analyses reported in this paper were run on the weekly grab-sample data, as is normal in monitoring work. Automated equipment at the drainage outlet was used to collect additional regular interval and storm event samples. To assess the measurement error, data from the automated system sample for the nearest neighbors in time can be paired to the next grab samples collected afterward. Table 1 summarizes reliability coefficients and reliability ratios (see e.g., pp. 1-9 in Fuller, 1987) for each set of basin observations. While there was a negligible bias in all comparisons and results were better than expected, we think all reported r and κ values should exceed 0.95 to be able to ignore measurement error considerations. Using the observations from the conventionally managed subbasin violates the premise of independence from treatments. The ANCOVA methodology will, therefore, not be formally compared to the AR methods.

Analysis with two AR model methods are compared instead. Time and past residual values (a.k.a. *lags*) are the only variables. The simpler approach develops a model based on the paired difference series constructed from the two individual series. Models consist of a trend and an AR component. Trend is modeled with splined or grafted polynomials. The independent time sequence is divided in two or more intervals then polynomial segments fit to each interval are developed with a continuity constraint at the join or transition points (a.k.a. *knots*). Join points are estimated both graphically and analytically with nonlinear models. Gallant and Fuller (1973) or Rivlin (1969) are among the many available references on this methodology. Residual autocorrelation can then be used to select lag terms. Alternatively, nonlinear routines can be used to simultaneously estimate both model components. An intrinsically linear trend component model can be done in many autoregression routines if estimating the knots and testing hypothesis about them is not of interest. In this work all approaches were employed.. In addition a 95% confidence interval is estimated over the period of comparison. While many formal tests are possible, for practical purposes, we can say that significant change is achieved when the confidence interval stops including the zero line. Our graphical presentation uses ideas suggested in Tufte (1983). The more involved second approach is similar, but models for each individual series are developed instead and the response difference is constructed from the difference in predictions with a joint 95% confidence band around the prediction differences. Here each trend is modeled with splined ordinary and rational polynomial segments with knots fixed at half year intervals. Modeling was done with SAS™ (company and any trade names are given for the benefit of the reader and imply no endorsement of the products by the USDA-ARS). PROC NLIN, PROC MODEL, PROC AUTOREG, and PROC GPLOT were used (SAS, 1990 and 1993).

3. Results and Discussion

A point of interest in these analyses is that the farmers were not able to fall apply N fertilizer to either subbasin in 1996 because of inclement weather. This factor probably resulted in a lack of difference during 1997. The Fig. 2 montage shows the individual time series and the results of the analysis on the difference series. Eq. [1] states the trend model with predetermined fixed

$$\Delta[NO_3\text{-}N] = \left\{ \begin{array}{rl} 0, & 0 < \text{week} \leq 59.8 \\ -0.101(\text{week}-59.8), & 59.8 < \text{week} \leq 98.9 \\ -4.3, & 98.9 < \text{week} \end{array} \right\} \text{ppm} \qquad [1]$$

knots. A lag-1 term with an AR coefficient of -0.322 is used for the model results shown in the bottom panel of the Fig. 2 montage. While the addition of other lag terms (3, 8 and 9) is weakly indicated, their inclusion only slightly reduces prediction error and the resulting confidence band width. Hence for simplicity, we exclude them. Ideally knots, especially the first one, could be fixed on the basis of N-application records. In this model, the first knot does not correspond to the week just after the start of fall N-application because not all farmers in the subbasins fertilized their fields in the same week and the transition also depends on weather and other environmental factors. Likewise, the second knot need not occur at a specific time. Hence we recommend that the knots be fit, preferably simultaneously, with a nonlinear modeling procedure. If not done this way, an intrinsically linear autoregression procedure like PROC AUTOREG can be used iteratively by varying the knot locations to get the best overall model. If hypothesis tests on the knots are not of interest, the latter method may suffice.

Fig. 3 shows the results of the second AR analysis. Graphically, the results for the trend of the predicted difference are very similar to the first analysis, but both of the latter method's individual AR models had comparatively much more complicated trend models each with three lag terms (including terms up to lag 6). The confidence band for the latter method is, on average, only 16% wider then that shown in Fig. 2. Hence, the simpler approach is recommended because it gives similar results, it is computationally easier, and it provides a simple and elegant interpretation corresponding to the trend phases. There are three distinct phases corresponding to the spline segments in Eq. [1]: a continued period of similarity, which makes sense given the lack of Fall, 1996 fertilizer application in the control subbasin; a transition period; and a sustained period of significant change.

Summary

ANCOVA methods are not recommended for several reasons. Conceptually the management practice on the control watershed is a treatment. Moreover, aside from other possible assumption violations and possible numerical problems, there are two major problems. (1) In this study the response phases were not certain and thus became part of the modeling job. Conventionally, ANCOVA requires the phases to be known. (2) Finally, in our study, the measurement error is appreciable and should not be ignored.

The AR approach avoids dealing with measurement error. Thus, a pure AR method, the simpler one, is recommended. Graphical presentation of the model provides good insight on the response difference over time; hence it is more likely to be used by nonstatistical researchers. Specifically, this analysis helps demonstrate that intense N management strategies can play an effective role in reducing ag-related nonpoint source [NO_3-N] inputs to water resources.

Acknowledgments

This work was supported by the USDA-ARS-MWA National Soil Tilth Laboratory (NSTL), Dr. J.L. Hatfield, Director. The dedication of the NSTL Nitrogen Management Team's Field Crew (J. Cook, F. Goodman, R. Hartwig, K. Heikens, and L. Pellack) and NSTL Analytical Laboratory Staff are greatly appreciated. The participation of the following farmer-cooperators was crucial: B. Bassett, Mr. and Mrs. B. Finch, S. Kaltenheuser, B. Muir, J. Oakland, R. Tripp, and T. Uthe.

References

Blackmer, A.M., Voss, R.D., and Mallarino,A.P., 1997. *Nitrogen fertilizer recommendations for corn in Iowa. PM-1714.* ISU Coop. Ext., Ames, IA.

Dinnes, D.L., Jaynes, D.B., Meek, D.W., Cambardella, C.A., Colvin, T.S., Hatfield, J.L., and Karlen, D.L.. 2000. Intensive N fertilizer management effects on water quality at the watershed scale. p. 95-99. In *Proceedings of The Watershed Approach to Improving Water Quality: Fact or Fantasy? Conference*, 28-30 Mar. 2000, La Crosse, WI. Soil Water Conserv. Soc., Ankeny, IA.

Fox, R.H., Roth, G.W., Iverson, K.V., and Peikielek, W.P., 1989. Comparison of soil and tissue nitrate tests for predicting soil nitrogen availability of corn. *Agron. J.*, 81, 971-974.

Fuller, W.A., 1987. *Measurement error models*. Wiley, New York, NY.

Gallant, A.R. and Fuller, W.A.., 1973. Fitting segmented polynomial models whose joint points have to be estimated. *Am Statis*., 68(341), 144-147.

Grabow, G.L., Spooner, J, Lombardo, L.A., and Line, D.E., 1998. Detecting water quality changes before and after BMP Implementation: Use of a spreadsheet for statistical analysis. *NWQEP Notes, The NCSU Water Quality Group Newsletter.* Nov. 1998, 92, 1-9. NCSU Coop. Ext., Raleigh, NC.

Grabow, G.L., Spooner, J, Lombardo, L.A., and Line, D.E., 1999a. Detecting water quality changes before and after BMP Implementation: Use of SAS for statistical analysis. *NWQEP Notes, The NCSU Water Quality Group Newsletter.* Jan. 1999, 93, 1-11. NCSU Coop. Ext., Raleigh, NC.

Grabow, G.L., Spooner, J, Lombardo, L.A., and Line, D.E., 1999b. Detecting water quality changes before and after BMP Implementation: Use of SAS for trend analysis. *NWQEP Notes, The NCSU Water Quality Group Newsletter.* Jul. 1999, 95, 1-8. NCSU Coop. Ext., Raleigh, NC.

Hatfield, J.L., Jaynes, D.B., Burkart, M.R., Cambardella, C.A., Moorman, T.B., Prueger, J.H., and Smith, M.A., 1999. Water quality in Walnut Creek watershed: Setting and farming practices. *J. Environ. Qual.*, 28, 11-24.

Jaynes, D.B., Hatfield, J.L., and Meek. D.W., 1999. Water quality in Walnut Creek watershed: Herbicides and nitrate in surface waters. *J. Environ. Qual.*, 28, 45-59.

Magdoff, F.R., Jokela, W.E., Fox, R.H., and Griffin, G.F., 1990. A soil test for nitrogen availability in the northeastern United States. *Commun. Soil Sci. Plant Anal.*, 21, 1103-1115.

Magdoff, F.R., Ross, D., and Amadon, J., 1984. A soil test for nitrogen availability to corn. *Soil Sci. Soc. Am. J.*, 48, 1301-1304.

Rivlin, T.J., 1969. Chapter 4: Polynomial and spline interpolation, pp. 87-119. *An introduction to the approximation of functions*. Blaisdell Pub. Co., Waltham, MA.

Tufte, E.R. 1983. *The visual display of quantitative information*. Graphics Press, Cheshire, CT.

Steel, R.G.D., and Torrie, J.H., 1980. *Principles and procedures of statistics, a biometrical approach, 2nd Ed.* McGraw-Hill Pub. Co., New York, NY.

SAS Inst. Inc., 1990. *SAS/STAT user's guide, version 6, 4th Ed.* SAS Inst., Cary, NC.

SAS Inst. Inc., 1993. *SAS/ETS user's guide, version 6, 2nd Ed.* SAS Inst., Cary, NC.

Table 1. [NO_3-N] reliability statistics.

Site[1]	n	r	κ
210	208	0.87	0.86
220	205	0.82	0.81
230	232	0.89	0.89

[1] Site: Watershed/Subsurface drain outlet ID.
n: No. observations.
r: Reliability coefficient (via the Pearson correlation coefficient).
κ: Reliability ratio.

Figure 1. Subbasins in the Walnut Creek Watershed near Ames, IA. The area in subbasin 210 is 491 ha; 220 is 366 ha; and 230 is 863 ha. The subsurface drainage system is denoted as *Tiles* in the map key.

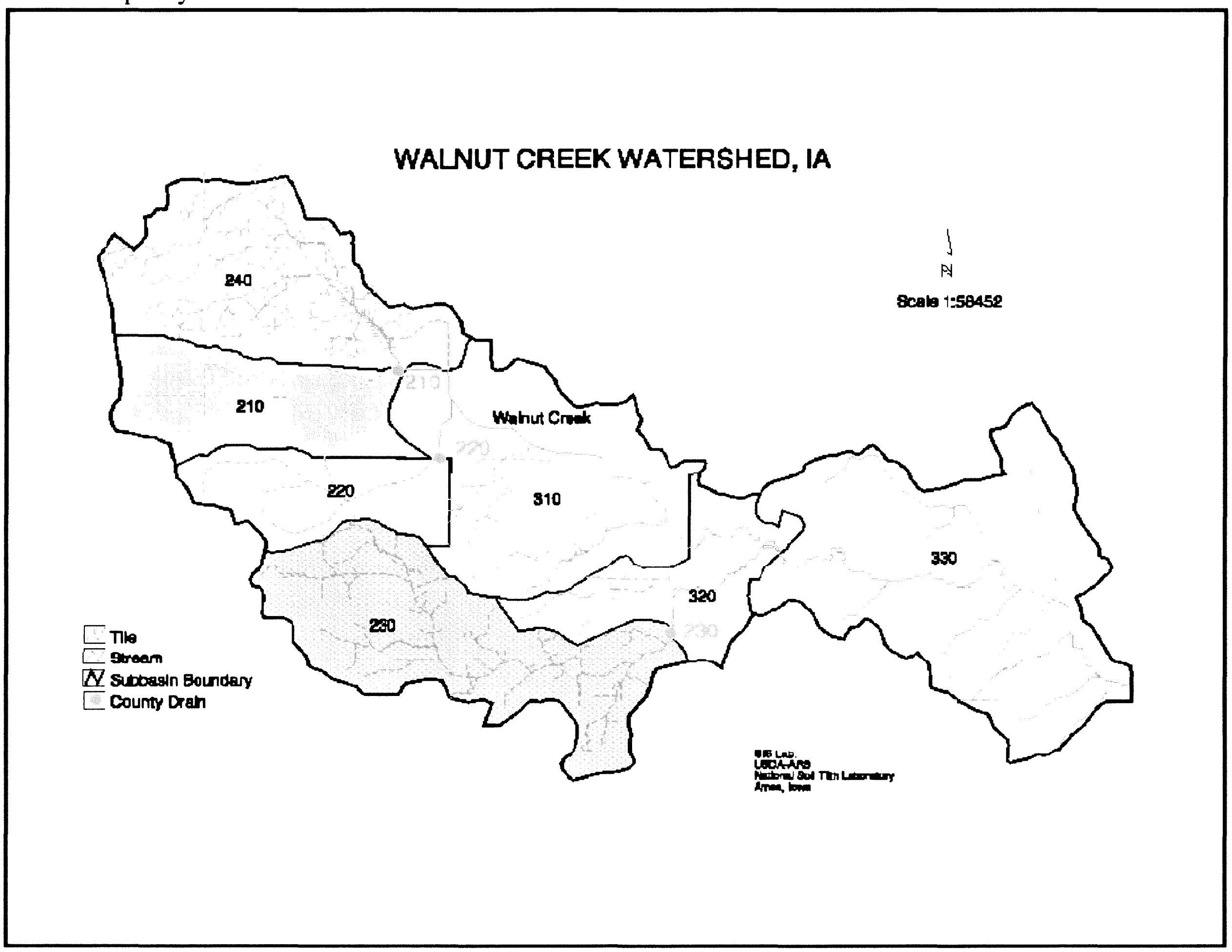

Figure 2. The top two panels show the individual time series for each subbasin. The bottom panel shows the difference series with Eq. [1] predictions (dashed line) and the 95% confidence band (shaded area) for the entire period of observation. Data are represented by the circles.

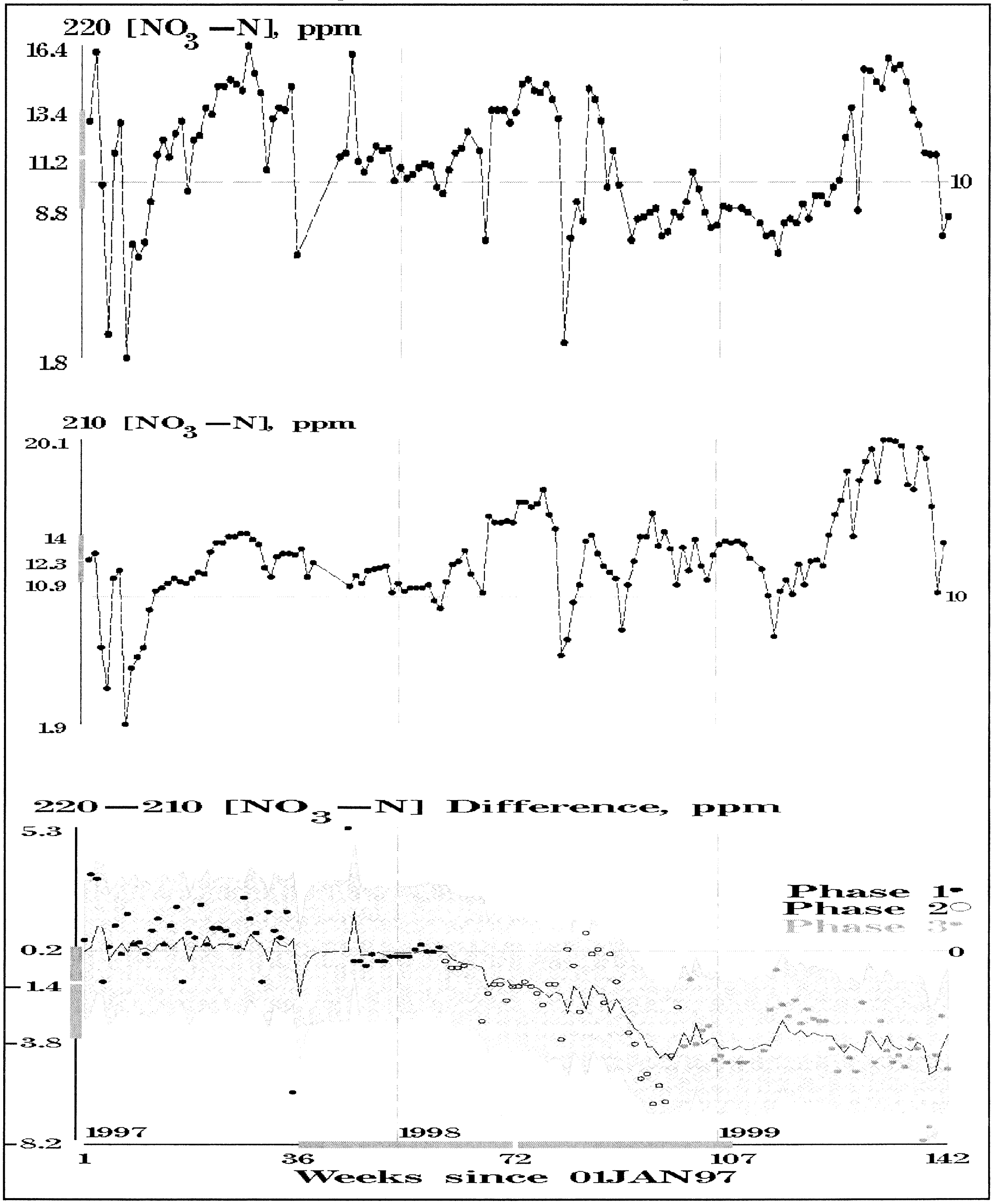

Figure 3. Difference of both individual model predictions (dashed line) with the jointly estimated 95% confidence band (shaded area) for the entire period of observation. Data are represented by the circles.