TEMPORAL AUTOCORRELATION IN MODELING SOIL POTENTIALLY MINERALIZABLE NITROGEN

by

FABIO SARTORI

(Under the direction of Daniel B. Hall)

ABSTRACT

In this simulation study, the precision and accuracy of two different methods for modeling net nitrogen mineralization data have been compared. Method 1 represents one traditional approach found in the soil science literature that does not account for the temporal autocorrelation present in the data. Method 2, based on the recent advancements in methods and software for nonlinear mixed effects models that account for this autocorrelation, has proven to be more precise than the traditional approach under a variety of different scenarios.

INDEX WORDS: Potentially mineralizable nitrogen, Temporal autocorrelation, Nonlinear Mixed effects models, Parameter estimation, Precision, and Accuracy.

TEMPORAL AUTOCORRELATION IN MODELING SOIL POTENTIALLY MINERALIZABLE NITROGEN

by

FABIO SARTORI

B.S., The University of Padova, Italy 1997

A Thesis Submitted to the Graduate Faculty of The University of Georgia, Department of Statistics in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2004

TEMPORAL AUTOCORRELATION IN MODELING SOIL POTENTIALLY MINERALIZABLE NITROGEN

by

FABIO SARTORI

Major Professor: Daniel B. Hall

Committee: Mary Meyers
Jaxk Reeves

Electronic version approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

DEDICATION

Alla memoria dei miei genitori, Ida e Carlo.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

The estimation of the soil potential capability of releasing available nitrogen is important for the efficient use of fertilizer and prevention of ground water pollution in agricultural and intensively managed forest ecosystems. The main aim of this work is to demonstrate that it is relevant to consider temporal autocorrelation when fitting net nitrogen mineralization data. Considering such autocorrelation allows one to obtain better estimates and inferences (e.g. unbiased standard errors) on soil potentially mineralizable nitrogen (N_0) and decomposition rate constant (k). N_0 and k parameters are commonly estimated by considering either the cumulative or the incremental amount of available nitrogen obtained after laboratory incubations repeated over time (Stanford and Smith, 1972). In most cases, two sub-samples (pseudo-replicates) are extracted for each soil type, at regular intervals under controlled conditions of temperature and moisture (Stanford and Smith, 1972; Cabrera and Kissel, 1988; Maimone et al., 1991). The average value of available nitrogen (AN) is computed each time. Incremental AN or cumulative AN data (sum of all increments obtained at different time intervals) for each sample is then modeled as a function of time without considering the temporal autocorrelation, assuming there is independence between the repeated measures on the same soil sample. Most authors have focused their attention on using models that account for measurement errors intrinsic to this laboratory technique. In their review of N_0 modeling approaches, Cabrera and Kissel (1994) suggest that one of the most precise methods for fitting net nitrogen mineralization data (NNM) data is the use of nonlinear regression procedures to fit incremental models, as proposed by Ellert

and Bettany (1988). These authors have found that using incremental rather than cumulative data has the advantage of reducing the interdependence between each observation without accumulating possible measurement errors.

Nevertheless, to my knowledge, no author has ever fitted a model to an entire data set, composed of all series of repeated measures obtained from a particular experiment, accounting for the within-run temporal autocorrelation. The need to consider autocorrelation when using longitudinal data has been discussed in other studies on soil respiration (Hess and Schmidt, 1995) and microbial ecology studies (Robinson, 1985). Several models have been compared to better represent the biological process underlying NNM as discussed by Cabrera and Kissel (1994). Most studies have focused on finding the best model for individual runs, and computing an average of the obtained parameter estimates for a certain set of soil (Stanford and Smith, 1972; Molina et al., 1980; Talpaz et al. 1981; Deans et al. 1986; Cabrera and Kissel, 1988). Analogously to what Hall and Bailey describe in their study about modeling and predicting forest growth (Hall and Bailey, 2001), this approach is a somewhat crude method of obtaining an estimate of the parameter for a certain soil type population. A better approach is available through the use of nonlinear mixed effects NNM curves. Such models include both fixed regression parameters (fixed effects) that describe the shape of the typical NNM curve over the entire population and which account for differences across common soil grouping factors, experimental treatments and other covariate effects, and random regression coefficients (random effects) that individualize the curve to the different levels of grouping or clustereing present among the experimental units. The major advantage of fitting a model to the entire data set is that it will allow better statistical inference, resulting in unbiased parameter standard errors and valid tests and confidence intervals.

This advanced statistical technique is often applied when using repeated measures or longitudinal data (Lindstrom and Bates, 1990). The increasing popularity during recent years of nonlinear mixed effects models (NLMMs) is due to their flexibility in modeling the within-group correlation often present in grouped data, their handling of balanced and unbalanced data in a unified framework, and the availability of reliable and efficient software for fitting them (Lindstrom and Bates, 1990; Pinheiro and Bates, 2000).

In this simulation study, data are generated from a known model to compare the performance of nonlinear mixed effects models that account for temporal auto-correlation relative to models that assume there is no temporal dependence among observations. Also as an application, I fit the Stanford and Smith data (SSD) set to show that there is indeed correlation when considering data sets grouped by run.

In this work I perform a simulation study based on data generated from a known model, rather than just comparing models on a single set of data, and I consider the advantage of a more complete evaluation under a variety of possible different scenarios. Knowing exactly what the underlying chosen model generating the data is, or in other words, what the population parameters we are trying to estimate are, affords us the opportunity to better judge the accuracy and precision of a specific model for parameter estimation.

CHAPTER 2

MATERIALS AND METHODS

In order to judge whether or not it is relevant to consider temporal autocorrelation when fitting NNM data, a simulation study was structured following a factorial design. An S-PLUS program was developed and adapted assuming different initial conditions for each simulated scenario. The program performs a series of iterations to generate and fit data sets under different conditions. The main structure includes one basic loop (1) to generate at the i_{th} iteration a single subset of data, composed of 6 observations (n_i=6), defined as a run. Loop 1 is repeated 40 times (K=40) to create a data set composed of 40 independent runs. A secondary loop (2) repeats this process (1) 500 times. During each simulation, data are generated and then fitted with two different methods: 1) without considering any correlation in the data and assuming independence among the model errors using weighted nonlinear least squares, and 2) accounting for the temporal autocorrelation using a nonlinear mixed effects model.

Initial Conditions and Data Generation

To compare methods 1 and 2 with respect to different data sets, two model functions with the same design, three (N_0,k) parameter combinations, and four possible different variance-covariance structures were combined in a 2×3×2×2 factorial design (Table 1). The choice of the different parameters was based on the data obtained from figure 1 and table 2 of the SSD (Stanford and Smith, 1972).

Expectation function f_1 is the single exponential function for the i_{th} run of the form:

(1) $E(N_{ij})=N_0\cdot(1-e^{k\cdot tij})$, i=1, . . ., K, j=1, . . .,n_i,

to fit cumulative data (Cabrera et al., 1994). Expectation function f_2 is a modified version of:

(2) $E(\Delta_{ij})=N_0\cdot e^{-k\cdot tij}\cdot(e^{k\cdot dij}-1)$, i=1, . . ., K, j=1, . . .,n_i,

to fit incremental data, proposed by Ellert and Bettany (1988). N_{ij} (mg Kg^{-1}) is the cumulative nitrogen mineralized in the i_{th} run at time t_{ij}. N_0 is the population potentially mineralizable N parameter, and k the population rate constant of mineralization (Fig. 1). The quantity Δ_{ij} is the simulated incremental amount of N released during the interval d_{ij}, preceding time t_{ij}. Each of the i_{th} runs is composed of n_i extraction dates corresponding to time t_{ij}'s as considered by Cabrera and Kissel (1988). For all possible cases it is always assumed K=40 and n_i=10, with exactly the same extraction dates: 2, 4, 8, 12, 16, 20, 24, 28, 32, and 36 weeks since the incubation start. In vector form $\mathbf{t}=(2, 4, 8, 12, 16, 20, 24, 28, 32, 36)^T$. Therefore the corresponding d_{ij} lags are respectively: 2, 2, 4, 4, 4, 4, 4, 4, 4, and 4 weeks or $\mathbf{d}=(2, 2, 4, 4, 4, 4, 4, 4, 4, 4)^T$. In this study two correlation coefficients were considered, ρ=0.9 and ρ=0, to represent respectively the scenarios with and without error autocorrelation. The 0.9 value was selected in order to obtain the desired autocorrelation structure given the number of observation of each run.

In the homoskedastic scenarios the variance was assumed to be 1.44 for both the cumulative and incremental function. An exponential variance function of time was used for the more complex heteroskedastic scenarios, similar to the general form described by Pinheiro and Bates (2000): (3) $Var(\varepsilon_{ij}) = \sigma_p^2\cdot\exp(\upsilon_{ij}) \equiv \sigma_{ij}^2$ i=1, . . ., K_j; j=1, . . .,n_i, where σ_p^2 is the population variance assumed to be 1.44, as for the homoskedastic scenarios.

I used $\upsilon_{ij}= -(1-t_{ij}/36)$ and $\upsilon_{ij}= +(1-t_{ij}/36)$, respectively, for the cumulative and incremental scenarios to create appropriate transformations in the "heterogeneous AR(1)" type variance covariance-structure (Littel et al., 1996) to represent the biological process of NNM.

The elements of the variance-covariance matrix Σ are given by:

(4) $\sigma_{ij}\cdot\sigma_{ik}\,\rho^{|j-k|}$, j=1, . .,10; k=1, . . .,10,

where j is the row and k column index for the variance-covariance matrix Σ. In the cumulative case σ_{ij}, j=1, . . ., n_i exponentially increases from its minimum at $t_{i1}=2$, $\sigma_{i1}=1.2\cdot\exp(1-(2/36)) \approx 0.47$, reaches its maximum value at $t_{i10}=36$, $\sigma_{i10}=1.2$. The corresponding $\boldsymbol{\sigma}_i$ vector containing the different values in time is:

(5) $\boldsymbol{\sigma}_i =(0.46, 0.49, 0.55, 0.62, 0.69, 0.77, 0.86, 0.96, 1.07, 1.2)^T$, i , j=1, . . .10

In the incremental case instead σ_{ij}, j=1, . . .,n_i, exponentially decreases from its maximum at $t_{i1}=2$, $\sigma_i =1.2\cdot\exp((4/36)-1) \approx 3.08$, reaches its minimum value at $t_{i10}=36$, $\sigma_{i10}=1.2$. The corresponding $\boldsymbol{\sigma}_i$ vector containing the different values in time is:

(6) $\boldsymbol{\sigma}_i =(3.08, 2.92, 2.61, 2.34, 2.091, 1.87, 1.67, 1.49, 1.34, 1.2)^T$, i , j= 1, . . .10. In general the cumulative data is obtained by adding the single incremental observations, I_{ij}'s.

Being $I_{ij}=f_2+\varepsilon_{ij}$, the corresponding cumulative model is given by: (7) $\sum_1^{ni} I_{ij} = \sum_1^{ni} \mathbf{f}_2 + \sum_1^{ni} \varepsilon_{ij}$.

The corresponding variance can be computed as

(8) $$\mathrm{Var}\left(\sum_1^{ni} I_{ij}\right) = \mathrm{Var}\left(\sum_1^{ni} \mathbf{f}_2 + \sum_1^{ni} \varepsilon_{ij}\right) = \mathrm{Var}\left(\sum_1^{ni} \varepsilon_{ij}\right) = \sigma^2 \sum_1^{ni} \nu_{ij}$$

The i_{th} and j_{th} elements of Σ are arranged in a manner, that determines an increasing and decreasing variance with time, respectively, for the cumulative and incremental model (Appendix A).

As a simplified example, considering six extraction dates temporally spaced as $\mathbf{t}=(4, 8, 12, 16, 22, 30)^T$ and being $\mathbf{R}$ the within-run correlation matrix, the i_{th} and j_{th} elements of Σ are obtained as follows:

$\Sigma=\text{diag}(\boldsymbol{\sigma}_i)\ \mathbf{R}\ \text{diag}(\boldsymbol{\sigma}_i)=$

$$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6 \end{pmatrix} \begin{pmatrix} \rho^0 & \rho^4 & \rho^8 & \rho^{12} & \rho^{18} & \rho^{26} \\ \rho^4 & \rho^0 & \rho^4 & \rho^8 & \rho^{12} & \rho^{18} \\ \rho^8 & \rho^4 & \rho^0 & \rho^8 & \rho^{12} & \rho^{18} \\ \rho^{12} & \rho^8 & \rho^4 & \rho^0 & \rho^4 & \rho^8 \\ \rho^{18} & \rho^{12} & \rho^8 & \rho^4 & \rho^0 & \rho^4 \\ \rho^{26} & \rho^{18} & \rho^{12} & \rho^8 & \rho^4 & \rho^0 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1\cdot\sigma_1\ \rho^0\cdot & \sigma_1\cdot\sigma_2\ \rho^4 & \sigma_1\cdot\sigma_3\ \rho^8 & \sigma_1\cdot\sigma_4\ \rho^{12} & \sigma_1\cdot\sigma_5\ \rho^{18} & \sigma_1\cdot\sigma_6\ \rho^{26} \\ \sigma_2\cdot\sigma_1\ \rho^4 & \sigma_2\cdot\sigma_2\ \rho^0 & \sigma_2\cdot\sigma_3\ \rho^4 & \sigma_2\cdot\sigma_4\ \rho^8 & \sigma_2\cdot\sigma_5\ \rho^{12} & \sigma_2\cdot\sigma_6\ \rho^{18} \\ \sigma_3\cdot\sigma_1\ \rho^8\cdot & \sigma_3\cdot\sigma_2\ \rho^4 & \sigma_3\cdot\sigma_3\ \rho^0\cdot & \sigma_3\cdot\sigma_4\ \rho^4 & \sigma_3\cdot\sigma_5\ \rho^8 & \sigma_3\cdot\sigma_6\ \rho^{12} \\ \sigma_4\cdot\sigma_1\ \rho^{12} & \sigma_4\cdot\sigma_2\ \rho^8 & \sigma_4\cdot\sigma_3\ \rho^4 & \sigma_4\cdot\sigma_4\ \rho^0 & \sigma_4\cdot\sigma_5\ \rho^4 & \sigma_4\cdot\sigma_6\ \rho^8 \\ \sigma_5\cdot\sigma_1\ \rho^{18} & \sigma_5\cdot\sigma_2\ \rho^{12} & \sigma_5\cdot\sigma_3\ \rho^8 & \sigma_5\cdot\sigma_4\ \rho^4 & \sigma_5\cdot\sigma_5\ \rho^0 & \sigma_5\cdot\sigma_6\ \rho^4 \\ \sigma_6\cdot\sigma_1\ \rho^{26} & \sigma_6\cdot\sigma_2\ \rho^{18} & \sigma_6\cdot\sigma_3\ \rho^{12} & \sigma_6\cdot\sigma_4\ \rho^8 & \sigma_6\cdot\sigma_5\ \rho^4 & \sigma_6\cdot\sigma_6\ \rho^0 \end{pmatrix}$$

In general an autoregressive first order AR(1) process can be written as (Wei, 1989) (9): $Z_t=\phi_1 Z_{t-1}+a_t$, where a_t is a series of uncorrelated random variables from a well-defined distribution with mean μ_a, usually assumed to be zero, and variance $\text{Var}(a_t)=\sigma_a^2$. The quantity, ϕ_1, is the partial autocorrelation coefficient between the two series Z_t and Z_{t-1}. In an AR1(1) process the value of the series Z_t is completely determined by the knowledge of Z_{t-1}, the one unit apart process. Given this formulation, it follows that the autocorrelation function (ACF) becomes $\rho_k=\phi_1\rho_{k-1}=\phi^k_1$, $k\geq 1$ (10), where k is the number of units apart from Z_t. The corresponding partial autocorrelation function (PACF) is (11): $\phi_{kk}=\rho_1=\phi_1$, when k=1. $\phi_{kk}=0$, $k\geq 2$. If we represent these functions graphically, we can say that for a stationary AR(1) process the ACF tails off and the PACF cuts off after lag 1.

As a result of these initial assumptions four different variance-covariance matrices for each function type (i.e. cumulative and incremental) were used in this study. For example, in scenarios 1 through 3, when $\rho=0$, or in other words when errors are independent and identically distributed (Homoskedasticity and Independence), $\sum$ corresponds to the 10×10 matrix $\sum_{C1}=1.44\ \mathbf{I}_{10}$, with 1.44 in all along diagonal and 0 in all off-diagonal positions for the cumulative case**.** Similarly for the corresponding incremental scenarios (1-3) $\sum_{I1}= 1.44\ \mathbf{I}_{10}$. The specific residuals pattern deriving from the use of the two model functions and four variance-covariance type are presented in figures A-1 through A-8 of Appendix A.

The choice of the (N_0, k) parameter combination was based on information obtained from the SSD. After fitting the 39 soil-data set as cumulative data using nonlinear least squares via the S-PLUS nlsList function, three parameter values (pairs) were arbitrarily selected to represent a soil with low (150, 0.09), medium (200, 0.07), and high (300, 0.08) capacities for supplying available nitrogen (AN). The different curves of this study were generated by varying only the k parameter according to a uniform distribution within ± 0.5 k to obtain a variability resembling that present in the SSD.

Data were computer generated using S-PLUS based on the theory regarding the generation of multivariate normal vectors (Stapleton, 1995). Based on this result given **Z**, a n-vector of standard normal random variables, then (12): $\mathbf{X}=\mathbf{BZ}+\boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is the desired expectation function and **B** is the triangular decomposition of $\sum$ (or Choleski's factor) such that $\mathbf{BB'}=\sum$, and **X** is the vector of simulated values with the desired distribution (Stapleton, 1995).

I applied this result to my programming effort as follows: I defined the expectation function *f* and the desired variance-covariance matrix $\sum$ of the specific model. I generated for the i_{th} iteration a 10×1 vector e_i of independent random normal values with mean 0 and standard

deviation 1 using the rnorm function. I determined the Choleski's factor $\sum^{1/2}$ such that the corresponding model error vector satisfies (13): $\boldsymbol{\varepsilon}_i=\sum^{1/2} \mathbf{e}_i$. Finally, I added the computed value of either $\boldsymbol{f}_1$ or $\boldsymbol{f}_2$, representing the fixed component of the data, to the corresponding random vector $\boldsymbol{\varepsilon}_i$.

Fitting the Data

At each iteration in loop (1), the run is fitted using Method 1. When a complete set of data composed of 40 runs is produced, a simple average of the parameter values is calculated. The same entire data set is also fitted using Method 2. In this case there is no need to take any average, because Method 2 model fits all of the simulated data at the same time, as one data set.

The two pairs of resulting values (M.1average and M.2 direct estimate) are saved and a new iteration (2) is performed. For each of the 24 possible scenarios, 500 iterations (i.e. 500 generated data sets) were performed. At the end of each simulation, summary statistics are computed on the resulting matrix of parameter estimates, composed of 500 rows corresponding to each iteration, and 4 columns containing the two parameter estimates and the corresponding standard errors for N_0 and k calculated by methods 1 and 2. The N_0 parameter root mean squared error (RMSE) were computed for both methods as follows:

Method 1

$$\mathrm{RMSE}_{N_0} = \left[\sum_{p=1}^{500} \frac{[(\sum_{i=1}^{40} \frac{\hat{N}_{0i}}{40})_p - N_0)]^2}{500} \right]^{1/2}$$

Method 2

$$\mathrm{RMSE}_{N_0} = \left[\sum_{p=1}^{500} \frac{(\hat{N}_{0p} - N_0)^2}{500} \right]^{1/2}$$

The quantity $\sum_i$ is the sum over i and $\sum_p$ over p with i=1, . . ., 40 and p=1,...,500. In method 1 $\hat{N}_{0pi}$ is N_0 fitted value during the i^{th} iteration of loop (1) and the p^{th} iteration of loop (2). In method 2, $\hat{N}_{0p}$ is the N_0 fitted value during the p^{th} iteration of loop (2). N_0 and κ are the population parameters used in the model to generate the different data sets.

Method 1: Separate Fits by Run

This method doesn't account for the within run temporal correlation, instead assuming that all the observations are independent of one another. It is the most commonly used by authors in the soil science literature to fit NNM data and therefore the most appropriate to be compared (Cabrera and Kissel, 1994). The considered expectation functions and design are

(14) $E(N_j)=N_0\cdot(1-e^{-k\cdot tj}) + \varepsilon_j,$

(15) $E(\Delta_j)=N_0\cdot e^{-k\cdot tj}\cdot(e^{k\cdot dj}-1) + \varepsilon_j, \quad j=1, . . .,10.$

Similar to (1) and (2), N_j (mg Kg^{-1}) is the cumulative nitrogen mineralized at the j^{th} extraction date for an individual run. N_0 is an initial estimate of potentially mineralizable N, and k the corresponding rate constant of mineralization. For the different scenarios, the initial parameter values were always set equal to the corresponding population N_0 and k used in the models to simulate the data sets [(1) and (2)] to facilitate the convergence of the algorithm. The quantity Δ_j is the incremental amount of N released at the j^{th} extraction date, and d_j is the interval preceding time t_j. As for equations (1) and (2), $\mathbf{t}=(2, 4, 8, 12, 16, 20, 24, 28, 32, 36)^T$ and $\mathbf{d}=(2, 2, 4, 4, 4, 4, 4, 4, 4, 4)^T$. For both model functions, the variance covariance matrix is $\boldsymbol{\Sigma}=\sigma^2\mathbf{I}_{10}$. The NLS function was used to fit the simulated individual runs for both (14) and (15) using, respectively, cumulative and incremental simulated data.

This function uses nonlinear least squares estimation performing iterations via the Gauss Newton (GN) algorithm. I decided to use this procedure with numerical solution (i.e. without specifying the model derivatives) and independence was assumed when using this function.

The Geometry of Nonlinear Least Square Estimation

In this section I give a description of the basic concepts regarding the NLS method commonly being used by authors in the soil science literature for fitting NNM data (Molina et al., 1980, Deans et al. 1986).

To fully understand the NLS function, the structure of the GN algorithm, and more importantly the theoretical foundations of methods 1 and 2, I recall some important concepts of linear least squares geometry that are applied to the fitting algorithms of this work.

Based on Stapleton (1995), in general for a linear model written in the form

(16) $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$,

where $\mathbf{Y}$ is a vector of response variables, $\mathbf{X}$ the model matrix, and $\boldsymbol{\varepsilon}$ the vector of corresponding errors, we assume the following assumptions to be valid:

1) (Additive error) $\mathbf{X}\boldsymbol{\beta}$ is the unknown mean of $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ is an unobserved random vector with mean 0.

2) (Linearity) $\mathbf{X}\boldsymbol{\beta}= \beta_1 x_1+. . .+\beta_k x_k$ where $\beta_1,. . ., \beta_k$ are unknown parameters. This assumptions says that $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \in V=C(\mathbf{X})$ lies in the column space of $\mathbf{X}$. i.e., it is a linear combination of explanatory vectors $\mathbf{x}_1, . . .,\mathbf{x}_k$ with coefficients the unknown parameters $\boldsymbol{\beta}=(\beta_1, . . ., \beta_k)^T$.

3) $var(\mathbf{y})=\sigma^2\mathbf{I}_n$, i.e., $\mathbf{y}$ has a spherical variance-covariance matrix.

4) (Normality) $\boldsymbol{\varepsilon}\sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

The principle of least squares estimation (LSE) minimizes the sum of squared errors for

$\mathbf{y}- \mathbf{X\beta}$, (17) $SSE=\|\mathbf{y}-\mathbf{X\beta}\|^2$.

Given these assumptions and if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent, then we estimate $\boldsymbol{\beta}$ as

(18) $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$.

These basic assumptions for a linear model determine an important series of considerations to approach and solve the problem of parameter estimation in nonlinear model theory. . The assumption $var(\epsilon)=\sigma^2 I$ indicates that the distance between $\mathbf{y}$ and $\mathbf{X\beta}$ is an Euclidean distance and the linearity assumption implies that the expectation surface $\boldsymbol{\eta}(\boldsymbol{\beta})$ is a P-dimensional subspace of N-dimensional Euclidean space which can be written as $\mathbf{X\beta}$, $\boldsymbol{\beta}$ being a P-dimensional parameter vector and $\mathbf{X}$ an N×P derivative matrix. The fitting of the model involves a two step process: obtain the possible response vector $\boldsymbol{\eta}(\boldsymbol{\beta})= \mathbf{X\beta}$ and then calculate the squared distance $\|\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\beta})\|^2$. All the possible vectors $\boldsymbol{\eta}(\boldsymbol{\beta})$ form a P-dimensional expectation surface, called expectation plane in the linear model. The minimizer $\hat{\boldsymbol{\eta}}$ of the least squares criterion $\|\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\beta})\|^2$ is the closest point on the expectation plance, which implies that the residual vector $\mathbf{y}-\hat{\boldsymbol{\eta}}$ will be normal to the expectation plane. This can be also expressed by the so called normal equation: $\mathbf{X}^T(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{0}$.

This theory and results are applied to the Gauss-Newton method for the nonlinear case where the vector of coefficients $\boldsymbol{\beta}$ is now indicated as $\boldsymbol{\theta}$. It performs the following two steps: 1) find on the expectation surface $\boldsymbol{\eta}(\boldsymbol{\theta})$ the closest $\hat{\boldsymbol{\eta}}$ to $\mathbf{y}$**.** In other words, we replace $\boldsymbol{\eta}(\boldsymbol{\theta})$ by the closest plane (planar assumption). 2) Using a linear coordinate system $\mathbf{V}^k(\boldsymbol{\theta}-\boldsymbol{\theta}^k)$, defined below, find the parameter vector $\boldsymbol{\theta}$ which is closest to the point $\hat{\boldsymbol{\eta}}$.

These two objectives are obtained by considering the Taylor expansion up to the first derivative around $\boldsymbol{\theta}^0$ (the initial set of parameter values) as follows:

(19) $\boldsymbol{f}(\mathbf{x}_n,\boldsymbol{\theta}) \approx \boldsymbol{f}(\mathbf{x}_n,\boldsymbol{\theta}^0)+\mathbf{v}_{n1}(\theta_1-\theta_1{}^0)+\mathbf{v}_{n2}(\theta_2-\theta_2{}^0)+ \ldots +\mathbf{v}_{nP}(\theta_P-\theta_P{}^0)$

where $\mathbf{v}_{np} = \partial \boldsymbol{f}(\mathbf{x}_n,\boldsymbol{\theta})/\partial\theta_p\,|_\theta{}^0$ and p=1,2, . . ., P.

Considering all N cases we can write: $\boldsymbol{\eta}(\boldsymbol{\theta})\approx\boldsymbol{\eta}(\boldsymbol{\theta}^0)+\mathbf{V}^0(\boldsymbol{\theta}-\boldsymbol{\theta}^0)$ where $\mathbf{V}^0$ is the N×P derivative matrix with elements $\{\mathbf{v}_{n1}\}$. This is equivalent to approximating the residuals, $\mathbf{z}(\boldsymbol{\theta})=\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\theta})$, by $\mathbf{z}(\boldsymbol{\theta})\approx\mathbf{y}-[\boldsymbol{\eta}(\boldsymbol{\theta}^0)+\mathbf{V}^0\boldsymbol{\delta}]=\mathbf{z}^0-\mathbf{V}^0\boldsymbol{\delta}$ where $\mathbf{z}^0=\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\theta}^0)$ and $\boldsymbol{\delta}=\boldsymbol{\theta}-\boldsymbol{\theta}^0$. The next step is to calculate the Gauss increment $\boldsymbol{\delta}^0$ that minimizes the approximate residual sum of squares,

(20) $RSS=\|\mathbf{z}^0-\mathbf{V}^0\boldsymbol{\delta}\|^2$. This expression is equivalent to (17) for the LSE of a linear model. Therefore in the nonlinear model, (18) becomes: (21) $\boldsymbol{\delta}^0=\{(\mathbf{V}^0)^T\mathbf{V}^0\}^{-1}(\mathbf{V}^0)^T\mathbf{z}^0$.

The model matrix $\mathbf{X}$ is now replaced by the derivative matrix $\mathbf{V}^0$. At this point the Gauss increment $\boldsymbol{\delta}^0$ can be computed in two different ways. One is to use the QR decomposition and the other is to solve (18) directly for the LSE. With the QR decomposition, the model matrix $\mathbf{X}$ is expressed as $\mathbf{X}=\mathbf{QR}$, with the N×N matrix Q and the N×P matrix R constructed so that Q is orthogonal (i.e: $\mathbf{Q}^T\mathbf{Q}=\mathbf{Q}\mathbf{Q}^T=\mathbf{I}$) and $\mathbf{R}$ is zero below the main diagonal. Geometrically, the columns of $\mathbf{Q}$ define an orthonormal basis for the response space, with the property that the first P columns span the expectation plane as defined in Bates and Watts (1988). As presented by these two authors, because of these properties and transformations, we can compute $\boldsymbol{\delta}^0$ without involving any matrix inverse operation. This has two major advantages in performing the algorithm: 1) it reduces the computational time 2) increases its stability. The S-plus NLME functions implement methods based on the QR decomposition. Assuming we have computed the Gauss increment $\boldsymbol{\delta}^0$ using (18), we can now compute $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0+\boldsymbol{\delta}_0$, and find a value $\hat{\boldsymbol{\eta}}^1$ that should be closer to $\mathbf{y}$ than $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$. We can now proceed to perform another iteration by calculating new

residuals $\mathbf{z}^1 = \mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\theta}^1)$, a new derivative matrix $\mathbf{V}^1$, and a new increment. This process is repeated until convergence, when the current increment is small enough to have no significant change on the parameter vector (Bates and Watts, 1988). In order to avoid the situation where the requested increment exceeds the region for which the linear approximation is valid, so that

(22) $S(\boldsymbol{\theta}^1)<S(\boldsymbol{\theta}^0)$,

the step factor λ was introduced (Box, 1960; Hartley, 1961) and used to create:

(23) $\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0+\lambda\boldsymbol{\delta}^0$.

The-step halving factor, λ, is initially set to 1, but can be reduced so that (22) is satisfied. The description of the GN algorithm presented here apply only to the case where there is a homoskedastic and independent model error structure.

Example 1.

To show an application of the GN method, I compute the first two iterations of the GN algorithm following the theoretical steps introduced above. In this example, I use the cumulative model $\boldsymbol{f}_1$ to fit one individual run. For this example, we consider the data for the Amarillo soil series taken from the SSD with $\mathbf{t}^T=(4,8,12,16,22,30)$ and $\mathbf{N}^T=\mathbf{y}^T = (15.8,32,47.7,63.1,78.2,92.6)$ and use as initial starting values $N_0=92.2$ and $k=0.03$. Using the Bates and Watts formulation, we have $\boldsymbol{\theta}^0=(92.2,0.03)^T$ and $\mathbf{N}=N_0\cdot(1-e^{--k\mathbf{t}}) + \boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon}\sim N(\mathbf{0}, \sigma^2\mathbf{I}_6)$.

Step 0 (Table 4): $\boldsymbol{\theta}^0=(92.2,0.03)^T$

Given $\partial\mathbf{N}/\partial N_0=1-e^{-kt}$ and $\partial\mathbf{N}/\partial k=N_0\cdot\mathbf{t}\cdot e^{-2k}\mathbf{t}$, $\mathbf{V}=(\mathbf{V}_{01} \mid \mathbf{V}_{02})$, and performing the following matrix operations in S-plus we calculate: $\boldsymbol{\delta}^0=\{(\mathbf{V}^0)^T\mathbf{V}^0\}^{-1}(\mathbf{V}^0)^T\mathbf{Z}^0=(65.05,\ 0.00079)^T$

Generally the algorithm starts with $\lambda=1$ in (23) and then halves it until (22) is satisfied. Now we can compute the new parameter value,

$\boldsymbol{\theta}^1 = \boldsymbol{\theta}^0+\lambda\boldsymbol{\delta}^0=(92.2,0.03)^T+(65.05,\ 0.00079)^T=(157.05, 0.03079)^T$, assuming $\lambda=1$.

Step 1 (Table 5): $\boldsymbol{\theta}^1=(157.05, 0.03079)^T$

From S-plus output we obtain $\boldsymbol{\delta}^1=\{(\mathbf{V}^1)^T\mathbf{V}^1\}^{-1}(\mathbf{V}^1)^T\mathbf{Z}^1=(0.1327,\ -0.000364)^T$

$\boldsymbol{\theta}^2 = \boldsymbol{\theta}^1+\lambda\boldsymbol{\delta}^1=(157.05, 0.03079)^T+(0.1327,\ -0.000364)^T=(157.18, 0.030428)^T$. To verify that these results are the same as these obtained by the S-PLUS NLS function, I have computed these two iterations using the NLS function with the trace=TRUE specification and have obtained exactly the same results.

Single Level Linear Mixed Effects Model

In order to present the NLMMs applied to this specific case study, I first introduce the theory of linear mixed effects models to create a background necessary to extend the same concepts and

notations to the nonlinear case. To motivate the use of LMM to this study related to NNM, we can consider that each soil sample being repeatedly leached over time is not of interest in itself, but as a representative of a broader population of soils, for example the taxonomic unit "Ultisol", a soil group with well defined chemical and physical properties different from other groups. Each set of repeated AN measures on a specific soil tube varies randomly around the population mean and thus should properly be considered as dependent measures on one another that share the same random specific effect.

To provide an example of a simple linear mixed effects model that may be appropriate for data as our NNM data, consider the one-way random effects model.

That is, (24) $y_{ij}=\mu+b_i+\varepsilon_{ij}$, $i=1, \ldots, K$, $j=1, \ldots, n_i$ with b_i random effect parameter specific for the i^{th} run, $\varepsilon_{ij}\sim N(0, \sigma^2)$, and $b_1,\ldots,b_k\sim N(0,\sigma^2_\beta)$. This model is a special case of the more general model

(25) $\mathbf{y}=\mathbf{X\beta}+\mathbf{Zb}+\boldsymbol{\varepsilon}$, $\mathbf{b}\sim \mathcal{N}(\mathbf{0},\boldsymbol{\Psi})$, $\boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\boldsymbol{\Lambda})$,

with $\mathbf{X}$ being the model matrix of $\boldsymbol{\beta}$ vector of fixed effects coefficients, $\mathbf{Z}$ the random effects design matrix of size $(K\times n_i)\times K$, $\mathbf{b}$ is the $K\times 1$-dimensional vector of random effects normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Psi}$, and $\boldsymbol{\varepsilon}$ the $(K\times n_i)\times 1$ error vector. The errors $\boldsymbol{\varepsilon}$ are assumed to be independent and also independent of the random effects b_i's. $\boldsymbol{\Lambda}$ is a positive-definite matrix parameterized by a fixed, generally small, set of parameters λ. (In the next section I provide a detailed description about the estimation process of $\boldsymbol{\Lambda}$).

Note that model (25) implies

$\text{corr}(y_{ij},y_{i,j+1})=\text{cov}(y_{ij}, y_{i,j+1})/ [\text{var}(y_{ij})\ \text{var}(y_{i,j+1})]^{1/2}$

$=\text{cov}(\mu+b_i+\varepsilon_{ij}, \mu+b_{i+1}+\varepsilon_{i,j+1})/(\sigma^2_\beta+\sigma)=\sigma^2_\beta/(\sigma^2_\beta+\sigma)$ and $\text{var}(y_{ij})=\text{var}(\mu+b_i+\varepsilon_{ij})=\text{var}(b_i+\varepsilon_{ij})=\sigma^2_\beta+\sigma$, for i,j values. These results indicate the general utility of LMMs particularly in modeling

data that are: 1) subject to multiple sources of error (or randomness or heterogeneity), since the variance of the response is assumed to come from multiple sources and 2) are correlated. For example, if we consider a data set of K independent runs each having $n_i=6$ observations and given the results above, we can formulate the symmetric variance-covariance matrix of the vector of responses on the same i^{th} run, $\mathbf{y}_i=(y_{1,}\ldots,y_{ni})^T$,

(26)
$$\mathrm{Var}(\mathbf{y}_i)=\begin{pmatrix} \sigma^2_\beta+\sigma & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta \\ & \sigma^2_\beta+\sigma & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta \\ & & \sigma^2_\beta+\sigma & \sigma^2_\beta & \sigma^2_\beta & \sigma^2_\beta \\ & & & \sigma^2_\beta+\sigma & \sigma^2_\beta & \sigma^2_\beta \\ & & & & \sigma^2_\beta+\sigma & \sigma^2_\beta \\ & & & & & \sigma^2_\beta+\sigma \end{pmatrix}$$

To estimate $\boldsymbol{\beta}$, the mixed model (31) can be written as a weighted least-squares model. Let

(27) $\mathbf{V}\equiv\mathrm{var}(\mathbf{y})=\mathrm{var}(\mathbf{X\beta}+\mathbf{Zb}+\boldsymbol{\varepsilon})=\mathrm{var}(\mathbf{Zb}+\boldsymbol{\varepsilon})=\mathbf{ZDZ}^T+\mathbf{R}$.

If $\mathbf{V}$ were known, then generalized least squares (GLS) estimation theory could be applied to this specific case obtaining $\hat{\boldsymbol{\beta}}$ as a solution of the equation

(28) $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X\beta}=\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$

Knowing $\hat{\boldsymbol{\beta}}$, we can then proceed and calculate the so called best linear unbiased predictor (BLUP) of $\mathbf{b}$. In other words we compute (Rencher, 2000): (29) $\hat{\boldsymbol{\beta}}=(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$

$\mathbf{b}=\mathbf{DZ}^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$ where $\mathbf{V}=\mathbf{ZDZ}^T+\mathbf{R}$. Equivalently we can say that to fit the LMM we need to solve the following mixed model equations:

(30)
$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} \end{pmatrix}\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{D}^{-1}+\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} \end{pmatrix}\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{pmatrix}=\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

However, since $\mathbf{V}$, in most cases, is unknown, it is necessary to use other methods such as the maximum likelihood estimation (MLE) within an iterative approach.

Example 2.

Assuming **V** is known, we fit LMM to a smaller data set similar to the SSD, with only 3 runs and each composed of 2 observations or measurements, y_{ij}, of AN. We assume that a linear mixed effects model (LMM) can be fitted to our data having a one level structure, with each observation grouped by run. Each run shares the same common random effect b_i. As before, we consider a one-way random effects model. We suppose that after each extraction, a certain amount of available nitrogen (AN) is measured, representing the corresponding mineralized nitrogen over a specific length of time (incremental approach). In this case our model can be formulated as: (31) $y_{ij}=\mu+b_j+\varepsilon_{ij}$, i=1,2,3. j=1,2, where $\varepsilon_{11}, \ldots, \varepsilon_{32} \sim$ iid $N(0, \sigma^2)$, b_1, b_2, $b_3 \sim$ iid $N(0, \sigma_\beta^2)$ and the b_j's and ε_{ij}'s are assumed uncorrelated. At the end of this experiment, we obtain the following values of AN, $\mathbf{y}^T=(y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32})=(29,34,27,25,18,23)$, and we assume that $\sigma^2=1$ and $\sigma_\beta^2=2$. For example, during the first extraction of the second run we obtain AN=27. In matrix form model (31) can be formulated as:

$$\underset{\mathbf{y}}{\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix}} = \underset{\mathbf{X}}{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \underset{\mathbf{Z}}{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}} \underset{\mathbf{b}}{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}} + \underset{\boldsymbol{\varepsilon}}{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}}$$

To find the best linear unbiased estimator, BLUE, we use generalized least squares estimation and compute first **V** using (29).

We know that $\mathbf{D}=\sigma_\beta^2\mathbf{I}_3$, and

$$\mathbf{Z}=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{pmatrix}\qquad \mathbf{D}=\begin{pmatrix}\sigma_\beta^2&0&0\\0&\sigma_\beta^2&0\\0&0&\sigma_\beta^2\end{pmatrix}\qquad \mathbf{R}=\begin{pmatrix}\sigma^2&0&0&0&0&0\\0&\sigma^2&0&0&0&0\\0&0&\sigma^2&0&0&0\\0&0&0&\sigma^2&0&0\\0&0&0&0&\sigma^2&0\\0&0&0&0&0&\sigma^2\end{pmatrix}$$

$\mathbf{V}=\mathbf{Z}\mathbf{D}\mathbf{Z}^T+\mathbf{R}$

$$\mathbf{V}=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{pmatrix}\begin{pmatrix}2&0&0\\0&2&0\\0&0&2\end{pmatrix}\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}+$$

$$+\begin{pmatrix}1&0&0&0&0&0\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}=\begin{pmatrix}3&2&0&0&0&0\\2&3&0&0&0&0\\0&0&3&2&0&0\\0&0&2&3&0&0\\0&0&0&0&3&2\\0&0&0&0&2&3\end{pmatrix}$$

We now can compute directly $\hat{\boldsymbol{\beta}}=(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$. A calculator returns: $\hat{\boldsymbol{\beta}}=(17.3,10.9,6.4)$.

The run specific random effects can be now computed as:

$\mathbf{b}=\mathbf{D}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})=$ (2.57, 0, -2.57).

Method 2: Single Level Nonlinear Mixed Effects Model

In the nonlinear settings model (25) can be rewritten as:

(32) $\mathbf{y}_i=\boldsymbol{f}_i(\boldsymbol{\theta}_i, \mathbf{x}_i)+\boldsymbol{\varepsilon}_i,\ \boldsymbol{\theta}_i=\mathbf{A}_i\boldsymbol{\beta}+\mathbf{B}_i\mathbf{b}_i$ for i= 1, . . ., K,

where $\mathbf{y}_i=(y_{i1}, . . ., y_{ini})^T$, $\boldsymbol{\theta}_i=(\theta_{i1}, . . ., \theta_{ini})^T$, $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1}, . . ., \varepsilon_{ini})^T$,

$\boldsymbol{f}_i(\boldsymbol{\theta}_i, \mathbf{x}_i)= [\boldsymbol{f}_i(\boldsymbol{\theta}_{i1}, \mathbf{x}_{i1}), . . ., \boldsymbol{f}_i(\boldsymbol{\theta}_{ini}, \mathbf{x}_{ini})]^T$, $\mathbf{x}_i=(\mathbf{x}_{i1}, . . ., \mathbf{x}_{ini})^T$,

$\mathbf{A}_i=(\mathbf{A}_{i1}, . . ., \mathbf{A}_{ini})^T$, $\mathbf{B}_i=(\mathbf{B}_{i1}, . . ., \mathbf{B}_{ini})^T$, assuming $\mathbf{b}_1, . . .,\mathbf{b}_K \sim$ iid $N_q(\mathbf{0}, \boldsymbol{\psi})$,

and $\boldsymbol{\varepsilon}_i \sim N_{ni}(\mathbf{0},\sigma^2\boldsymbol{\Lambda}_{ni})$. More specifically the corresponding models applied in method 2 can in general be represented as

(33) $N_{ij}=N_{0ij}\cdot[1-\exp(-k_{ij}\cdot t_{ij})] + \varepsilon_{ij}$

(34) $\Delta_{ij}=N_{0ij}\cdot\exp(-k_{ij}\cdot t_{ij})\cdot[\exp(k_{ij}\cdot d_{ij})-1] + \varepsilon_{ij}, \qquad i=1, . . ., K, j=1, . . .,n_i,$

where

$$\begin{bmatrix} N_{0ij} \\ k_{ij} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} N_0 \\ k \end{bmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} (b_{ki})$$

$$\boldsymbol{\phi}_{ij} \qquad \mathbf{A}_{ij} \qquad \boldsymbol{\beta} \qquad \mathbf{B}_{ij} \quad \mathbf{b}_i$$

$\mathbf{b}_i=b_i$ is a scalar and $b_1, . . ., b_K \sim$iid $N(0, \sigma^2_b)$. In this case, the coefficients N_0 and k are indexed for convenience as i and j but they don't actually vary with time t_{ij} and / or lag d_{ij}. N_0 is the initial estimate of potentially mineralizable-N, k is a fixed effect parameter representing the rate constant of mineralization. In this case, $\mathbf{0}$ and $\boldsymbol{\psi}$ are 1×1 matrixes, and Δ_{ij} is the incremental amount of N released during the interval d_{ij}, preceding time t_{ij}. As was the case for method 1, each of the i^{th} runs is composed of $n_i=10$ extraction dates $\mathbf{t}=(2, 4, 8, 12, 16, 20, 24, 28, 32, 36)^T$, and there are K=40 runs.

During the actual simulation, either models (33) or (34) were fitted to the simulated data set using the NLME function, specifying an AR(1) within-run correlation structure with the variance specification

weights=varExp(~Time/36-1) and weights=varExp(~1-Time/36), respectively, for the cumulative and incremental scenarios. To provide a graphical description of the autocorrelation and heteroskedasticity present in the simulated data sets I fitted NLME M0 using both the incremental and cumulative functions in scenarios 1, 4, 8, and 12. These are displayed as figures A-1, A-2, A-3, A-4, A-5, A-6, A-7, and A-8, respectively, in Appendix A.

Considering all of the $\mathbf{y}_i$, we can write in a more general vector form $\mathbf{y}$ (Lindstrom and Bates, 1990):

$\mathbf{y} \mid \mathbf{b} \sim \mathbf{N}(f(\phi,\mathbf{x}), \sigma^2\Lambda)$, $\phi=\mathbf{A}\beta + \mathbf{B}\mathbf{b}$.

This derives from knowing that the conditional density $p(\mathbf{y} \mid \mathbf{b}; \beta, \sigma^2, \mathbf{D})$ given $\mathbf{b}$ is a normal density because the error vector ε_i is normal. $p(\mathbf{y} \mid \mathbf{b}; \beta, \sigma^2, \mathbf{D})$ is given by a multivariate normal density with mean $\boldsymbol{f}(\phi)$ and variance-covariance matrix $\sigma^2\Lambda$.

$\mathbf{b} \sim N(\mathbf{0}, \sigma^2 \widetilde{\mathbf{D}})$, where $\mathbf{B}= \text{diag}(\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K)$, $\mathbf{b}=(\mathbf{b}_1,\mathbf{b}_2, \ldots, \mathbf{b}_K)^T$, $\mathbf{A}=(\mathbf{A}_1,\mathbf{A}_2, \ldots,\mathbf{A}_K)^T$. and $\mathbf{B}_{ij}$ are the model matrices respectively for the fixed and random effects.

Estimation of β and b

Using this estimation method, we try to incorporate the estimation of the nonlinear mixed effects into the context of the linear case. In fact, when the variance components Λ and $\mathbf{D}$ are known and $\boldsymbol{f}$ is a linear function of β and $\mathbf{b}$, the standard estimators of β and $\mathbf{b}$ are respectively the generalized least squares (GLS) estimators

[cf. (29)] $\hat{\beta}=(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$, $\mathbf{b}=\mathbf{D}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\beta})$ where $\mathbf{V}=\mathbf{Z}\mathbf{D}\mathbf{Z}^T+\mathbf{R}$.

These estimates are those that maximize the log likelihood

(35) $l(\boldsymbol{\beta}, \mathbf{b} \mid \mathbf{y}) = -(1/2)\sigma^{-2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})^{T}\boldsymbol{\Lambda}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})-(1/2)\sigma^{-2}\mathbf{b}^{T}\tilde{\mathbf{D}}^{-1}\mathbf{b}.$

We can, for example, fix $\boldsymbol{\beta}$ to obtain the profile loglikelihood of $\boldsymbol{\Lambda}, \tilde{\mathbf{D}},$ which is the marginal density of **b**. Vice-versa, we can fix **b** and obtain the log-likelihood function for $\boldsymbol{\beta}$. The two terms in (35) are a sum of squares and a quadratic term in **b**. By transforming the quadratic term in **b** to an equivalent sum of squares term, we then can treat the optimization purely as a least squares problem, which can be adapted in the nonlinear setting. This is done by augmenting the data vector with "pseudo-data" of the form

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\mathbf{b} + \tilde{\boldsymbol{\varepsilon}},$$

where

$$\tilde{\mathbf{y}} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1/2}\mathbf{y} \\ \mathbf{0} \end{pmatrix} \qquad \tilde{\mathbf{X}} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1/2}\mathbf{X} \\ \mathbf{0} \end{pmatrix} \qquad \tilde{\mathbf{Z}} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1/2}\mathbf{Z} \\ \tilde{\mathbf{D}}^{-1/2} \end{pmatrix} \qquad \tilde{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \sigma^{2}\mathbf{I}),$$

where $\tilde{\mathbf{D}}^{-1/2}=\text{diag}(\mathbf{L}^{-T}, \mathbf{L}^{-T}, \ldots, \mathbf{L}^{-T})$, and **L** is the Cholesky's factor of **D** ($\mathbf{D}=\mathbf{L}^{T}\mathbf{L}$ and **L** is upper-triangular). Similarly, $\boldsymbol{\Lambda}^{-1/2}$ contains the Cholesky factor of $\boldsymbol{\Lambda}_{i}$. Expression (35) can now be rewritten as:

(36) $l(\boldsymbol{\beta}, \mathbf{b} \mid \mathbf{y}) = -(1/2)\sigma^{-2}(\mathbf{y}-\boldsymbol{f}(\mathbf{A}\boldsymbol{\beta}-\mathbf{B}\mathbf{b})^{T}\boldsymbol{\Lambda}^{-1}(\mathbf{y}-\boldsymbol{f}(\mathbf{A}\boldsymbol{\beta}-\mathbf{B}\mathbf{b}))-(1/2)\sigma^{-2}\mathbf{b}^{T}\tilde{\mathbf{D}}^{-1}\mathbf{b}$

The $\hat{\boldsymbol{\beta}}$ which maximizes the objective function in (36) is the maximum likelihood estimate relative to an approximate marginal distribution of **y** commonly defined as *the profile likelihood function*.

Assuming that we have obtained an estimate $\hat{\boldsymbol{\phi}}=\mathbf{A}\hat{\boldsymbol{\beta}}+\mathbf{B}\hat{\boldsymbol{\phi}}$ by using this procedure, we can then proceed and calculate **y** as before by augmenting the data with "pseudo-data".

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{f}}\,(\mathrm{A}\hat{\boldsymbol{\beta}}+\mathrm{B}\hat{\mathbf{b}})+\tilde{\boldsymbol{\varepsilon}},$$

where

$$\tilde{\mathbf{y}} = \begin{bmatrix} \boldsymbol{\Lambda}^{-1/2}\mathbf{y} \\ 0 \end{bmatrix} \qquad \tilde{\boldsymbol{\varepsilon}} \sim \mathrm{N}(\mathbf{0},\sigma^2\mathbf{I}) \qquad \tilde{\boldsymbol{f}}\,(\mathbf{A}\hat{\boldsymbol{\beta}}+\mathbf{B}\hat{\mathbf{b}}) = \begin{bmatrix} \boldsymbol{\Lambda}^{-1/2} f(\mathbf{A}\hat{\boldsymbol{\beta}}+\mathrm{B}\hat{\mathbf{b}}) \\ \tilde{\mathbf{D}}^{-1/2}\hat{\mathbf{b}} \end{bmatrix}$$

Estimation of θ

Our goal in this case is to use the maximum likelihood function of $\mathbf{y} \mid \boldsymbol{\beta},\mathbf{b}$, as stated before. The general relationship between a marginal density $p(\mathbf{y}, \mathbf{b}; \boldsymbol{\beta}, \sigma^2, \mathbf{D})$ and a conditional one is of the form:

$$p(\mathbf{y}, \mathbf{b}; \boldsymbol{\beta}, \sigma^2, \mathbf{D}) = \int p(\mathbf{y}| \mathbf{b}; \boldsymbol{\beta}, \sigma^2, \mathbf{D})\, p(\mathbf{b},\mathbf{D})\mathrm{d}\mathbf{b}$$

a multidimensional integral with respect to **b**. Because $\boldsymbol{f}(\boldsymbol{\beta},\mathbf{b})$ is nonlinear this integral has no closed-form expression and to solve this problem we approximate the conditional distribution of $\mathbf{y} \mid \boldsymbol{\beta}$, **b** using the Taylor's expansion as follows.

Taking the first-order linear Taylor approximation of $\boldsymbol{f}(\boldsymbol{\beta},\mathbf{b})$ about $\hat{\mathbf{b}}$, a predictor of **b**,

(37) $\boldsymbol{f}(\mathbf{A}_i\boldsymbol{\beta}+\mathbf{B}_i\mathbf{b}_i) \approx f(\mathbf{A}_i\boldsymbol{\beta} +\mathbf{B}_i\hat{\mathbf{b}}_i)+ \hat{\mathbf{Z}}\ \mathbf{b}_i- \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i$

where $\hat{\mathbf{Z}}_i=\partial\boldsymbol{\eta}_i/\partial\mathbf{b}_i^{\mathrm{T}}|\,\hat{\boldsymbol{\beta}}_i,\hat{\mathbf{b}}_i$, $\hat{\mathbf{Z}}_i$ is a function of $\hat{\boldsymbol{\phi}}_i=\mathrm{A}_i\hat{\boldsymbol{\beta}}_i+\mathrm{B}_i\hat{\mathbf{b}}_i$ because $\hat{\boldsymbol{\beta}}_i$ and $\hat{\mathbf{b}}_i$ too are functions of $\hat{\boldsymbol{\phi}}_i$. By treating $\hat{\mathbf{b}}_i$, as fixed we now calculate the expected value of (37) as:

$$\mathrm{E}[\boldsymbol{f}(\mathbf{A}_i\boldsymbol{\beta}_i+\mathbf{B}_i\mathbf{b}_i)] \approx f(\mathbf{A}_i\boldsymbol{\beta}_i +\mathbf{B}_i\hat{\mathbf{b}}_i)+ \hat{\mathbf{Z}}_i\,\mathbf{b}_i- \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i \approx f(\mathbf{A}_i\boldsymbol{\beta}_i +\mathbf{B}_i\hat{\mathbf{b}}_i)- \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i$$

Considering that $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\Lambda_i)$, we have

$\mathrm{Var}(\boldsymbol{f}(\mathbf{A}_i\boldsymbol{\beta}_i + \mathbf{B}i\hat{\mathbf{b}}_i) + \hat{\mathbf{Z}}_i\mathbf{b}_i + \mathbf{e}_i) = \mathrm{Var}(\hat{\mathbf{Z}}_i\mathbf{b}_i + \mathbf{e}_i)$, with

$\boldsymbol{f}(\mathbf{A}_i\boldsymbol{\beta}_i + \mathbf{B}i\hat{\mathbf{b}}_i)$ being a fixed component. Then, $\mathrm{Var}(\hat{\mathbf{Z}}_i\mathbf{b}_i + \mathbf{e}_i) = \hat{\mathbf{Z}}_i\mathbf{D}\hat{\mathbf{Z}}_i^T + \sigma^2\Lambda_i$. To ease the computation we decompose $\mathbf{D}$ as $\mathbf{D} = \sigma^2\Delta^{-1}(\Delta^{-1})^T$ and we can write

$$\mathrm{Var}(\hat{\mathbf{Z}}_i\mathbf{b}_i + \mathbf{e}_i) = \hat{\mathbf{Z}}_i\sigma^2\Delta^{-1}(\Delta^{-1})^T\hat{\mathbf{Z}}_i^T + \sigma^2\Lambda = \sigma^2(\hat{\mathbf{Z}}_i\Delta^{-1}(\Delta^{-1})^T\hat{\mathbf{Z}}_i^T + \Lambda) = \sigma^2\Sigma_i\Delta_i$$

by letting $\Sigma_i = (\hat{\mathbf{Z}}_i\Delta_i^{-1}(\Delta_i^{-1})T\hat{\mathbf{Z}}_i^T + \Lambda_i)/\sigma_i^2\Delta_i$.

The likelihood corresponding to the approximate marginal distribution in (35) can now be expressed as:

$$\prod_K(2\pi\sigma^2)^{-ni/2}||\Sigma(\Delta)||^{-1/2} \times \exp[-1/(2\sigma^2)\{\mathbf{y}_i - + \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \hat{\mathbf{b}}_i)\}^T\Sigma_i(\Delta)^{-1}\{\mathbf{y}_i - + \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \hat{\mathbf{b}}_i)\}],$$

with K representing as before the total number of runs.

By taking the log this product quantity is transformed into a sum, representing the corresponding log-likelihood function:

$$-(1/2)\sum_M[n_i \times \log(2\pi\sigma^2) + \log||\Sigma(\Delta_i)|| + \{\mathbf{y}_i - + \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \hat{\mathbf{b}}_i)\}^T\Sigma_i(\Delta)^{-1}\{\mathbf{y}_i - + \hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i - \boldsymbol{\eta}(\boldsymbol{\beta}, \hat{\mathbf{b}}_i)]$$

This loglikelihood is maximized through a series of iterations, alternating between a step to estimate β and obtain the predictor $\hat{\mathbf{b}}_i$, for fixed Δ and a step to estimate Δ for fixed values of $\boldsymbol{\beta}$ and $\hat{\mathbf{b}}_i$.

Iteration Main Structure

In this section I provide a brief explanation on how the different calculations presented merge together in one iteration. In step 1, also defined the linear mixed effects model (LME) step of one iteration, we estimate $\boldsymbol{\beta}$ and update the predictor $\hat{\mathbf{b}}_i$. By ignoring the dependence of $\Sigma(\Delta)$ on β,

we obtain these quantities by maximizing the log-likelihood function with respect to $\boldsymbol{\beta}$ and $\hat{\mathbf{b}}_i$ with i=1, . . .,K. The objective function in (36) becomes

$$\Sigma_K\{ y_i-+\hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i-\eta(\beta,\hat{\mathbf{b}}_i)\}^T\Sigma_i(\Delta)^{-1}\{ y_i-+\hat{\mathbf{Z}}_i\hat{\mathbf{b}}_i-\eta(\beta,\hat{\mathbf{b}}_i)\}==\Sigma_K[||(y_i-\eta(\beta,\hat{\mathbf{b}}_i)||^2+||\Delta\hat{\mathbf{b}}_i||^2]$$

The first part $\sum_M[||(\mathbf{y}_i-\boldsymbol{\eta}(\boldsymbol{\beta},\hat{\mathbf{b}}_i)||^2$ can be viewed as the component present in the nonlinear least squares (NLS) function. Instead the term $||\boldsymbol{\Delta}\hat{\mathbf{b}}_i||^2$ represents a sort of added "penalty" and for this reason step 2 called the *penalized least squares* (PNLS) *criterion*.

In the PNLS step, we first substitute our estimate $\hat{\boldsymbol{\beta}}(\boldsymbol{\Delta})$, the estimator of $\boldsymbol{\beta}$ based on the current $\boldsymbol{\Delta}$, into the log-likelihood function. By substitution we obtain:

$$-(1/2)\sum_M[n_i\times\log(2\pi\sigma^2)+\log|\sum(\Delta_i)^{-1}|+\{w_i-\hat{\mathbf{X}}_i\beta\}^T\Sigma_i\Delta_i\{ w_i-\hat{\mathbf{X}}_i\beta\}]_{\beta=\beta(\Delta)}$$

where $\hat{\mathbf{X}}_i=\partial f_i(\beta,\hat{\mathbf{b}}_i)/\partial\beta^T$, and $w_i=y_i-f_i(\beta,\hat{\mathbf{b}}_i)$.

The obtained objective function is the log likelihood of a linear mixed effects model with:

$w=(w_1^T,\ldots,w_M^T)^T$ the response vector, and $\hat{\mathbf{X}}=(\hat{\mathbf{X}}_1,\ldots,\hat{\mathbf{X}}_M)^T$ and

$\hat{\mathbf{Z}}=(\hat{\mathbf{Z}}_1,\ldots,\hat{\mathbf{Z}}_M)$ respectively, being the design matrices for the fixed and the random effects. This part of one iteration is then solved using the standard techniques for a linear mixed-effects model. Therefore a residual sum of squared errors is computed each time.

These two steps are repeated until convergence.

Example 3. Fitting one simulated data set in scenario 12.

In this example, I fit two different models to a simulated data set of 40 runs in scenario 12 (i.e., N_0=300, k=0.08, ρ=0.9) for the cumulative function (f_1). Initially I use the NLME-M0cum function without any particular specification about the structure of the variance-covariance matrix (Appendix A, figures A4 and A8). M0 is not satisfactory because of the presence of autocorrelation and heteroskedasticity (Figure 2.1). I then fit NLME-M1cum (Figure 2.2) with the appropriate variance-covariance structure: an AR1 type and heteroskedastic transformation (3). M2 improves the previous fit, eliminating the significant autocorrelation and reducing the presence of heteroskedasticity.

Preliminary Analyses of the SSD

To gather some preliminary information about the presence or absence of autocorrelation in the SSD, I initially fitted a series of 20 arbitrarily selected runs independently using the cumulative approach (14) similar to the approach used in other studies (Molina et al., 1980; Talpaz et al., 1981; Deans et al. 1986). Successively using the NLME function without any particular variance-covariance specification, I fitted model M0 only to the "Ultisol" data grouped by run (this data consists of 9 runs) for both the incremental and cumulative approach.

CHAPTER 3

RESULTS AND DISCUSSION

The Ultisol Soil Order in the SSD

As a result of this preliminary investigation when modeling SSD by run, no significant autocorrelation was found for the 20 selected soil series at a significance level α=0.05 (the autocorrelation was significant at α=0.1) in agreement with the results and modeling assumptions of other authors (Molina et al., 1980; Deans et al. 1986; Talpaz et al. 1981; Cabrera and Kissel, 1988; Maimone et al. 1991).

When fitting M0 to the Ultisol group as a unique data set there was significant autocorrelation (α=0.05) and presence of heteroskedasticity as is shown in the M0cum ACF, residual, and QQ-plot (Figure 3). Because of the violation of the assumptions regarding error homoskedasticity and independence, M0cum doesn't adequately fit the data as presented in the fixed- and random-effect curve plot.

Similar results have been found when fitting the corresponding M0inc (Figure 4). There is significant autocorrelation at different lags and heteroskedasticity, resulting in a poor fit of the data. The residuals plot suggests that this model is inappropriate to fit this data set. The purpose of this analysis is to show evidence of autocorrelation in the SSD and underscore that it is necessary to account for the temporal autocorrelation when modeling NNM data.

When correlation is present it is necessary to introduce new parameters into the model to be estimated (e.g the partial autocorrelation coefficients of the AR(1)). Hess and Schimdt (1995), testing different models on soil respiration data, have found that fitting incremental data rather than the corresponding cumulative can reduce the presence of autocorrelation, avoiding the estimation of additional correlation parameters. In fact, using cumulative rather than incremental data induces serial correlation of the residuals. However these authors recognize that choosing the appropriate model function (i.e. the incremental rather than cumulative model) doesn't necessarily eliminate the presence of autocorrelation if the observations are indeed correlated. The "Ultisol" data example, whether modeled as cumulative or incremental data, shows a significant correlation. Because the same soil sample is being repeatedly measured over time, these measures should properly be considered as dependent upon one another. When fitting a larger sample size, the present autocorrelation becomes more evident than when modeling individual runs. In using NLMMs there is the advantage of introducing new parameters into the model to account for autocorrelation without as drastic a loss in terms of degrees of freedom as when using individual run models based on a much smaller sample.

Autocorrelation

The choice of a run composed of 10 observations such as used in the Cabrera's and Kissel's study (1988) rather than just 6 as in the SSD was for convenience in obtaining the desired AR(1) residual autocorrelation pattern varying with lag in the simulated data. In fact, when the number of lags was small (i.e. 6) it was impossible to produce the characteristic ACF tailing off after lag 1. As a consequence, fitting an AR(1) model was not a satisfactory choice for many 6 observation-data sets, having a different and not well-defined autocorrelation structure. The appropriateness of fitting an AR(1) model to the 10 observation-simulated data was evident

when considering runs composed of 100 observations. These simulations were performed to check if indeed fitting an AR(1) model to the simulated data was the most appropriate choice.

Heteroskedasticity and Homoskedasticity

Based on these considerations, the variance-covariance structure for the different heteroskedastic scenarios was assumed to increase with time as if the measurement (laboratory) errors were "summed up" [as in (6)] over time in the cumulative data (Appendix A: figures C8 and C12 show a right-opening megaphone shape).

The opposite variance transformation has been used for the incremental heteroskedastic scenarios, in the attempt of describing the biological processes involved (Appendix A, figures I8 and I12: a left-opening megaphone shape). During the initial phases of the incubation the variance is larger because of large flushes of AN due to the decomposition of the so called "light" organic fraction. As other authors have reported when modeling NNM data, it is necessary to account for these initial flushes due to the most labile organic matter pool (Deans et al., 1986; Ellert and Bettany, 1988).

Based on the results reported by Hess and Schimdt (1995) for soil respiration data, a homoskedastic residual pattern is most likely an appropriate choice to represent the residual distribution for an incremental model. To my knowledge, no author has ever discussed the problem of estimating the variability associated with NNM incremental models. The initial choice of creating different curves for both model functions by varying the k parameter and maintaining a relatively small variance (1.44), was driven by the necessity of generating realistic data sets in all possible cases. Larger variance values (e.g. 16-25) would generate decreasing patterns in the cumulative (i.e. final extraction date values smaller than previous and adjacent one) and negative values in the incremental case. This limitation in the choice of the variance

was mainly for the cumulative function. Given this result, a second set of simulations was performed with σ^2=4 only for the incremental scenarios, with the intent of investigating whether a larger population variance determines any difference compared to the other case.

Method 1 versus Method 2

Under the heteroskedastic and correlated scenarios, modeling the individual runs separately can be acceptable, but the fitted parameters are not the best unbiased estimators of the population parameters, as shown by the results contained in Tables 2 and3. The same results are summarized in Figure 5, representing the regression of RMSE of the estimated N_0 using method 2 versus the corresponding values of method 1-RMSE for all the 24 simulations. The diagonal represents the line for which there is a 1:1 correspondence (correlation coefficient r=1) between root mean squared error for method 2 (RMSE_2) and the corresponding computed with method 1 (RMSE_1). In other words points on this line would represent simulations for which there was no difference in RMSE between method 1 and method 2. Points below this line in the lower triangle are simulations for which RMSE_2 is greater than RMSE_1, and vice-versa for the dots above the r=1-line in the upper triangle. It is evident for both the cumulative and incremental cases that RMSE_2 is lower than RMSE_1 (Figure 5), indicating that the error in estimating N_0 is greater when using method 1. Overall, RMSE for the incremental cases is larger than the corresponding cumulative RMSE. It has the smallest values in the homoskedastic and uncorrelated scenarios (1, 2, and 3) and largest in scenario 11. The incremental scenarios 10-12 are those with the largest error. These results agree in that scenarios 1-3 represent the simplest conditions to compare the two methods, while 10-12 are the most complex. Considering the same variance for cumulative and incremental scenarios induces a relatively larger variation in the simulated data for this second case, because the fixed component of the generated data with

the incremental approach is relatively lower than the corresponding cumulative one with respect to time. Based on Tables 2, 3, and Figure 5, both methods produce N_0 parameter estimates very close to the population target. However method 2 is relatively more precise and accurate than method 1.

In creating the different data sets, particular attention was given to generating curves very different from one another as in a real experiment, maintaining the parameter of major interest N_0 as the fixed target. The k parameter was forced to vary according to a uniform distribution, to avoid possible biases towards the mixed effects model that assumes a random normal distribution of errors and random effects. Method 1 has proven to be highly precise and accurate in parameter estimation, but the use of NLMMs allow for even a relatively greater precision and accuracy. In absolute terms, estimating N_0 with either method is approximately the same. Differences in terms of a decimal g of N Kg^{-1} soil are not very meaningful to the soil scientist, considering the large variability associated with this laboratory technique. However the major advantage of using method 2 as opposed to 1 is in obtaining unbiased estimates of the standard error (i.e. root mean squared error).

The larger and biased root mean squared errors of method 1 depend on the unaccounted temporal dependency. Method 2 accounts for this factor and produces unbiased standard errors to form confidence intervals around the N_0 parameter. No dependence was found between the overestimation of the N_0 parameter and increasing the population variance: in fact the resulting N_0 estimates and corresponding RMSEs of the additional simulations for the incremental scenarios were very similar to those obtained during the first set of simulations.

CHAPTER 4

CONCLUSIONS

This simulation study has proven the importance of accounting for the presence of autocorrelation to obtain better parameter estimates when modeling NNM data under a variety of possible scenarios including independence, correlation, and heteroskedasticity in the observed data.

The use of nonlinear mixed effects models offers the opportunity to formulate models relatively more precise and accurate then those commonly applied in these studies, obtaining parameter estimates based on unbiased standard errors. When fitting models based on wrong assumptions, we reduce the precision of our estimation and in general over estimate the true N_0 parameter value. This study has proven that when we do not consider important model assumptions our standard error estimation (or root mean squared error) is biased. Whether or not NLMMs offer an improvement in absolute terms in estimating N_0, there is not any valid reason for not accounting for the temporal autocorrelation in one's modeling effort when it is indeed present.

REFERENCES

Bates, D.M., and D.G. Watts. 1988. Nonlinear regression analysis and its applications. John Wiley & Sons, New York.

Box, G.E.P. 1960. Fitting empirical data. Annals of the New York Academy of Sciences. 86: 792-816.

Cabrera, M.L., and D.E. Kissel. 1988. Length of incubation time affects the parameter values of the double exponential model of nitrogen mineralization. Soil Sci. Soc. Am. J. 52: 1086-1087.

Cabrera, M.L., and D.E. Kissel. 1988. Potentially mineralizable nitrogen in disturbed and undisturbed soil samples. Soil Sci. Soc. Am. J. 52: 1010-1015.

Cabrera, M.L., D.E. Kissel, M.F. Vigil. 1994. Potential mineralization: laboratory and field evaluation. Pages 15-30 *in* Soil testing: Prospects for Improving Nutrient Recommendations, SSSA Special Publication 40.

Deans, J.R., J.A.E. Molina, and C.E. Clapp. 1986. Models for predicting potentially mineralizable nitrogen and decomposition rate constants. Soil Sci. Soc. Am. J. 50: 323-326.

Ellert, B.H., and J.R. Bettany. 1988. Comparison of kinetics models for describing net sulfur and nitrogen mineralization. Soil Sci. Soc. Am. J. 52: 1692-1702.

Hall, D.B., R.L. Bailey, 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. Forest Science

Hartley, H.O. 1961. The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. Technometrics 3: 269-280.

Hess, T.F., and S.K. Schmidt. 1995. Improved procedure for obtaining statistically valid parameter estimates from soil respiration data. Soil Biol. Biochem. 27(1): 1-7.

Lindstrom, M.J., and D.M. Bates. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics. 46: 673-687.

Littel, R.C., G.A. Miliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS(c) system for mixed models. SAS Institute Inc., Cary, NC. 633 p. [Online, URL:http://www.rcr.uga.edu/sasdoc/onldoc.htm.]

Maimone, R.A., L.A. Morris, and T.R. Fox. 1991. Soil nitrogen mineralization potential in a fertilized loblolly pine plantation. Soil Sci. Soc. Am. J. 55: 522-527.

Molina, J.A.E., C.E. Clapp, and W.E. Larson. 1980. Potentially mineralizable nitrogen in soil: The simple exponential model does not apply for the first 12 weeks of incubation. Soil Sci. Soc. Am. J. 44: 442-443.

Pinheiro, J.C. and D.M. Bates. 2000. Mixed-effects models in S and S-PLUS. Springer, New York. 528 p.

Rencher, A.C. 2000. Linear Models in Statistics. John Wiley & Sons, New York.

Robinson, J.A. 1985. Determining microbial kinetic parameters using nonlinear regression analysis. Pages 61-114 *in* Advances in Microbial Ecology Vol. 8.

S-PLUS 2000. Guide to Statistics. Volume 1. Data Analysis Products Division, MathSoft, Seattle, WA.

S-PLUS 2000. Programmer's guide. Data Analysis Products Division, MathSoft, Seattle, WA.

Stanford, G., and S.J. Smith. 1972. Nitrogen mineralization potentials of soils. Soil Sci. Soc. Am. Proc. 36: 465-472.

Stapleton, J.H. 1995. Linear Statistical models. John Wiley & Sons, New York

Talpaz, H., P. Fine, and B. Bar-Yosef. 1981. On the estimation of N mineralization parameters from incubation experiments. Soil Sci. Soc. Am. J. 53: 91-99.

Wei, W.W-S. 1990. Time Series Analysis. Addison-Wesley Publishing Company, Redwood City, CA.

Table 1. Simulation study factorial design applied to both the cumulative (f_1) and incremental model function (f_2).

Scenario	N_0 , k‡	Var-Cov. Structure	
		ρ†	Var. structure type
1	150, 0.09	0	Homosk; σ^2=1.44§
2	200, 0.07	0	Homosk; σ^2=1.44
3	300, 0.08	0	Homosk; σ^2=1.44
4	150, 0.09	0.9	Simple AR(1); σ^2=1.44
5	200, 0.07	0.9	Simple AR(1); σ^2=1.44
6	300, 0.08	0.9	Simple AR(1); σ^2=1.44
7	150, 0.09	0	Heteroskedasticity
8	200, 0.07	0	Heteroskedasticity
9	300,0.08	0	Heteroskedasticity
10	150, 0.09	0.9	AR(1) Hybrid*
11	200, 0.07	0.9	AR(1) Hybrid
12	300,0.08	0.9	AR(1) Hybrid

‡ Parameter combination. † Correlation coefficient. § The variance term of the variance-covariance matrix was always set to be 1.44. * The chosen $\sum$ determines an increasing and decreasing variance with time respectively for the cumulative and incremental model.

Table 2. N_0 parameter estimates and root mean squared errors (RMSE) of the 12 cumulative (C) simulations computed with the two different methods.

Scenario	§ $\hat{N}_0$ (1)	†RMSE1	$\hat{N}_0$ (2)	RMSE2
C1	150.0191	0.23069	150.0187	0.166243
C2	200.0265	0.34137	200.0203	0.223958
C3	300.0093	0.27475	300.0065	0.18886
C4	150.0589	0.332593	150.0404	0.246543
C5	200.0847	0.483374	200.047	0.320241
C6	300.0414	0.389002	300.0219	0.273713
C7	150.0093	0.188665	150.0075	0.122127
C8	200.0128	0.27485	200.0069	0.163821
C9	300.0035	0.223529	300.002	0.139132
C10	150.0275	0.264404	150.0231	0.189416
C11	200.2305	0.94813	200.1608	0.639929
C12	300.0195	0.306163	300.0145	0.212916

§ Estimated N_0 parameter computed with both method (1) and (2). † Corresponding RMSE for N_0.

Table 3. N_0 parameter estimates and root mean squared errors (RMSE) of the 12 incremental (I) simulations computed with method 1 (M1) and method 2 (M2).

Scenario	§ $\hat{N}_0$ (1)	†RMSE1	$\hat{N}_0$ (2)	RMSE2
I1	150.1586	0.780077	150.1507	0.694976
I2	200.2112	0.980798	200.1587	0.816116
I3	300.1166	0.850686	300.0989	0.73615
I4	150.3999	1.485588	150.3674	1.211657
I5	200.5203	1.829722	200.436	1.465213
I6	300.2775	1.582531	300.2603	1.28368
I7	150.3235	1.28835	150.2659	1.22012
I8	200.4664	1.509959	200.1806	1.3101
I9	300.2379	1.335408	300.1461	1.243283
I10	150.7982	2.446561	151.0108	2.330947
I11	201.1309	2.861787	200.8644	2.484665
I12	300.5401	2.471982	300.6257	2.285935

§ Estimated N_0 parameter computed with both method (1) and (2). † Corresponding RMSE for N_0 computed using (13).

Table 4. Step by step calculations in iteration 1 of the Guass-Newton algorithm fitting the Amarillo soil series data of SSD.

				V_{01}	V_{02}
$\boldsymbol{\eta}(\boldsymbol{\theta}^0)$	$N_0 \cdot (1-e^{-kt})$	$\mathbf{z}^0 = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^0)$	$\mathbf{z}^0$	$1-e^{-kt}$	$t \cdot e^{-2kt}$
10.4	$92.2 \cdot (1-e^{-(0.03)(4)})$	15.8-10.4	5.4	0.11	326.4
19.7	$92.2 \cdot (1-e^{-(0.03)(8)})$	32-19.7	12.3	0.21	578.9
27.9	$92.2 \cdot (1-e^{-(0.03)(12)})$	47.7-27.9	19.8	0.30	770.2
35.1	$92.2 \cdot (1-e^{-(0.03)(16)})$	63.1-35.1	27.9	0.38	910.8
44.5	$92.2 \cdot (1-e^{-(0.03)(22)})$	78.2-44.5	33.6	0.48	1046.1
54.7	$92.2 \cdot (1-e^{-(0.03)(30)})$	92.6-54.7	37.9	0.59	1122.1

Table 5. Step by step calculations in iteration 2 of the Guass-Newton algorithm fitting the Amarillo soil series data of SSD.

			V_{11}	V_{12}
$\eta(\theta_1)$	$N_0 \cdot (1 - e^{-kt})$	z_1	$1 - e^{-kt}$	$t \cdot e^{-2kt}$
18.2	$157.05 \cdot (1 - e^{-(0.03079)\,(4)})$	-2.4	0.11	555.4
34.3	$157.05 \cdot (1 - e^{-(0.03079)\,(8)})$	-2.3	0.21	982.1
48.5	$157.05 \cdot (1 - e^{-(0.03079)\,(12)})$	-0.8	0.31	1302.4
61.1	$157.05 \cdot (1 - e^{-(0.03079)\,(16)})$	2.0	0.38	1535.3
77.3	$157.05 \cdot (1 - e^{-(0.03079)\,(22)})$	0.9	0.49	1754.9
54.7	$157.05 \cdot (1 - e^{-(0.03079)\,(30)})$	37.9	0.60	1870.5

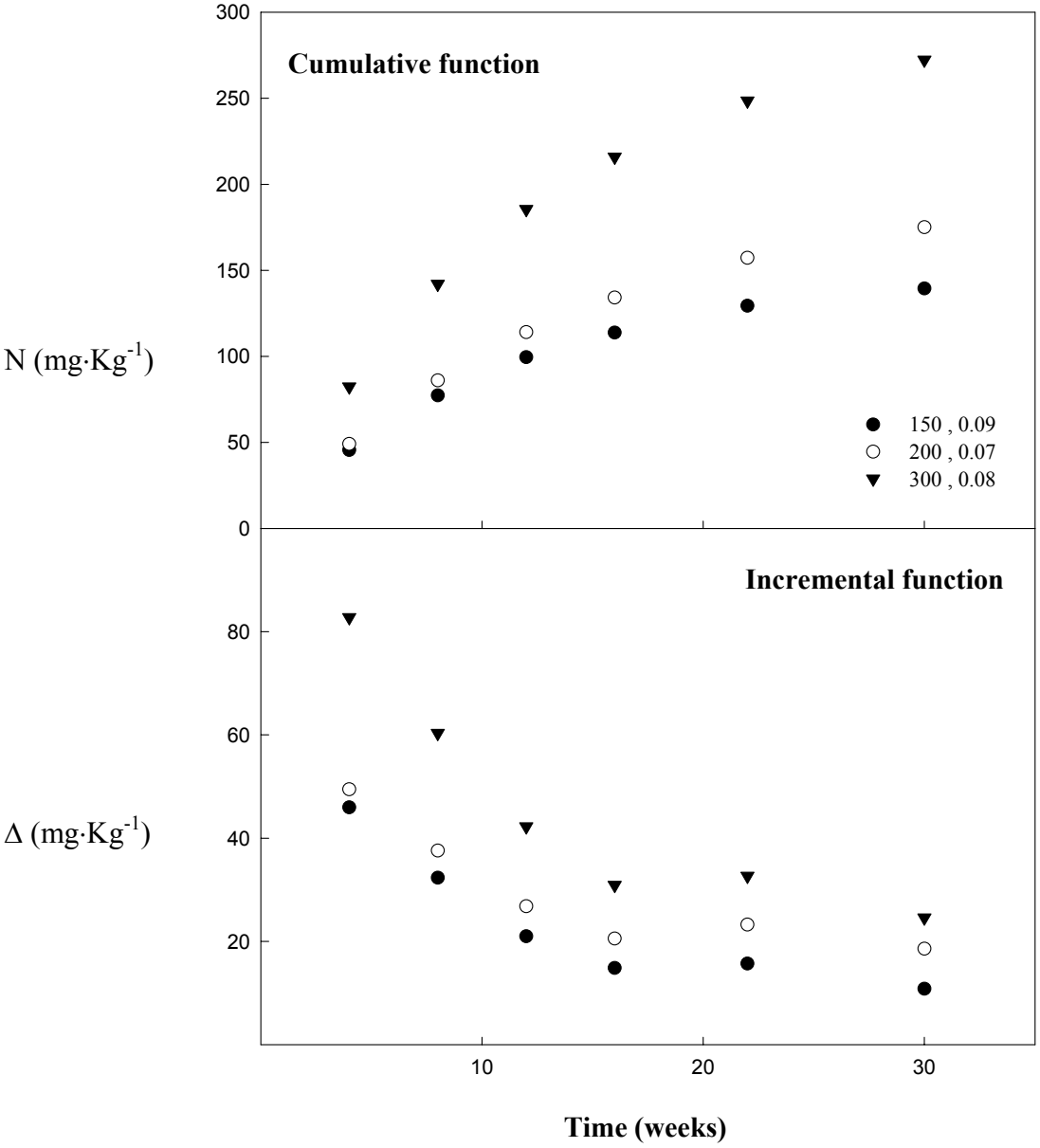

Fig. 1. Plot of cumulative (N) versus time and of incremental nitrogen (Δ) versus time based on simulated data for the three selected parameter combinations with only six observations similarly to the SSD.

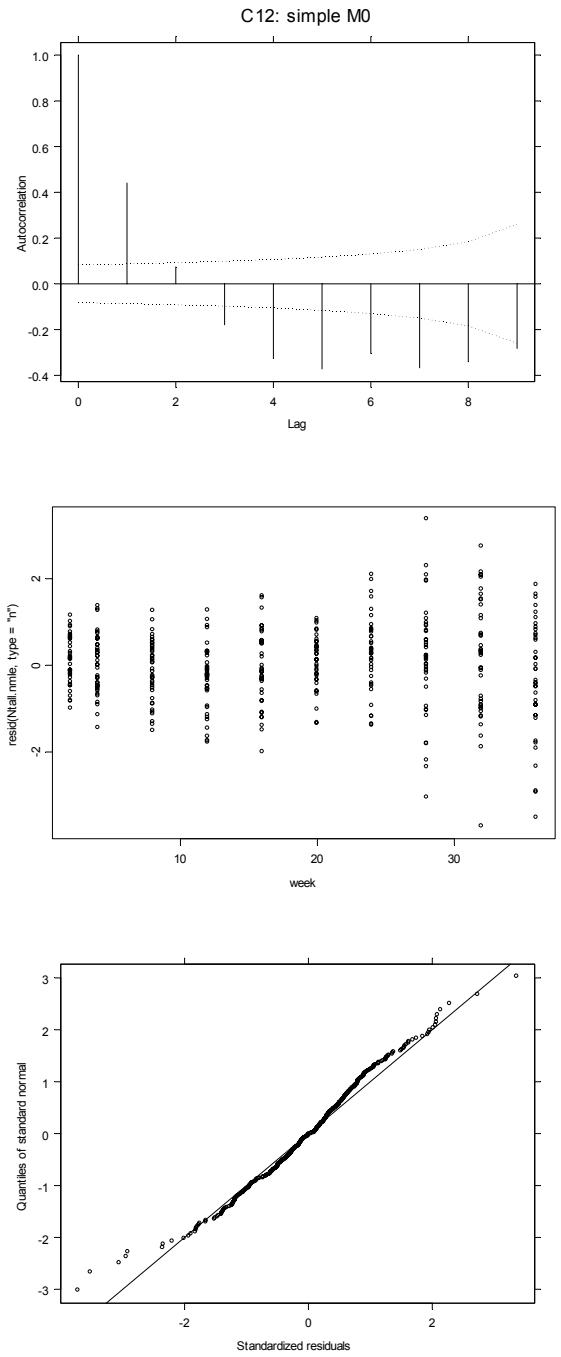

Fig 2.1. Diagnostic plots for NLME-M0cum fitted to a simulated data set under scenario 12. The ACF plot indicates the presence of significant autocorrelation (α=0.1) at lag 1. The QQ- and residuals plot shows the presence of increasing variance with time.

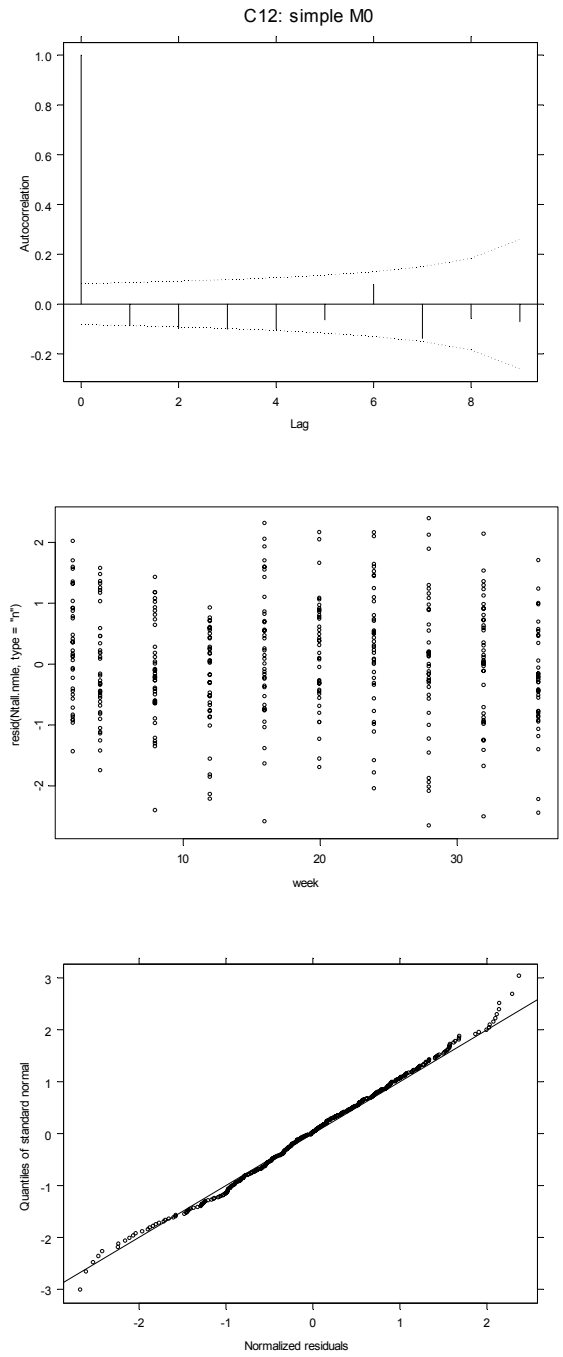

Fig 2.2. Diagnostic plots for NLME-M1cum fitted to the same simulated data set as for M0 under scenario 12. NLME-M1 eliminates the presence of significant autocorrelation (α=0.1) at lag 1 and reduces the presence of heteroskedasticity.

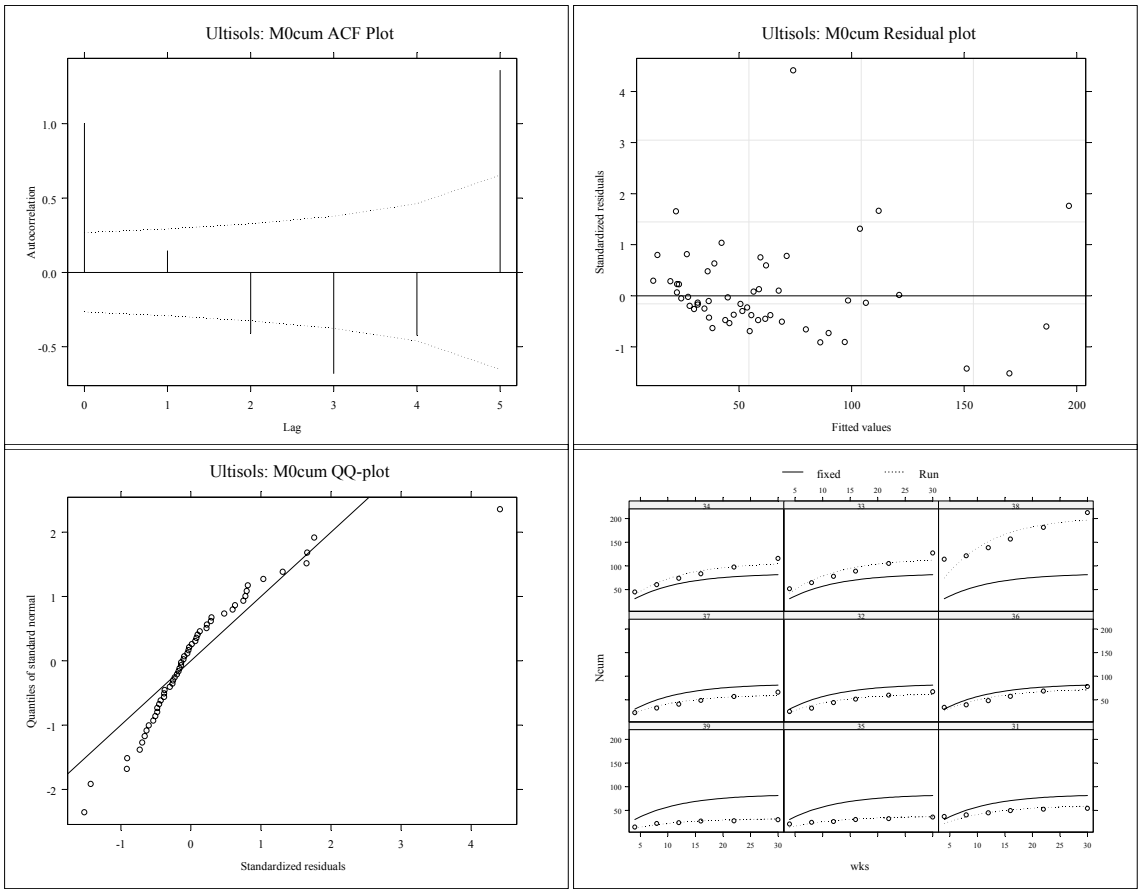

Fig 3. Diagnostic plots and fitted value curves for M0cum. Significant ACF for almost all lags and presence of heteroskedasticity, indicated by the poor fit of the random specific effect curves individualized for each individual run in the Ultisol group.

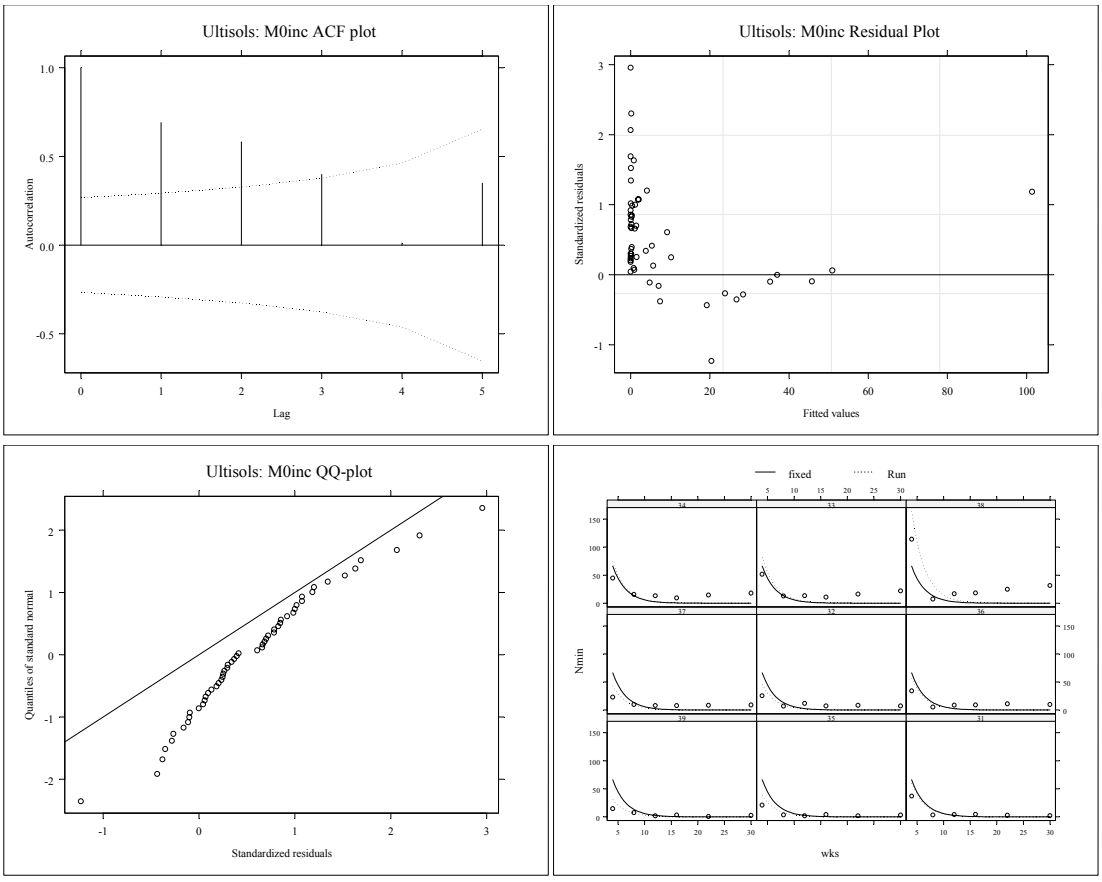

Fig 4. Diagnostic plots and fitted value curves for M0inc. The significant ACF and presence of heteroskedasticity indicates the inappropriateness of this model.

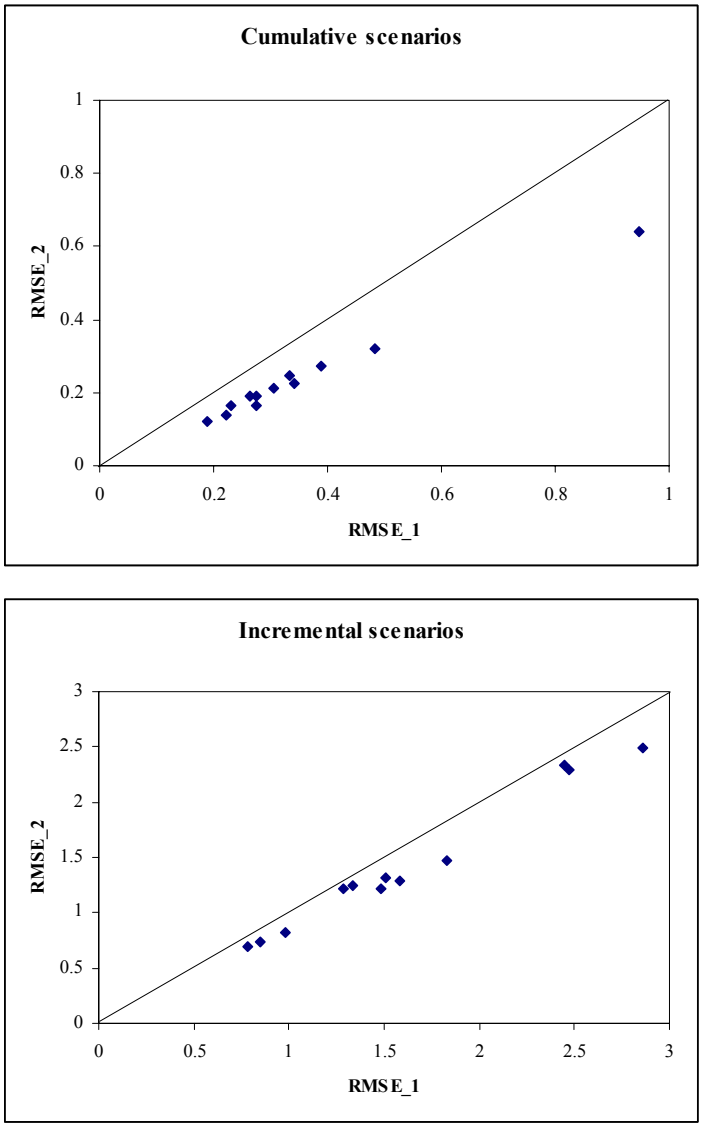

Figure 5. Comparison of the relationships between root mean squared errors for the estimate of N_0 parameter considering all possible simulation scenarios using the r=1-line. Each dot represents one specific simulation scenario for the cumulative and incremental functions.

APPENDIX A

DIAGNOSTIC PLOTS OF MODEL M0

Diagnostic plots of model M0 (nonlinear mixed effects model without any particular specification regarding the autocorrelation and heteroskedasticity structure) fitted to four possible scenarios of this study for both the incremental and cumulative model. Model M0 is a nonlinear mixed effects model that fits the entire simulated data set during each iteration of the program. The header of each individual plot specifies the simulation scenario as described in Tab. 1 for either the cumulative (A-1 to A-4) or the incremental (A-5 to A-8) approach. In each case the autocorrelation function, residuals, and standardized residuals are presented to detect the presence of autocorrelation and/or heteroskedasticity in the simulated data sets.

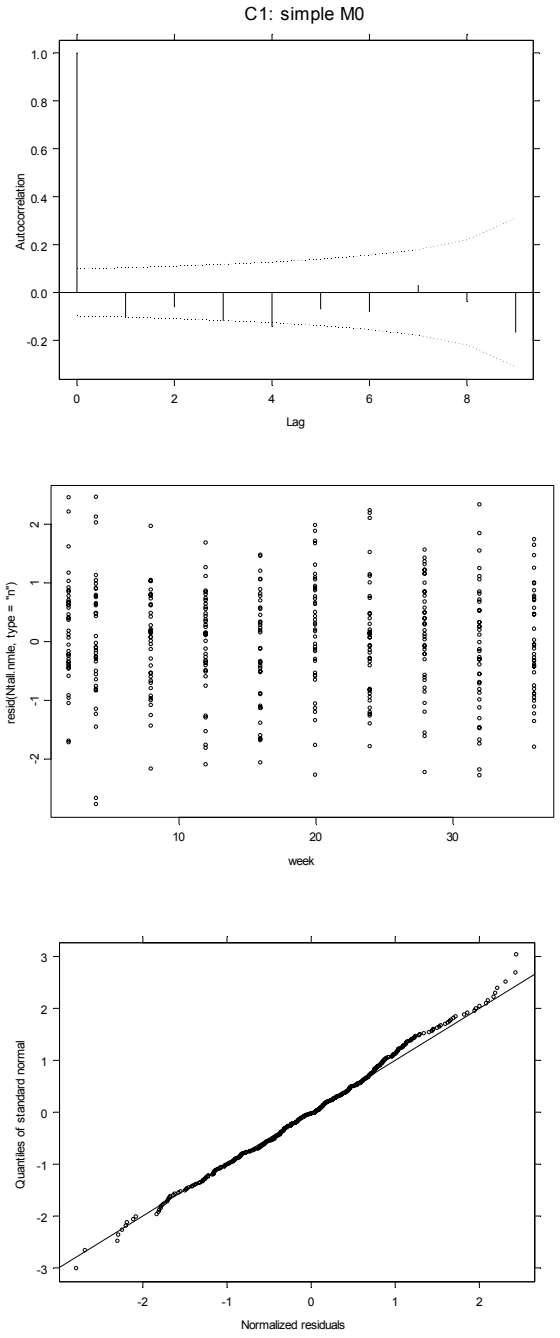

Figure A-1: Cumulative scenario 1.

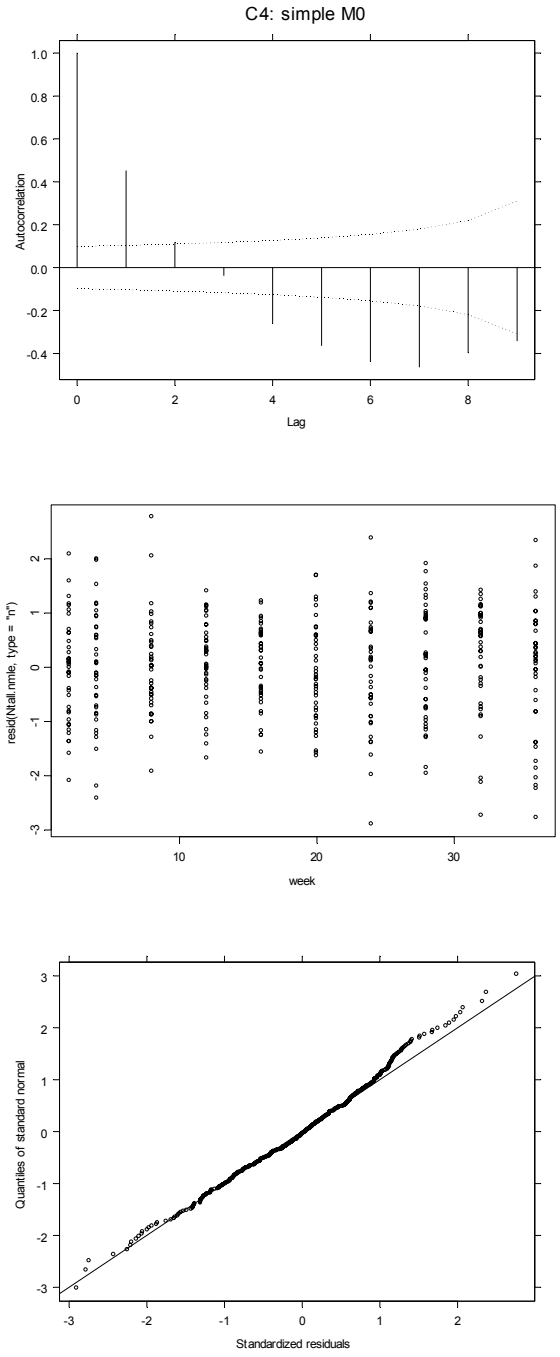

Figure A-2: Cumulative scenario 4.

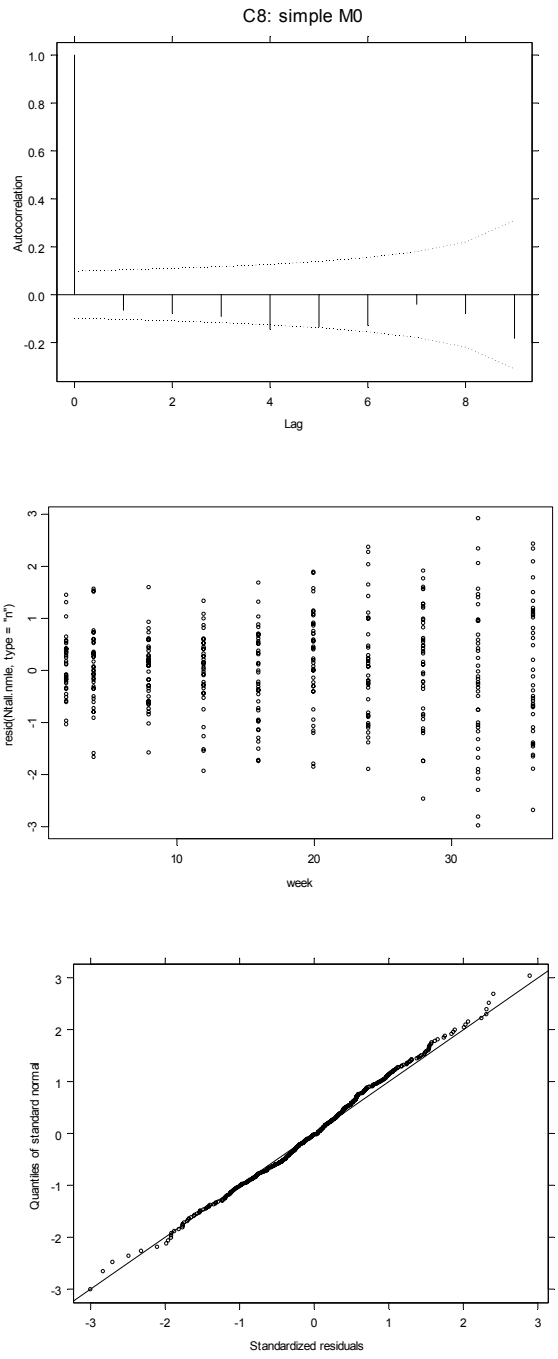

Figure A-3: Cumulative scenario 8.

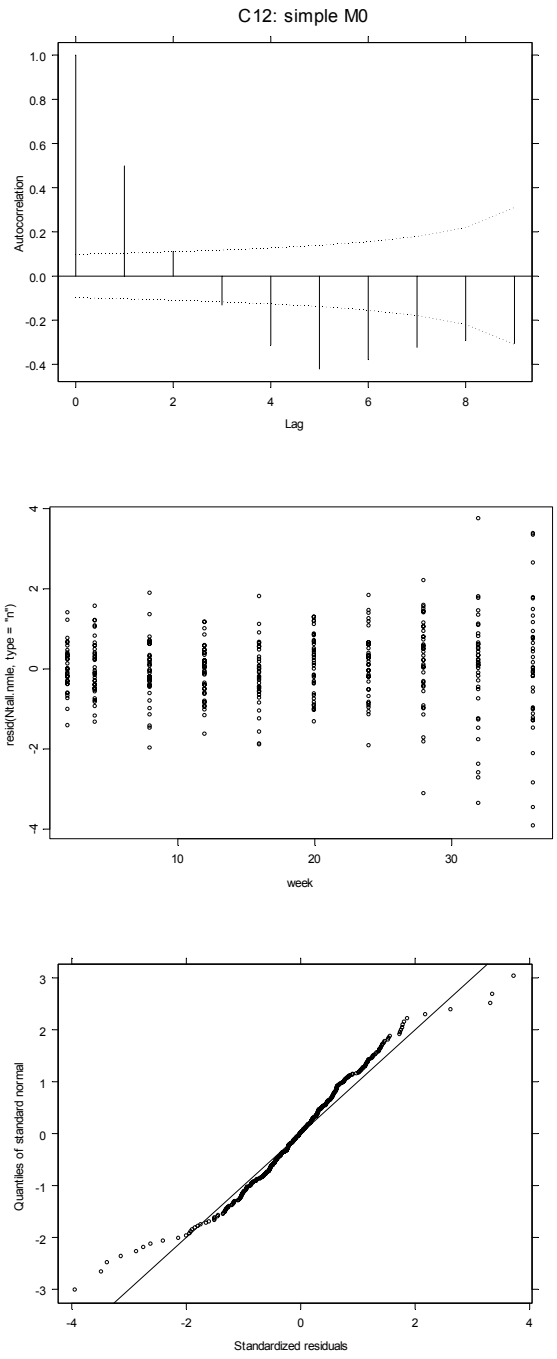

Figure A-4: Cumulative scenario 12.

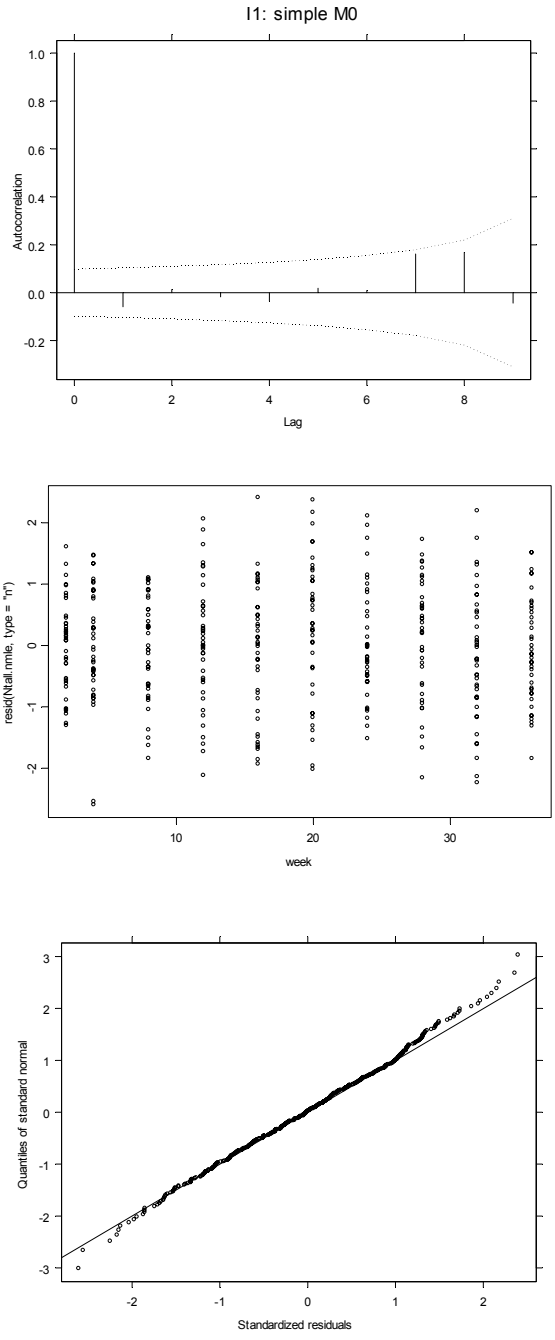

Figure A-5: Incremental scenario 1.

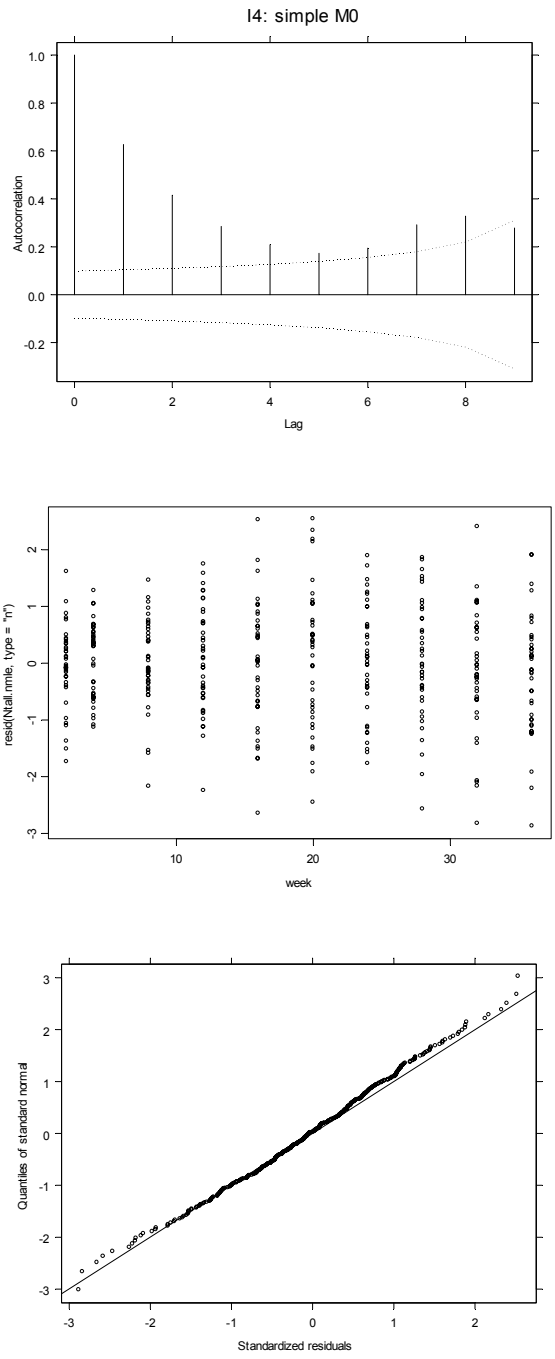

Figure A-6: Incremental scenario 4.

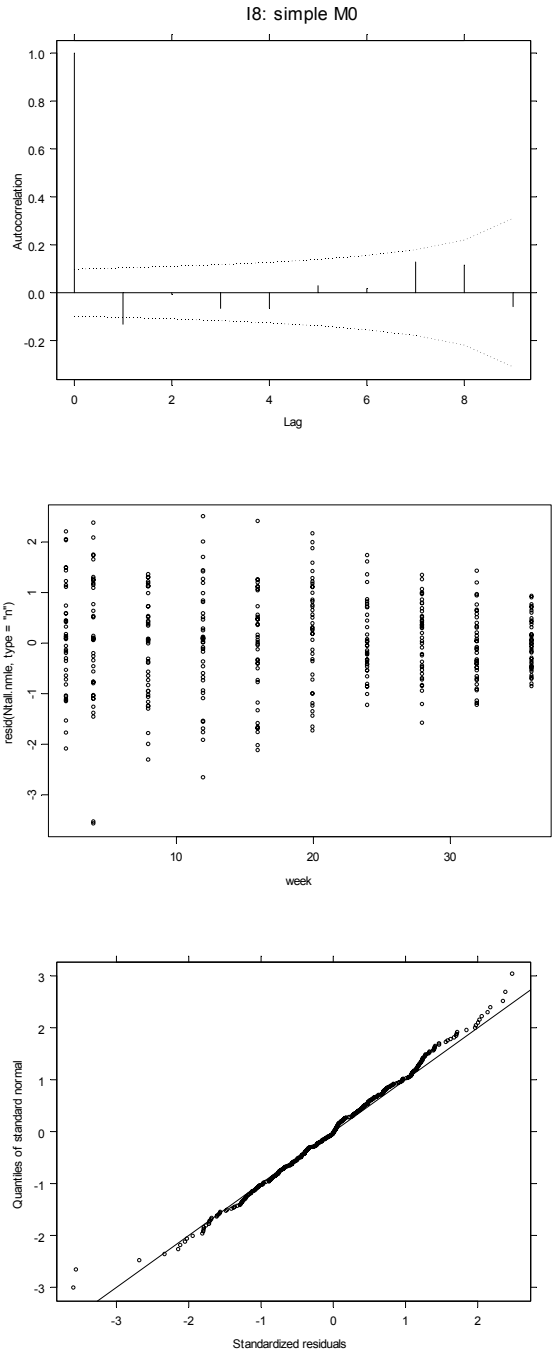

Figure A-7: Incremental scenario 8.

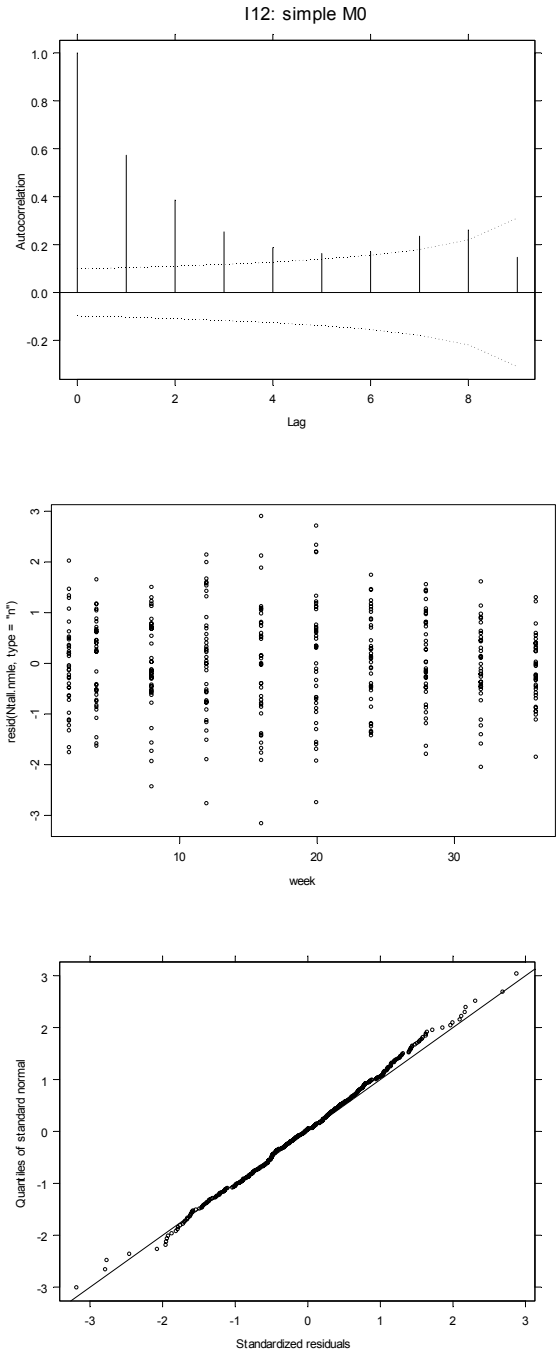

Figure A-8: Incremental scenario 12.