Linearisation of Highly Resolved Substance Transport Models by Temporal Aggregation of Model Outputs

Å. Forsman and **A. Grimvall**

Department of Mathematics, Linköping University, SE-581 83 Linköping, Sweden (asfor@mai.liu.se)

Abstract: Process-oriented models that involve strongly non-linear responses and are driven by highly resolved meteorological inputs are often used to predict substance transport over fairly long periods of time. Therefore, it is of great interest to ascertain whether the non-linear structure is also essential for predicting temporally aggregated model outputs. Here, ordinary least squares regression was employed to identify linear relationships between model outputs and inputs expressed on different temporal scales. Examination of two simple theoretical models showed that aggregation of outputs and inputs improved the linearity of models, provided the models included memory effects. This was confirmed in the analysis of the soil-water and nitrogen transport model *SOIL/SOILN*. Moreover, linear predictors that involved highly resolved inputs proved to be superior to linear predictors based on temporally aggregated inputs. The presence of linear structures facilitates the extrapolation of point models to larger areas, because it allows the use of spatially aggregated inputs to calculate spatially aggregated outputs.

Keywords: Temporal aggregation; Linearity; *SOIL/SOILN* model; Substance transport; Drainage

1. INTRODUCTION

Several of the basic processes that control the transport and transformation of substances in soil are highly non-linear. However, the impact of non-linear processes on model outputs varies with the temporal scale of the model inputs and outputs. Furthermore, process-oriented models are often used to compute total or average outputs over much longer periods than the internal time step of the model. Therefore, we conducted a study to determine under what conditions temporal aggregation of model outputs can improve the linearity of a model.

The presence of linear structures in substance transport models has several important implications. First, impulse-response weights in linear systems are easy to comprehend. Second, the presence of linear structures makes it possible to use spatially aggregated inputs to predict spatially aggregated outputs. Third, incorporation of process-oriented substance transport models in decision support tools is greatly facilitated, if the original model can be substituted by a linear statistical meta-model [Forsman et al., 2002].

The first part of the present study was devoted to clarifying the effect of temporal aggregation on two models with relatively simple structures. Subsequently, we examined the more complex process-oriented *SOIL/SOILN* model, which describes water flow and substance transport in a soil column. Work was focused on drainage in response to precipitation, and nitrogen leaching in response to precipitation and temperature. The linearity of the input-output relationship was evaluated by fitting linear regression models to data at different levels of aggregation.

2. NOTATION AND SIMULATION DESIGN

2.1 Notation

All models we considered were deterministic, i.e., a given vector $\boldsymbol{\omega}$ of inputs and model parameters produced a uniquely determined model output $\boldsymbol{y}$. From a mathematical point of view, all such models can be regarded as functions

$$\mathbf{y} = f(\boldsymbol{\omega})$$

The output $\boldsymbol{y}$ can be a scalar or a vector, whereas $\boldsymbol{\omega}$ is normally a vector with a very large number of components. These components represent all types of inputs and parameters that describe the modelled system. Since we were interested in the temporal dynamics of the models, we distinguished between the components $\mathbf{x}_t$ that depend on time, t, and the rest of the components

$\boldsymbol{\delta}$. Moreover, the output $\mathbf{y_t}$ at time t was assumed to be a function of the inputs at that time and also of s previous inputs

$$\mathbf{y}_t = f(\boldsymbol{\delta}, \mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-s})$$

If $s > 0$, the model was said to have memory effects.

Throughout this study, we examined a single output, y_t, in response to a single time-dependent input, x_t, with a basic time step of one day. Ordinary least squares regression was employed to determine linear relationships between model outputs and inputs. For a model with memory effects, the regression model based on daily inputs and outputs was formulated as

$$y_t = \alpha + \sum_{i=0}^{s} \beta_{t-i} x_{t-i} + \varepsilon_t$$

where α, β_t, ..., β_{t-s} are regression coefficients and ε_t is the residual at time t. Aggregated outputs in response to aggregated or daily inputs were analysed in an analogous manner. In all cases, inputs during the current period and previous periods were considered. The number of previous periods that was included depended on the memory effects in the studied model.

2.2 Model Structures

The previously mentioned *SOIL/SOILN* model comprises features such as a non-linear response to inputs, memory effects and seasonal dependence. To clarify the influence of specific features of the model on the effect of temporal aggregation, we constructed two simple, theoretical models. Both of these describe a non-linear response to independent inputs, and the second model also includes memory effects. The first model was analysed theoretically, whereas Monte Carlo simulations were performed to examine the second model, as well as the *SOIL/SOILN* model.

2.3 Simulation Study

The general principles of the simulation studies are shown in Figure 1. The simulations were first carried out with a time step of one day, after which inputs and outputs were aggregated to monthly and annual values. Scatter plots of predicted and observed (simulated) values and R^2 values were compared to evaluate linearity of the models at different levels of aggregation.

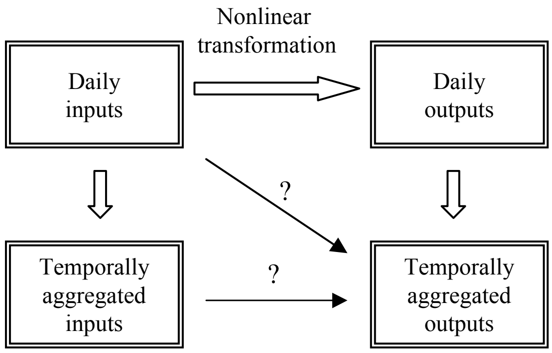

Figure 1. General principles of the simulation studies. Question marks indicate the relationships that were examined.

3. NON-LINEAR RESPONSE WITH NO MEMORY EFFECTS

Let $\{x_t, t = 1, \ldots, n\}$ be a series of independent, identically distributed random variables, and let $y_t = f(x_t)$, where f is a non-linear function. Consider the linear regression model

$$y_t = a + bx_t + \varepsilon_t \qquad (1)$$

where a and b are regression coefficients, and the error terms ε_t are uncorrelated. If $a + bx_t$ is the best linear predictor of y_t, i.e. the predictor that minimises $E(\varepsilon_t^2)$, then x_t and ε_t are also uncorrelated. Now, if $\bar{x}_T$ and $\bar{y}_T$ denote the average inputs and outputs over a period T = [t_1, t_2], it is easily shown that $a + b\bar{x}_T$ is the best linear predictor of $\bar{y}_T$, where

$$\bar{y}_T = \frac{1}{n_T} \sum_{t \in T} y_t = a + b\bar{x}_T + \bar{\varepsilon}_T \qquad (2)$$

Because the variances of the response variables and residuals decrease proportionally between (1) and (2), the R^2 value will not change. Furthermore, since $a + b\bar{x}_T$ is the best linear predictor of $\bar{y}_T$, the linearity of the model can not be improved by replacing aggregated inputs with daily inputs.

4. NON-LINEAR RESPONSE WITH MEMORY EFFECTS

A simulation study was conducted to analyse a simple model that included both a non-linear response and memory effects.

4.1 Data Simulation