Combining Measurement Quality into Monitoring Trends in Foliar Nutrient Concentrations

Mika Sulkava[1], Pasi Rautio[2], and Jaakko Hollmén[1]

[1] Helsinki University of Technology,
Laboratory of Computer and Information Science,
P.O. Box 5400, FI-02015 HUT, Finland
{Mika.Sulkava, Jaakko.Hollmen}@hut.fi
[2] The Finnish Forest Research Institute, Parkano Research Station,
Kaironiementie 54, FI-39700 Parkano, Finland
Pasi.Rautio@metla.fi

Abstract. Quality of measurements is an important factor affecting the reliability of analyses in environmental sciences. In this paper we combine foliar measurement data from Finland and results of multiple measurement quality tests from different sources in order to study the effect of measurement quality on the reliability of foliar nutrient analysis. In particular, we study the use of weighted linear regression models in detecting trends in foliar time series data and show that the development of measurement quality has a clear effect on the significance of results.

1 Introduction

Analyzing chemical characteristics in samples collected from different components of ecosystems (e.g. biological samples, soil, water, etc.) are key methods in environmental monitoring. Chemical analyses are, however, prone to many errors which has brought up concern about the reliability of the analyses. Great improvements in laboratory quality have been achieved in the past two decades due to, for example, use of international reference material and interlaboratory comparisons (so-called ring tests). Despite the general improvement in laboratory quality some of the latest ring tests still reveal problems in quality [1].

The aim of this paper is to study how large of an impact laboratory quality has on detecting changes in environment. We use data from the ring tests, where laboratories analyzing foliar samples have been surveyed (see e.g. [1]). Further we link the data of conifer foliar concentrations measured in samples collected from 36 Finnish ICP Forests Level I plots (International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests, see http://www.icp-forests.org/) and the results obtained by the Finnish laboratory in internal quality tests and the above mentioned ring tests.

Both theoretical computations and real-world data were used to study the effect of changing data quality on trend detection. Foliar nutrient data from Finland were analyzed using weighted regression. In our previous research the

use of sparse linear models for finding other linear dependencies in the data have been briefly discussed in [6].

2 Data

2.1 Foliar Nutrient Data

Annual nutrient concentration data of conifer needles (Norway spruce [Picea abies (L.) Karsten] and Scots pine [Pinus sylvestris L.]) collected from 36 Finnish ICP Forests Level I stands were available for years 1987–2002. Foliage from 20 pine stands and 16 spruce stands around the country were collected yearly in October or November. Concentrations of 12 elements were measured from the needles, but in this study we focus on two elements: nitrogen (N) and sulfur (S). For details concerning the sampling procedure, see [5]. A more comprehensive characterization of the data using nutrition profiles is presented in [3].

2.2 Laboratory Quality Data

The quality of measurements of the laboratory analyzing ICP Forests foliar samples in Finland was studied in national calibration tests and international interlaboratory tests. The test data can be used to estimate the accuracy and precision of the nutrient measurement data. Between 1987 and 1994 the quality of measurements was surveyed in laboratory comparisons arranged by IUFRO (International Union of Forest Research Organizations). Since 1993 the measurement quality was tested in seven ICP Forests biennial ring tests (see e.g. [1]). In addition to interlaboratory tests, starting from 1995 the quality of the Finnish laboratory was measured in repeated measurements of certified reference samples (CRM 101). Before 1995 the methods were more varied. The quality control of the laboratory is discussed in more detail in [3].

3 Methods

3.1 Weighted Regression

If the precision of the observations is not constant, fitting an ordinary least squares linear regression model in order to analyze the data is not well justified, because homoscedasticity[1] is one of the basic assumptions of the model. Instead, weighted regression [4] is an effective method with heteroscedastic data. The regression model with heteroscedastic data can be expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n \qquad (1)$$

where β_0 and β_1 are regression coefficients ($\boldsymbol{\beta} = [\beta_0 \ \beta_1]^T$), X_i are known constants and error terms ϵ_i are independent $N(0, \sigma_i^2)$. If X_i denote time steps, the model assumes that there is a linear trend in the time series data.

---

[1] Homoscedasticity = property of having equal variances.

The weight w_i is defined as the inverse of the noise variance and thus, the method gives weights to observations according to their uncertainty

$$w_i = \frac{1}{\sigma_i^2}. \tag{2}$$

For example, completely uncertain ($\sigma_i = \infty$) measurements are eliminated from the model. $\mathbf{Y}$ and $\mathbf{X}$ the dependent and independent variables expressed in vector and matrix terms and $\mathbf{W}$ is a diagonal matrix containing the weights w_i

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \tag{3}$$

The maximum likelihood estimators of the regression coefficients are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \tag{4}$$

The statistical significance of $\hat{\beta}_j \neq 0$ can be evaluated using the F-test [2].

3.2 Parameter Estimation

The error in the measurement of laboratory j in year i is assumed to be normally distributed with standard deviation (precision) σ_{ij} and mean (accuracy) μ_{ij}. The ICP Forests ring tests and a part of the IUFRO tests contained repeated measurements of the same sample. This makes it possible to estimate both the accuracy and precision of a tested laboratory. The estimated accuracy is the average deviation of n_{ij} repetitions from the average of all laboratories

$$\hat{\mu}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Z_{ijk} - \hat{\mu}_i, \tag{5}$$

where Z_{ijk} is the value of kth repetitive measurement and

$$\hat{\mu}_i = \frac{\sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} Z_{ijk}}{\sum_{j=1}^{m_i} n_{ij}}. \tag{6}$$

Above, m_i is the number of laboratories with acceptable results (i.e. laboratories fulfilling the quality requirements set by ICP Forests [1]). The estimate of the precision of a laboratory is the unbiased estimate for the standard deviation

$$\hat{\sigma}_{ij} = \sqrt{\frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (Z_{ijk} - \hat{\mu}_{ij})^2}. \tag{7}$$

In case there are no repetitions, i.e. $n_{ij} = 1$, the precision is estimated to be the standard deviation of all acceptable laboratories

$$\hat{\sigma}_{ij} = \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (Z_{ij1} - \hat{\mu}_i)^2}. \qquad (8)$$

The minimum relative standard deviation (RSD) of the methods was determined to be 0.7% for N and 1.5% for S between 1987 and 2000. The precision estimated using Equation 7 may be lower than these minimum values, because n_{ij} is too small to make reliable estimates. If this is the case, the minimum RSD is used instead of the estimated value.

4 Experiments

4.1 Theoretical Computations

The effect of measurement quality (i.e. accuracy and precision of measurements) on trend detection was studied. Here we assume that there is a linear trend (see Equation 1) in the time series data (e.g. decreasing foliar sulfur concentrations in the course of time) and that this trend will also continue in the future.

The development of quality in different laboratories was visually inspected. According to that two simple scenarios roughly corresponding to typical development of real measurement precision were constructed. Either the precision does not change with time or the precision changes linearly from initial precision level c in a time steps to level b and then stays constant

$$\sigma_i = \begin{cases} \frac{b-c}{a} X_i + c & \text{if } X_i \leq a \\ b & \text{if } X_i > a \end{cases} \qquad (9)$$

Trend detection with weighted regression was studied using the scenarios explained above. The hypothesis H_1 that there exists either an increasing or a decreasing trend in the data was tested against the null hypothesis H_0 that there is no trend. That is, H_0: $\beta_1 = 0$, H_1: $\beta_1 \neq 0$. The different parameter values, i.e. a, b, c, β_1, and n were varied and the p-value was calculated using the F-test. Significance level 0.05 was used to reject the null hypothesis.

The results for linearly changing precision are shown in Figure 1. The time needed to detect a trend with improvement in precision ($b < c$) can be seen in Figure 1 above the diagonals of the subfigures. The initial precision c, final precision b, and parameter a notably affect the time needed for detecting a trend. For example, when we look at the subfigure in the second column and top row in Figure 1, we can see (in the center of the subfigure) that the time needed to detect a trend with slope $\beta_1 = 0.1$ is greater than 10, if the precision is constant 0.5. However, if the final precision b is improved to 0.25 or 0.05 in two time steps ($a = 2$), the time n needed to detect the trend decreases to $8 < n \leq 10$ or $4 < n \leq 6$, respectively. We also studied the effect of an exponential change in precision, and found that the results are very similar to linear change in precision.
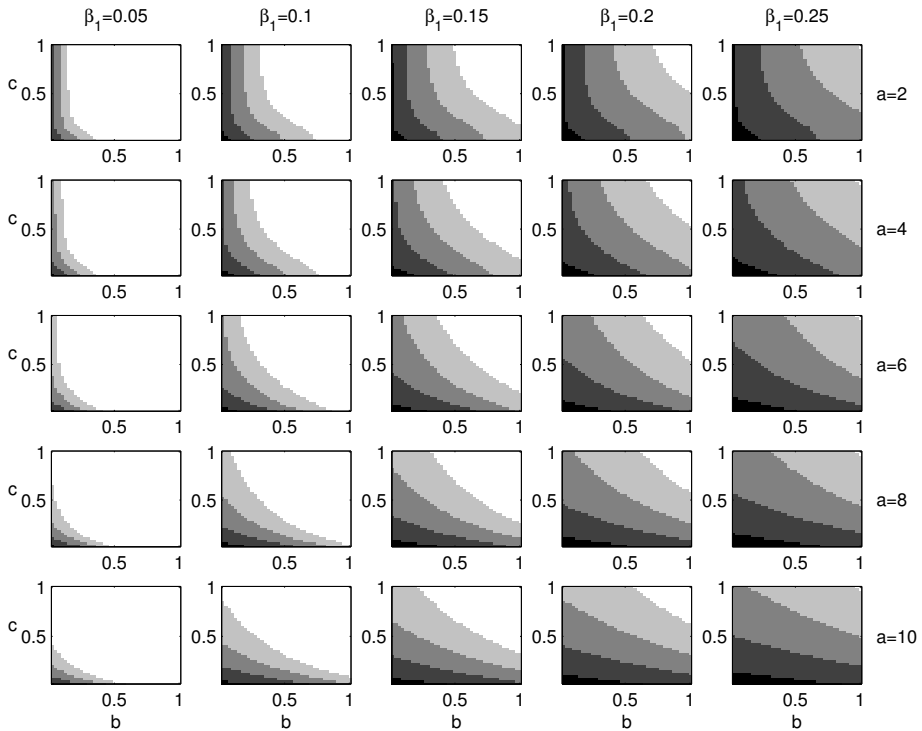
Fig. 1. Trend detection for linearly changing precision with different parameter values. The colors indicate the length n of time series that is needed to detect a trend with significance level 0.05. Black: $n \leq 4$, dark gray: $4 < n \leq 6$, medium gray: $6 < n \leq 8$, light gray: $8 < n \leq 10$, white: $n > 10$. In a subfigure the abscissa represents the final precision b and ordinate the initial precision c. Different columns of subfigures ($\beta_1 = 0.05, \ldots, 0.25$) represent the slope of the trend. Different rows of subfigures ($a = 2, \ldots, 10$) represent the speed of change in measurement precision.

4.2   Trends in Measurement Data

Quality data from all different sources was combined and the accuracy and precision of the Finnish laboratory was estimated for years 1988–2003 using methods described in Section 3.2. Measuring of the needles is done the year after the sample collection and therefore, quality data of year $t + 1$ was used for needles of year t. Because weighted regression requires that the noise in the measurements has zero mean, the accuracies μ_{ij} were subtracted from the measurements.

First, the average N and S foliar concentration data of pine between 1987–2002 were studied. Using weighted regression, a significant ($p < 0.05$) decreasing trend was found in S concentration and an insignificant ($p \geq 0.05$) weakly increasing trend in N concentration. The trend in average N concentration would become significant in three years if the data quality stays the same as in year

2003. The trend could be detected the following year, if the precision would be improved to $b = 0.06$.

It was also studied if there are significant trends in N and S concentrations in pine needles collected from the 20 individual pine stands between years 1987 and 2000. Out of 20 time series in 3 a significant trend was found in the N data and in 7 in the S data. In both cases there were three time series, where too many missing values made fitting a regression model unsubstantial.

We also experimented how long we would have to continue measuring in the stands, where a significant trend was not found, assuming that the trend continues, to be able to tell that the trend is significant. If the precision of measurements stays the same as in year 2001, the trends would become significant in 1–146 years depending on the estimated slope and number of measurements. However, if the precision of measurements is improved to approximately 0.5% RSD, the time needed decreases clearly. If the standard deviation of N measurements decreases linearly in $a = 3$ years to value $b = 0.06$, the time needed to detect a trend decreases on average 46%. Similarly, if the standard deviation of S measurements decreases linearly in $a = 3$ years to value $b = 0.005$, the time needed to detect a trend decreases on average 15%.

5   Conclusions

The results show that measurement precision strongly affects trend detection. Improving data quality can decrease clearly the time needed for finding statistically significant trends in environmental monitoring. Even though the Finnish laboratory analyzing the foliar samples has always fulfilled the quality demanded by the ICP Forests programme for both nitrogen (less than 10% deviation from the mean values of all labs) and sulfur (less than 20% deviation) it can still take many years to detect a possible ongoing trend with this measurement precision. For a laboratory not meeting the criteria set by ICP Forest programme (e.g. showing deviation greater than 20% in case of sulfur) it can take years or even decades to detect possible ongoing changes in the state of the environment. In all our results from theoretical computations and real world data clearly highlight the importance of quality in laboratory analyses.

Acknowledgements

References

1. Alfred Fürst. 7th needle/leaf interlaboratory comparison test 2004/2005. Technical report, United Nations Economic Commission for Europe, European Union, 2005.
2. Bent Jørgensen. *The theory of linear models*. Chapman & Hall, 1993.

3. Sebastiaan Luyssaert, Mika Sulkava, Hannu Raitio, and Jaakko Hollmén. Evaluation of forest nutrition based on large-scale foliar surveys: are nutrition profiles the way of the future? *Journal of Environmental Monitoring*, 6(2):160–167, February 2004.

4. John Neter, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*. McGraw-Hill, 4th edition, 1996.

5. Klaus Stefan, Alfred Fürst, Robert Hacker, and Ulrich Bartels. Forest foliar condition in Europe - results of large-scale foliar chemistry surveys 1995. Technical report, European Commission, United Nations Economic Commission for Europe, Brussels, Geneva, 1997.

6. Mika Sulkava, Jarkko Tikka, and Jaakko Hollmén. Sparse regression for analyzing the development of foliar nutrient concentrations in coniferous trees. In Sašo Džeroski, Bernard Ženko, and Marko Debeljak, editors, *Proceedings of the Fourth International Workshop on Environmental Applications of Machine Learning (EAML 2004)*, pages 57–58, Bled, Slovenia, September/October 2004.