

Detection of the change point in the winter wheat experiment

Anna Szczepańska, Ewa Bakinowska

Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, e-mail: sanna6@wp.pl

SUMMARY

This paper concerns detection of the change point, which is treated as an abrupt change in the response function or one of its derivatives. The change point is identified using the semiparametric model and the theory given by Speckman (1994). The theory is illustrated by a real experiment in which the dry biomass of winter wheat is studied.

Key words: biomass of winter wheat, change point, kernel function, semi-parametric model, smoother matrix

1. Introduction

The change point occurring in various processes has been considered in many research experiments. Horvath and Kokoszka (1997) and Gijbels and Gordeniaux (2004) detected the climatic changes that took place over a span of several years or decades using nonparametric regression. Muller (1992) estimated change points in data on the annual flow volume of the Nile river between 1871 and 1970. In the present paper the change point of dry biomass accumulation of winter wheat is determined, with analysis based on the theory given by Speckman (1994). The accumulation of dry biomass in the process of vegetation provides the economic value of cereals (Evans and Wardlaw 1976, Reiner 1992, Spiertz and Vos 1985). The intensity of this process and the final effect, as grain yield, are determined by the availability of water and minerals (Dzierżyc and Donev 1989, Kleemola 1998). The important factors influencing cereal crops are minerals which are taken from the soil during plant growth. Sometimes the soil is not rich in these substance, so appropriate supplements are used to stimulate the growth of plants. The minerals which most effectively influence on the plant and its development are nitrogen, phosphorus and potassium.

The one most responsible for the dynamic of plant biomass is nitrogen and to achieve an optimal nitrogen fertilizer, the optimum fertilizer dose and the proper time for its application should be known.

Of interest to the practitioner are the detection of the time point at which an intensive change of growth is observed, and study of the influence of dose of nitrogen on changes in dry biomass.

2. Model and method

The change point is a point at which there appears a discontinuity or an abrupt change in the mean response function or in one of its derivatives. The model of an experiment where such a point is detected can be considered as a semiparametric model of the form

$$\mathbf{y} = \mathbf{f} + \beta \boldsymbol{\phi}_{k\tau} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{y} \in R^n$ is the vector of observations, $\mathbf{f} = [f(t_1), f(t_2), \dots, f(t_n)]^T$ is an unknown smooth function with at least k derivatives and $t_i, i = 1, 2, \dots, n$ are equally spaced time points of the form $t_i = \frac{(2i-1)}{2n}$. The vector $\boldsymbol{\phi}_{k\tau} = [\frac{(t_1-\tau)^{k-1}}{(k-1)!}, \frac{(t_2-\tau)^{k-1}}{(k-1)!}, \dots, \frac{(t_n-\tau)^{k-1}}{(k-1)!}]^T$, where $t_i > 0$, depends on the change point $\tau \in (0, 1)$. If $t_i \leq 0$ then the i th element of vector $\boldsymbol{\phi}_{k\tau}$ is equal to zero. Moreover the unknown parameter β is the size of the change in the $(k-1)$ th derivative. The vector $\boldsymbol{\epsilon} \in R^n$ is the vector of errors, which are independent and identically distributed, $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$, where the matrix \mathbf{I}_n is the $n \times n$ identity matrix.

To detect the change point in model (1) the null hypothesis $H_0 : \beta = 0$ is tested against the alternative hypothesis $H_1 : \beta \neq 0$. From Speckman (1994) the test statistic is

$$T_n(\tau) = \frac{\hat{\beta}(\tau)}{\sqrt{Var(\hat{\beta}(\tau))}}, \quad \tau \in (h, 1-h), \quad (2)$$

where h is the bandwidth.

Observe that from Eubank and Speckman (1994) the estimator of parameter β depends on the change point τ and can be written as

$$\hat{\beta}(\tau) = \frac{\boldsymbol{\phi}_{k\tau}^T (\mathbf{I} - \mathbf{S})^2 \mathbf{y}}{\boldsymbol{\phi}_{k\tau}^T (\mathbf{I} - \mathbf{S})^2 \boldsymbol{\phi}_{k\tau}},$$

where the matrix $\mathbf{S} = (s_{ij})$ is the smoother matrix with

$$s_{ij} = \frac{1}{nh} \left[K \left(\frac{t_i - t_j}{h} \right) \right], i, j = 1, 2, \dots, n \quad (3)$$

and $K(\cdot)$ is the kernel function.

The variance of the estimator $\hat{\beta}(\tau)$ has the form

$$\text{Var}(\hat{\beta}(\tau)) = \sigma^2 \mathbf{a}^T \mathbf{a}, \quad \mathbf{a}^T = \frac{\boldsymbol{\phi}_{k\tau}^T (\mathbf{I} - \mathbf{S})^2}{\boldsymbol{\phi}_{k\tau}^T (\mathbf{I} - \mathbf{S})^2 \boldsymbol{\phi}_{k\tau}},$$

and from Przystalski and Krajewski (2007) the estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n - \text{tr}\mathbf{H}} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

where $\mathbf{H} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\boldsymbol{\phi}_{k\tau}\mathbf{a}^T$ and $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]^T$.

It is stated in Speckman (1994) that based on the test statistic (2) the change point is detected if and only if $\sup_{\tau} |T_n(\tau)| > c_{\alpha}$, where c_{α} is the critical value and $\tau \in [2h, 1 - 2h]$.

3. Analysis of the experiment

3.1. Material

Using the theory of Section 2, an analysis was performed on data from an experiment conducted in 2003 at the Brody experimental farm of Poznań Agricultural University. The experiment was planned in a randomized complete block design (Trętowski and Wójcik, 1988) and experimental plots of 20.25 m² were used. The growth in the dry biomass of winter wheat was measured on 1 meter length of a plot. The observations consist of measurements performed in four replications at nine time points with equal weekly intervals, starting from the 45th day of vegetation. In the experiment four doses of nitrogen were applied: 0 kg/ha, 60 kg/ha, 120 kg/ha and 180 kg/ha. Here these are denoted by Dose 0, Dose 60, Dose 120 and Dose 180. A graphical presentation of the data from the experiment is given in the figure below.

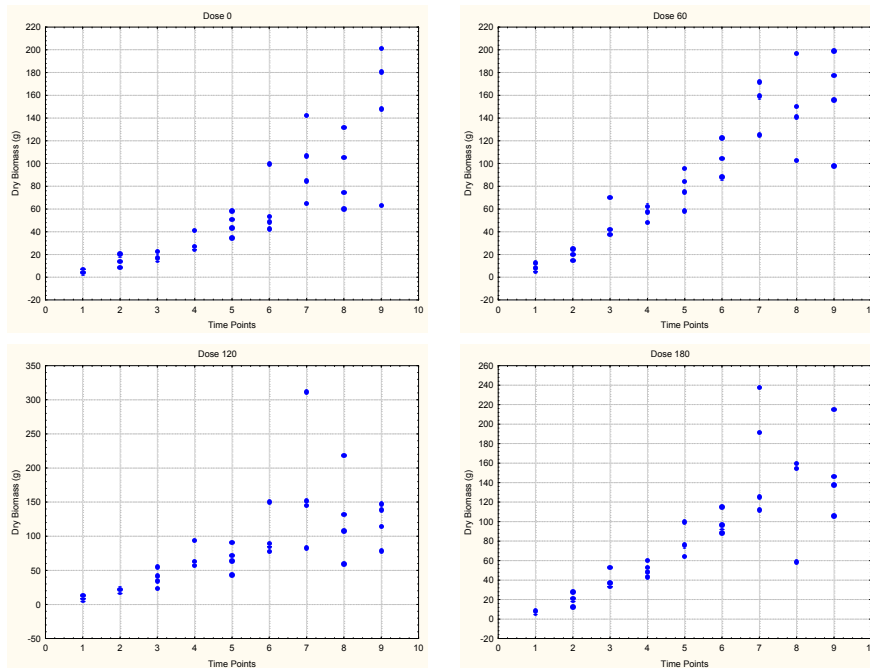


Figure 1. Data on dry biomass of winter wheat.

The analysis of the experiment is based on means (compare Speckman, 1994) calculated for each of the nine time points.

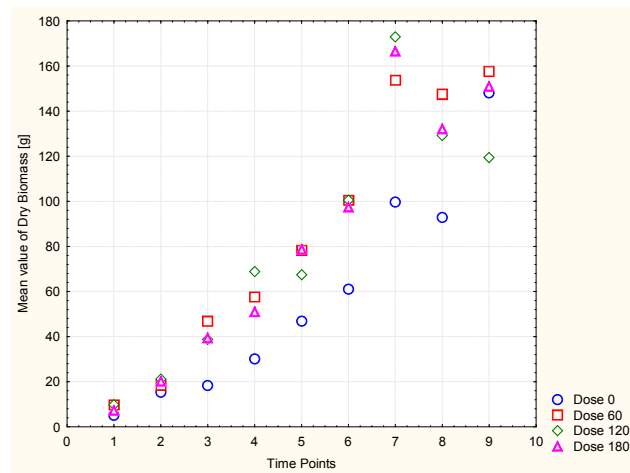


Figure 2. Data on mean value of dry biomass of winter wheat.

3.2. Numerical calculations

In this work an attempt was made to find the change point in the first derivative of the mean response function which illustrated the rapidity of change in dry biomass of winter wheat. Thus in the calculations we took $k = 1$. From Section 2 one can observe that the test statistic given in (2) indirectly depends on the kernel function and the bandwidth. The choice of kernel function and bandwidth and their influence on the smoothness and the shape of \hat{y} have been studied in the literature. It is stated by Härdle et al (2004) that using the same value of h and different kernel functions, the estimated curves do not have the same degree of smoothness. This inconvenience can be avoided by multiplying one of the bandwidths by an adjustment factor. Moreover, they observed that for practical purposes the choice of the kernel function is almost irrelevant for the efficiency of the estimate. Thus for the analysis of the experiment we choose the Epanechnikov kernel of the form

$$K\left(\frac{t_i - t_j}{h}\right) = \begin{cases} \frac{3}{4}\left(1 - \left(\frac{t_i - t_j}{h}\right)^2\right) & |t_i - t_j| \leq h \\ 0, & |t_i - t_j| > h \end{cases}.$$

In contrast to the kernel function, the value of bandwidth, h , has a large influence on the estimation. The best bandwidth can be found by applying optimality criteria. One method given in the literature selecting the optimal bandwidth is cross-validation (see Härdle et al (2004), Kayri and Zirhlioglu, (2009)). It is stated by Härdle et al (2004) that the chosen bandwidth in this method automatically adapts to the smoothness of the function. The cross-validation criterion involves minimizing, with respect to h , the following function

$$CV(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K * K\left(\frac{t_i - t_j}{h}\right) - \frac{2}{nh(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K\left(\frac{t_i - t_j}{h}\right), \quad (4)$$

where $K * K(u) = \int K(u-v)K(v)dv$.

Using the above formula we find that the optimal bandwidth for the data under consideration is $h = 0.7$. Unfortunately we cannot consider this value for the analysis of experiment, because detection of the change point is based on the interval $[2h, 1 - 2h]$, and for $h = 0.7$ the theory given in Section 2 cannot be used. For our calculations, based on the cross-validation

criterion we select the best value of bandwidth from the interval $[0.1, 0.24]$. In the considered interval the bandwidth which minimizes (4) is $h = 0.22$. The change point is detected by comparing the maximum of the absolute value of statistic (2) for $\tau \in [2h, 1 - 2h]$ with the critical values, which were calculated from Theorem 2 in Speckman (1994) as $c_{0.05} = 2.54$ and $c_{0.1} = 2.77$, where $c_\alpha = z_{\alpha/2n}$, $z_\alpha = \phi^{-1}(1 - \alpha)$ and $\phi(z)$ is the cumulative distribution function of the standard normal distribution.

All calculations were performed using the program Mathematica 7.0.

3.3. Results

On graphs, given below, the absolute value of test statistic (2) for the given doses of nitrogen is shown.

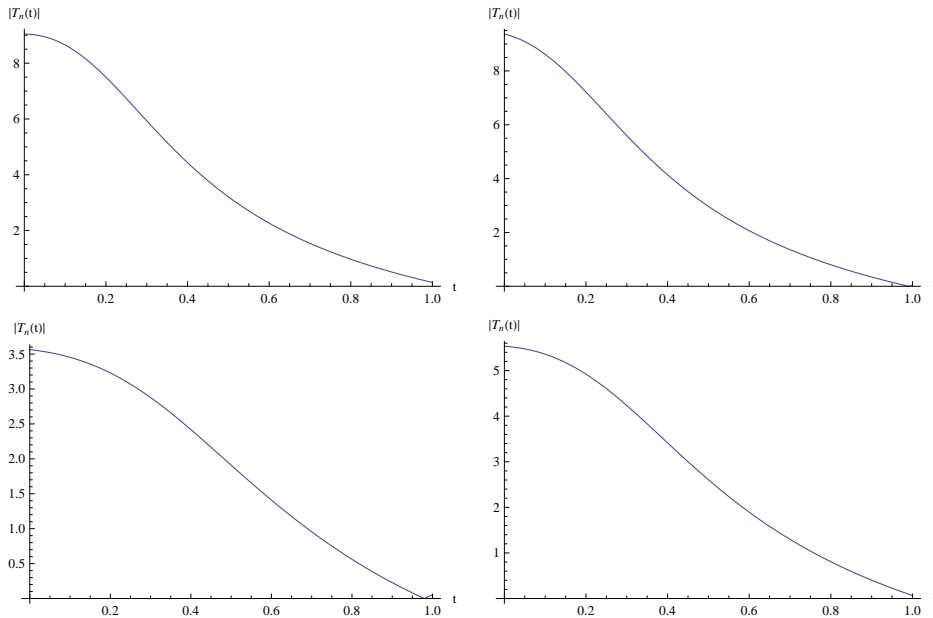


Figure 3. Test statistics.

Observe that the above functions are decreasing and have almost the same shape. Moreover, the maximum of the absolute value of statistic (2) on the interval $[0.44, 0.56]$ for $h = 0.22$, which is larger than the critical values (5), determines the change point. In the table below we present the values of the change point obtained for the different doses.

	Dose 0		Dose 60		Dose 120		Dose 180	
h	$\sup_{\tau} T_{\tau}(n) $	t_{chp}	$\sup_{\tau} T_{\tau}(n) $	t_{chp}	$\sup_{\tau} T_{\tau}(n) $	t_{chp}	$\sup_{\tau} T_{\tau}(n) $	t_{chp}
0.22	3.90535	4.46	3.63632	4.46	2.21761	-	3.08239	4.46

Here t_{chp} denotes the time point at which the change point is located.

From the above table it can be observed that for three of the nitrogen doses (Dose 0, Dose 60 and Dose 180) the change point appears at time point $4.46 \approx 5$ while for Dose 120 there is no the change point. Further numerical results for Dose 120 show that for $h < 0.14$ the shape of the test statistic is changed and a change point for Dose 120 exists between time points 2 and 3, for example $(h = 0.1; t_{chp} = 2.3)$, $(h = 0.11; t_{chp} = 2.48)$, $(h = 0.12; t_{chp} = 2.66)$, $(h = 0.13; t_{chp} = 2.84)$. Unfortunately these values of h are not optimal and do not give the best estimation of the smooth function. On analysis of the whole data set using the bandwidth $h = 0.22$, the applied test does not detect a change point for Dose 120.

4. Conclusions

In this paper the theory of Speckman (1994) was used to detect the change point in an experiment with winter wheat. Based on the Epanechnikov kernel and a bandwidth equal to 0.22, the change points in the rate of change of dry biomass were determined. From the data for the three nitrogen doses of 0 kg/ha, 60 kg/ha and 180 kg/ha, it was found that the change point occurs at about 70th day of growth. However, for doses of 120 kg/ha, there was no change point in the growth of dry biomass. Based on the data (Figure 1) it can be seen that for all nine time points the nitrogen has an impact on the average amount of dry biomass. However, the above analysis of the experiment shows that the dose of nitrogen does not significantly affect the change point of increase in dry biomass. This paper may provide an illustration of the theory for semiparametric models, and of the concept of change point, through their application to the analysis of a real agricultural experiment.

Acknowledgments

We are grateful to Dr Marta Molińska-Glura for making available data from the experiment measuring dry biomass of winter wheat. (M. Molińska-Glura, 2007, *Examination of the relative growth rate of cereal on the example of wheat and triticale (in Polish)*, PhD dissertation, The Main Library at Poznań University of Life Sciences).

REFERENCES

- Dzieżyc J., Donev A.N. (1989): Water needs of crops. Collective work, (in Polish) PWN, Warszawa.
- Eubank R.L., Speckman P.L. (1994): Nonparametric estimation of functions with jump discontinuities. IMS Lecture Notes - Monograph Series vol. 23.
- Evans J., Wadrow J. (1976): Aspects of the comparative physiology of grain yielding cereals. *Adv. Agron.* 28: 301–358.
- Gijbels I., Goderniaux A.C. (2004): Bandwidth Selection for Change-point Estimation in Nonparametric Regression. *Technometrics* 46(1): 76–86.
- Horvát I., Kokoszka P. (1997): Change-Point Detection With Nonparametric Regression. Technical Report University of Liverpool.
- Härdle W., Müller M., Sperlic, S., Werwatz A. (2004): Nonparametric and Semiparametric Models. Springer.
- Kayri M., Zirhlioglu G. (2009): Kernel Smoothing Function and Choosing Bandwidth for Non-Parametric Regression Methods. *Ocean Journal of Applied Sciences* 2(1).
- Kleemola J. (1998): Modeling crop growth and biomass partitioning to shoots and roots in relation to nitrogen and water availability, using a maximization principle, II. *European Journal of Agronomy* 8(3-4): 191–204.
- Müller H.G., (1992): Change-Points in Nonparametric Regression Analysis. *Ann. Statist.* 20(2): 737–761.
- Przystalski M., Krajewski P. (2007): Constrained estimators of treatment parameters in semiparametric models. *Statistics & Probability Letters* 77: 914–919.
- Reiner L. (1992): Weizen aktuell. DLG-Verlag, Frankfurt am Main.
- Speckman P.L. (1994): Detection of change-point in nonparametric regression. unpublished manuscript.
- Spiertz J., Vos J. (1985): Grain growth of wheat and its limitation by carbohydrate and nitrogen supply, *Wheat growth and modeling*. Plenum Press, New York. 129-141.
- Trętowski J., Wójcik A.R. (1988): Methodology of agricultural experiments (in Polish). WSRP Siedlce.