THE EXTENSION OF TIME SERIES ANALYSIS OPPORTUNITIES AND ITS APPLICABILITY IN DECISION SUPPORT OF FORESTRY AND WOOD SCIENCES

Ph.D. Dissertation Summary

by

Zoltán Pödör

Institute of Informatics and Economics

The Simonyi Károly Faculty of Engineering, Wood Sciences and Applied Arts
University of West Hungary

Research Supervisor:

Dr. László Jereb

University of West Hungary

Sopron, Hungary 2014.

1. Motivations and Purposes

The decision support systems related to various professional fields, with forestry [27, 28] among them, help make scientifically based professional decisions. A decision support system offers innumerable approaches to a problem as well as tools and methodologies for the solution of the given problem, among other things, the possibilities to study time series, and to identify and analyse their interrelations.

There are many fields in which the time sequence of collected data holds significance. A special subset of time series comprises periodic time series that possess some natural or artificial cycles. Such time series are generated by forestry research in large numbers and can typically be related to different calendar-based periods. Forestry gives rise to a whole range of tasks and problems, which, when analysed, allowed for the identification of three major groups of problems as follows,

- the study of the systems of relationships between dependent and independent variables (the fundamental problem was defined by systems of relationships between growth and climate data),
- the issue of the time wise stability of independent parameters (typically of climatic components),
- making discrete growth data continuous by fitting a curve to them, and their further application in preparing data series and in analytical problems.

Within the scope of the dissertation, a general methodological solution and an analytical process were worked out for the solution of the above-mentioned problems, which significantly widen the scope of the studies that can be completed, and address these issues in a uniform way. The applicability of the method was demonstrated in terms of forestry problems. However, the applicability of the method is not limited to the forestry problems that had originally motivated its development. The procedures developed can also be applied in any other field where the analysis of relationships between periodic time series is of fundamental importance.

2. Research Methodology

The study of time series and the analysis of their interrelations are highly important and major areas of statistics and data mining, and therefore the results presented herein may be related to these professional fields. The basic aims in analysing relationships are to define the strength of the relationships between variables and to define functions to describe these interrelations. The first type of relationship analysis is correlation analysis and the latter is regression analysis. Apart from the two basic methods, obviously innumerable other techniques to detect relationships are known, and even the referenced basic methods themselves have many special versions (rank correlation, linear-, and nonlinear regression) [30]. Forestry literature suggests that other often applied methods include principal component analysis [24], response function analysis [23] and, with respect to grouping data, cluster analysis [18]. Bootstrap and jackknife methods [22] are applied to improve the output stability of the above-mentioned procedures used for identifying interrelations. The evolutionary and the moving interval techniques [19, 20] allow for examining the changes in relationships over time. These are generally used with some analytical methods.

The whole data series may not necessarily be homogenous for time series of an adequate length, there may be significant erratic changes in it. The point at which the time series is divided in two and differences between the data in the two subintervals can also statistically be confirmed in terms of one of their properties, e.g. average, deviation etc. can be defined as a breakpoint. Such relationships between the time series, exactly because of the breakpoints detected in the dependent or independent variable, may not therefore be considered stabile over the whole studied period. Several statistical procedures are known to detect breakpoints in time series [25].

The data used in the analyses are usually discrete data series that are often inaccurate and noisy. Further problems may be generated by lack of data, outliers or defective data. By fitting the right function to the data, these problems can be resolved since, by making the data line continuous, we can filter noisy data, missing data can be supplemented for by good approximation, and obvious data defects can be detected and repaired. The process of curve-fitting is mathematically a multivariable extreme value problem and in order to solve this problem, we must define the approximate solutions of a set of generally non-linear multivariable equations [31]. The task is made even more difficult by the fact that the curve to be fitted must be defined before starting the procedure, in the light of the basic data. This problem, as a forestry issue, arose in connection with the intra annual growth data of trees. This also narrowed down the range of applicable functions to S growth curves with one inflection point [29, 32]. Further curve parameters can be generated in an objective and mathematically well-defined way through the application of fitted functions. The Gauss-Newton method was used to fit the curves, and which model can be considered better for the various fittings was decided in consideration of the Akaike information criterion and the modified coefficient of determination [21].

The analyses completed as part of the research and our own methods were implemented in the free R software environment for statistical computing.

3. New Results

The new results described in the dissertation relate to two main topics that also define the two sets of theses. Chapter 2 of the dissertation presents the newly developed and created analytical method and its applicability through forestry examples while the procedures in Chapter 3 can be considered as special preparatory steps to the analytical process relying on the CReMIT method. In Chapter 3.1 studies of breakpoints and the related results and findings are presented while Chapter 3.2 describes the research and valuations regarding curves that can be fitted to growth data series. Both studies can be considered as organic parts of the uniform analytical process developed in connection with the CReMIT method, which also expand the range of parameters available for the CReMIT method, and thereby the range of possible studies to be realised. The results can be categorised into the following sets of theses:

1. Thesis: CReMIT method

The study of forestry literature made the range of relevant analytical methods to investigate the relationships between the growth of trees as dependent and climatic parameters as independent variables circumscribable [4, 5]. It could be ascertained that the analysis of the effects of the various time-shifted and delayed intervals is highly important. The methods learnt, however, are not adequate for complete and systematic analysis since the studied periods so far were primarily set by observation and experience, and the aspiration for completeness was missing. The applied methods usually look for linear correlations and deterministic approach to time series is typical. Experiences show that set of theses no. 1 defines results that are able to make the study of periodic time series deeper and more complex irrespective of the analytical methods applied.

- 1.1 Thesis: Derived from the periodicity of the basic data, the CReMIT (Cyclic Reverse Moving Interval Techniques) method [5, 7, 15, 17], applied on a time series as independent variable, was developed to generate windows of various width and time shifting within optional parameter limits (Chapter 2.1.1).
- 1.2 Thesis: Over the windows, transformed time series can be interpreted that can identify relationships systematically and in a wider scale than before [5, 7, 15, 17], (Chapter 2.1.2).
- 1.3 Thesis: This procedure elevates the opportunities granted by evolutionary and moving intervals techniques to a higher level of abstraction [17] (Chapter 2.1.3).

- 1.4 Thesis: The extension of the basic CReMIT method to several variables was developed and presented. Such an extension involves the application of the basic method for the dependent or several independent variables, thereby further widening the range of variables that can become part of the analysis and therefore the range of possible analytical processes [17] (Chapter 2.2).

2. Thesis: The support of CReMIT application

The CReMIT method was applied as part of an analytical process in order to ensure its application in practice. The process consists of three main parts, the data preparation, the CReMIT method and the analytical modules. After data preparation is completed, the data enter the second transformation module that executes the CReMIT method. The third, the analytical module of the shell receives the derived time series and executes the pre-defined analytical process. The derived time series generated by the CReMIT module can be used in any analytical process, which ensures a high level of flexibility.

The first module in the process, besides preparing the data in the traditional sense of the word, prepares the time series to be studied also in different ways, such as detecting breakpoints and fitting the right curves to the data series. Those latter expand the range of parameters transferable to the CReMIT module, and consequently, the range of possible analytical processes.

Within the procedures defining the data of the time series such factors may appear that can cause abrupt and statistically demonstrable changes in the points of the data series. In the case of long time series, this questions the time-wise stability of the data series, which is especially important in studies of parameters describing climatic features (precipitation, temperature). Although several papers deal with the study of abrupt changes in environmental parameters [25, 26], the evaluation of the completed analyses confirmed that this issue was generally not addressed properly in relationship analyses. In order to avoid unfounded conclusions and fine tune analytical results we therefore paid attention also to detecting abrupt changes in the data series in our studies.

The other preparatory step involves fitting a matching curve to growth data series. It allows for the execution of several procedures as well that form the part of traditional data preparation, such as data purification, noise reduction etc. The growth process of trees per annum can be characterised by sigmoid or S-curves. This family of curves with one inflection point includes an innumerable number of functions [29, 32]. The range of curves typically used for fitting can be narrowed down and identified based on the study of forestry literature (logistic, Richards and Gompertz functions).

Set of theses no. 2 formulates statements regarding the developed analytical process and the special preparatory steps of the same.

- 2.1 Thesis: By the application of the CReMIT method, a general analytical method embedded in a specific framework was implemented. The system comprised of the data preparation, the CReMIT and the analytical modules allows for the completion of the whole analytical process [5, 7, 15, 17] (Chapter 3.1).
- 2.2 Thesis: The preparatory module includes the breakpoint detecting function that allows for detecting statistically confirmable breakpoints in the data series, and for this fact to be included in correlation analyses (Chapter 3.2).
- 2.3 Thesis: The preparatory module furthermore includes the curve that fits the growth data. The fits of logistic, Gompertz and Richards functions were compared on the sample data and it was shown that for the fitting results and practical considerations, the Gompertz function is the most appropriate to describe the growth process [6, 8, 16] (Chapter 3.3).

4. Aplication of the Results and Development Opportunities

The practical applicability of the CReMIT method and the analytical process developed to expand the study of periodical time series are widely confirmed by the studies of correlations between tree growth, climate [7, 9, 12], and butterfly catching data [1, 2, 3], the health of trees and climatic features [10, 11, 13, 14]. The varied applications definitely underline the general approach of the method, and highlight its wide-scale applicability (Chapter 2.3.). In the preparatory module of the analytical process developed to support the application of the CReMIT method, two special techniques were also developed. Breakpoint detection allows for identifying abrupt changes in the studied time series, while with the help of the curves fitting the growth data, the generation of continuous data series that are able to generate further properties becomes possible.

The amount of processed data varied depending on the task, from growth data sets of 7-10 individual trees to those describing the health of trees containing over 360,000 records. The macrolepidoptera catching data covering 24 locations and more than 20 species date back to more than 50 years were used in univariable and multivariable CReMIT analyses.

The correlation analyses performed with the use of the breakpoints detected in the studied meteorological data series showed significant differences as compared to each other in the two sub-periods (in strength and even in direction), and in comparison with the total period, stronger relationships could be confirmed by the application of the CReMIT method in the various sub-periods (Periods 3.2.).

The curves fitting discrete growth data may be used to reduce random effects and to manage defective or missing data [6, 8]. The Gompertz function used for fitting and its three basic parameters can be used to define further properties in a mathematically objective way that may be related to the properties of the curve, and to study the possibility of their inclusion in analyses [16] (Chapter 3.3.).

As part of research and development, the fundamental aim was to develop tools to support the work of professionals, and mostly of forestry professionals. Considering that it is always the duty and up to the competence of professionals to achieve professional results, we define supporting our existing partners further as a priority in achieving valuable professional results.

The method, however, lends itself to application even in other scientific fields or in industrial environments, and in many cases, these applications also define new IT or statistics-related challenges. Consequently, the results of the dissertation can be further developed in three major directions:

1. By the extension of the range of TR transformation functions applied in the CReMIT module: the currently used TR functions define continuous window functions (sum, average, minimum and maximum). An important developmental direction towards indices that are often used also in forestry is the application of various weight factors that might range from simple binary weights to more complex weights.
2. By methodological development: Since the currently implemented analyses detect univariable and multivariable linear correlations, a natural step forward is to make the detection of non-linear correlations or the application of other multivariable analytical methods used in the field of data mining possible.
3. By expanding the range of application fields: The method can obviously be applied with minimum modifications in fields other than forestry too. As an example, we could mention various industrial and production processes in which, due to the steps built on one another, there naturally are time-shifted relationships. The method, among others, allows for studying the relationships between reject percentage and the different technological phases or analysing the delayed impacts of various events in the case of financial or economic time series.

In summary we can state that a general methodology and analytical process was worked out for addressing the initial forestry problems that systematically widens the scope of variables allowable to the analysis while complying with the requirements of forestry studies. The resulting technique is, however, not specific to one professional field since it is based on general principles that appear in any or at least a large number of other periodic time series too. In the preparatory module of the analytical process, by detecting breakpoints, the study of the relationships between time series can be made more accurate while the procedure of fitting curves helps the preparation of the studied data series and with generating newer parameters that can be used in the analysis.

The searchings were support by the TÁMOP-4.2.2.C–11/1/KONV-2012-0015 (Earth-system) project too.

Bibliography

The own publications appearing in Ph.D. theses

[1] Csóka, Gy., Pödör, Z., Hirka, A., Szőcs L., (2012a): Az időjárási tényezők hatása a tölgy búcsújáró lepke (*Thaumetopoea processionea* L.) populációinak fluktuációjára. Magyar Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.

[2] Csóka,Gy., Pödör, Z., Hirka, A., Führer, E. and Szőcs, L. (2012b). Influence of weather conditions on population fluctuations of the oak processionary moth (*Thaumetopoea processionea* L.) in Hungary. Joint IUFRO 7.03.10 – "Methodology of forest insect and disease survey" and IUFRO WP 7.03.06 – "Integrated management of forest defoliating insects" Working Party Meeting, Palanga, 10–14 September 2012.

[3] Csóka, Gy., Pödör, Z., Hirka, A., Führer E., Móricz, N., Rasztovics, E., Szőcs, L., (2013): Időjárásfüggő fluktuáció a tölgy búcsújáró lepke nyugat-magyarországi populációinál. Növényvédelmi Tudományos Napok, Budapest, 2013. 02. 19-20.

[4] Manninger M., Edelényi, M., Pödör, Z., Jereb, L. (2011a): Alkalmazott elemzési módszerek a környezeti tényezők fák növekedésére gyakorolt hatásának vizsgálatában. Erdészettudományi Közlemények, 1(1), 59-70.

[5] Edelényi, M., Pödör, Z., Jereb, L. (2011b): Speciális elemzési megközelítés a fák növekedése és az időjárási paraméterek közötti kapcsolatok vizsgálatában. Agrárinformatika / Agricultural Informatics (2011) Vol. 2, No. 1, 39-48.

[6] Edelényi, M., Pödör, Z., Jereb, L., Manninger M. (2011c): Telítődési görbék alkalmazása korlátos növekedési adatsorok vizsgálatában. Országos Gazdaságinformatikai Konferencia, OGIK'2011, Győr

[7] Edelényi, M., Pödör, Z., Jereb, L., Manninger M. (2011d): Másodlagos idősorokat származtató módszer kifejlesztése és bemutatása erdészeti adatokon. Acta Agraria Kaposváriensis (2011) Vol. 15 No. 3, 39-49.

[8] Edelényi M., Pödör Z., Jereb L., Manninger M. (2012): Erdei fák éves növekedémenetének közelítő leírása függvényekkel. Meteorológiai Társaság

XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.

[9] Führer, E., Edelényi, M., Jagodics, A., Jereb, L., Horváth, L., Móring, A., Pödör, Z., Szabados, I., (2012): Az átmérő-növekedés és az időjárás közötti összefüggés egy idős bükkösben. Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.

[10] Jereb, L., Edelényi, M., Pödör, Z. (2013a): EVH I. szintű mintafák levélvesztés, koronaelhalás és rovarkárosítás intenzítás adatainak vizsgálata. Kutatási jelentés, pp. 91.

[11] Jereb, L., Edelényi, M., Pödör, Z. (2013b): Mortalitás visszamenőleges vizsgálata az FNM adatokon, Kutatási jelentés, pp. 77.

[12] Manninger, M., Edelényi, M., Pödör, Z., Jereb, L. (2011): The effect of temperature and precipitation on growth of beech (*Fagus sylvatica* L.*)* in Mátra Mountains, Hungary. Applied Forestry Research in the 21st Century conference, Prága-Pruhonice, 2011.09.13-15.

[13] Pödör, Z., Kiss, B., Csóka, Gy., Jereb, L. (2013a): Egyes nagylepkefajok Magyarországi fogási adatainak lehetséges klímafüggése – vizsgálati módszertan és előzetes eredmények. Kari Tudományos Konferencia, Sopron, 2013.12.10.

[14] Pödör, Z., Csóka, Gy., Kiss, B. (2013b): Simple- and Multivariate data analysis of light trap catching data by a systematic window procedure. Decision Support System Workshop and ForestDSS Community of Practice, Lisbon, 4.-6. 12. 2013.

[15] Pödör, Z., Jereb, L. (2013c): Data analysis approach to investigate the periodic impact of climate on forestry parameters. Decision Support System Workshop and ForestDSS Community of Practice, Lisbon, 4.-6. 12. 2013.

[16] Pödör, Z., Manninger, M., Jereb, L. (2014a): Application of Sigmoid Models for Growth Investigations of Forest Trees. Applied Mathematics and Applications Conference, Conference Proceedings, Advanced Computational Methods for Knowledge Engineering, Advances in Intelligent Systems and Computing 282, 353-364.

[17] Pödör, Z., Edelényi, M., Jereb L. (2014b): Systematic Analysis of Time Series. Infocommunication Journal, VI(1), 16-21.

Other references:

[18] Abonyi, J. (2006): Adatbányászat a hatékonyság eszköze, Computerbooks, Budapest 2006, pp.400.

[19] Biondi, F. (1997): Evolutionary and moving response functions in dendroclimatology. Dendrochronologia 15, 139-150.

[20] Biondi, F., Waikul, K. (2004): DENDROCLIM2002: AC++ program for statistical calibration of climate signals in tree-ring chronologies. Computers & Geosciences 30 (2004), 303–311.

[21] Burnham, K. P. és Anderson, D. R. (2002): Model Selection and Multi-Model Inference. Springer, 2. kiadás, pp. 485.

[22] Efron, B. (1979): Bootstrap methods: another look at jackknife. Ann. Stat., 7, 1-26.

[23] Fritts, H. C. (1976): Tree rings and climate. Academic Press, London, pp. 582.

[24] Joliffe, I. T. (2002): Principal Component Analysis, second edition. Springer, 2002, pp. 518.

[25] Molnár, J. (2003):A légnyomási mező szerkezete és módosulása a Kárpát-medence térségében. Doktori értekezés, Debrecen, pp. 170.

[26] Molnár, J., Izsák, T. (2011) Trendek és töréspontok a léghőmérséklet Kárpátalja idősoraiban. Légkör 56(2), 49-54.

[27] Packalen, T., Marques, A. F., Rasinmäki, J., Rosset, C., Mounir. F., Rodriguez, L. C. E., Nobre, S. R. (2013): Review. A brief overview of forest management decision support systems (FMDSS) listed in the FORSYS wiki. Forest Systems 2013 22(2), 263-269.

[28] Reynolds, K. M., Twery, M., Lexer, M. J., Vacik, H., Ray, D., Shao, G., Borges, J. G. (2008): Handbook on Decision Support Systems 2, 499-533.

[29] Sit, V. and Poulin-Costello, M. (1994): Catalog of curves for curve fitting. BC Ministry of Forests Biometrics Information Handbook No. 4. pp. 110.

[30] Tusnádi Gábor és Ziermann Margit (szerk): Idősorok analízise, Műszaki Könyvkiadó, Budapest, 1986, pp 339.

[31] Veress, K. (2007): A Newton és Gauss-Newton módszerek alkalmazása egyenletrendszerek megoldására és nemlineáris optimalizálásra. Egyetemi jegyzet, Szeged, pp. 46.

[32] Zeide, B. (1993) Analysis of Growth Equations. Forest Science 39(3), 594-616.