Comparison of three signal analysis methods for modelling of GHG emissions and uncertainties

W. Jęda, Z. Nahorski

Systems Research Institute
Polish Academy of Sciences
Newelska 6, PL-01-447 Warsaw, Poland
E-mail: Zbigniew.Nahorski@ibspan.waw.pl

Abstract

The parties that signed the Annex I to the Kyoto Protocol report their emissions of the greenhouse gases each year. Uncertainty of reported values assessed so far for several countries is very high. One of the aim of this paper is to investigate procedures for independent calculation of this estimates.

We consider three methods for empirical estimation of the standard deviation of the accounting errors. These are signal processing methods: a smoothing procedure based on the spline functions, a parametric model with a time-varying parameter, and a model of geometric Brownian motion. They are verified on historical observations of greenhouse gas emissions from combustion of the fossil fuels.

1 Introduction

The parties who signed the Annex I have to monitor their emissions starting from the base year, which is mainly 1990. This way a dozen of emission observations for each country are already available. This redundancy in observations could be possibly favourably used to improve estimates of individual emissions in the commitment period 2008-2012, using statistical inference. In this paper we consider three methods for estimating emissions and their variances: the smoothing splines [13], [5], a parametric model with a time variable coefficient, and the Brownian motion model. Such models could be, in principle, also used for prognosis of the emission and risk of noncompliance.

Other methods could be tried to solve the problem. A simple alternative would be to use another smoothing method. Those based on the wavelets might be promising ones [2], [14]. Popular methods in the automatic control literature use the parametric models with calculation of the state errors following an earlier phase of the parameter estimation. In some of them, like in the extended Kalman filter, the parameters and the states are estimated simultaneously. Similar results can be obtained using the method of Cox & Bryson [1] where the control theory approach is explicitly applied. To use this kind of methods, the

parametric model is needed. Apart of that, at least some of them require quite long data samples to converge.

2 Notation used

By $x(t)$, as a function of time, we denote the integral of the real emission calculated on the interval $(t-1,t]$ where t is expressed in years. Thus, the integral is calculated over the one-year-back period. In the sequel we call $x(t)$ the emission. The integer value of t is assigned to the end of a year. The emission in the basic year t_0 will be denoted $x(t_0) = x_0$. The emission balances provided by the Annex I Parties are prepared by inventory of emissions from all involved activities during a year. Due to uncertainties in assessing the exact quantities and coefficients, they are in errors. We denote the observed (reported) values $y(t_i)$ or shortly y_i. The index i begins here at 0 and takes the consecutive integer values. The real emissions are unknown and can be only estimated. Hats will mark the estimated values, so $\hat{x}(t)$ is the estimated emission.

By δ we denote the fraction of the emission to be reduced within the Kyoto obligations until the commitment period. Thus at the commitment period the emission should be not greater than $(1-\delta)x_0$. The value of δ is not greater than few percents.

As it is common to express obligations in percents, it is useful to work not with the straight observations but with their logarithms. Let us denote $\hat{X}(t_i) = \ln\big(\hat{x}(t_i)/\hat{x}_0\big)$, thus $\hat{X}(t_i)$ is the logarithm of the normalized emission. As in our case $\hat{x}(t_i)/\hat{x}_0$ is close to 1, then it approximately holds

$$\hat{X}(t_i) = \ln\frac{\hat{x}(t_i)}{\hat{x}_0} \approx \frac{\hat{x}(t_i)}{\hat{x}_0} - 1 = \frac{\hat{x}(t_i) - \hat{x}_0}{\hat{x}_0} \tag{1}$$

Thus, $\hat{X}(t_i)$ may be interpreted as the relative change of $\hat{x}(t_i)$ with respect to $\hat{x}_0$ and may be expressed in percents.

3 A nonparametric method

3.1 Basic assumptions and simplifications

The function $x(t)$, as an integral of a positive function, is continuous and positive. In the paper we assume that $x(t)$ is a smooth enough function. The emissions can be only observed with errors in time instants t_i.

We assume that the real process x_i is observed with multiplicative errors $\varepsilon_i = u_i x_i$, where

$$E(u_i) = m_i, \qquad E[(u_i - m)^2] = \sigma_i^2, \qquad \operatorname{cov}(u_i, u_j) = \gamma_j$$

Thus, the observations can be modelled in the following way

$$y_i = x_i + u_i x_i = (1 + u_i)x_i, \qquad i = 0, 1, \ldots, N$$

where y_i are the observed emissions, x_i the (unknown) real emissions, and u_i their relative uncertainties.

The above dependencies are also true for $i = 0$. Dividing sides and taking the logarithms we get

$$Y_i = X_i + \ln \frac{1 + u_i}{1 + u_0}$$

where $Y_i = \ln y_i/y_0$ and $X_i = \ln x_i/x_0$. For small u_0 and u_i it approximately holds

$$\ln \frac{1 + u_i}{1 + u_0} \approx u_i - u_0$$

resulting in the expression

$$Y_i = X_i + u_i - u_0$$

The errors $v_i = u_i - u_0$ have the zero mean, $E(v_i) = 0$, and the variance $\sigma^2_{v_i} = \sigma^2_i + \sigma^2_0 - 2\gamma_{i0} = \sigma^2_i + \sigma^2_0 - 2\rho_{i0}\sigma_i\sigma_0$, where $\rho_{i0} = \gamma_{i0}/\sigma_i\sigma_0$ is the cross correlation of u_0 and u_i. The covariance is equal to

$$\mathrm{cov}(v_i, v_j) = E[(u_i - u_0)(u_j - u_0)] = \gamma_{ij} - \gamma_{i0} - \gamma_{j0} + \sigma^2_0$$

It equals zero, if all summands are equal. But generally the sequence is correlated, even if the original errors u_i are not. We assume, however, that the correlation is negligibly small.

3.2 Smoothing and uncertainty analysis

Let us consider some abstract data z_i generated by the following system

$$z_i = f(t_i) + e_i, \qquad i = 0, 1, 2, \ldots, N$$

The vector

$$\mathbf{e} = (e_0, \ldots, e_N) \propto \mathcal{N}(0, \sigma^2\mathbf{I})$$

contains the set of observation errors. We want to recover the function $f(t)$, assumed to be smooth enough, knowing only the erroneous observations $z_i, i = 0, 1, \ldots, N$. For this we use the smoothing splines. Their idea is to find the function $\hat{z}(t)$ that does not need to go directly through the observed points z_i, allowing for estimation of the observation errors.

The method described by Wahba with the generalized cross validation [13] was used. Consistency of the estimate for the smoothing splines has been proved theoretically for it ([5], th. 3.4), while other good statistical properties have been checked on numerical simulations.

The above analysis was applied for smoothing the values $Y_i = \ln y_i/y_0$ using the data on emissions from the fossil fuels provided by Marland et al. [9], in the periods 1950-1998 and 1970-1998. The estimates of standard deviations depend on the number of data used. This dependence is visible, although mostly not crucial, in the results presented in Table 1 for different time periods.

Wahba ([13], sec. 4.9) recommends using at least 25-30 observations when applied the smoothing splines. The data used in calculating the values in the years 1970-1999 in Table 1 contain 29 points for each country, just satisfying the recommendations. However, for many countries, the corresponding standard deviations differ for different length of data. Additional symptoms suggest that the data in the shorter sequence may be too short.

Table 1: Estimated standard deviations of observation errors in % for different countries and two time periods for two methods

Years	1950 - 1998		1970 - 1998		$\sim$ 2000
Country	smooth.	param.	smooth.	param.	reported
Argentina	2.3	0.7	0.4	0.1	
Australia	1.8	0.5	0.9	0.5	
Austria	2.7	0.9	1.1	1.0	
Belgium	2.3	3.3	2.3	3.3	1.1
Brazil	1.9	1.1	1.3	1.7	
Canada	1.9	0.8	0.5	1.8	
China	4.7	7.1	1.4	1.7	
Cuba	6.6	2.2	1.9	1.4	
Egypt	3.4	1.4	2.6	1.1	
Finland	4.8	1.3	3.8	3.6	3.0
France	2.3	3.0	2.3	1.1	< 2.5
Greece	2.8	0.9	2.2	0.9	
Iceland	3.5	1.4	2.7	1.4	
Ireland	4.3	1.2	2.2	2.2	< 1.0
Israel	3.4	2.2	2.0	0.9	
Italy	1.6	2.3	1.3	0.7	
Japan	2.7	4.8	1.8	2.4	
Luxembourg	2.9	4.3	2.8	4.0	
Mexico	1.7	2.1	1.7	2.0	
Netherlands	2.8	0.9	3.7	1.4	2.5
New Zealand	1.8	0.8	2.9	2.1	
Norway	4.2	2.0	5.2	3.3	
Poland	1.5	1.8	1.8	2.2	
Portugal	1.9	0.9	1.9	1.2	
Romania	1.9	2.4	2.1	2.9	
Spain	3.0	1.2	1.7	1.0	
Sweden	2.5	1.1	2.3	1.4	1.0
Switzerland	3.3	4.3	1.9	1.0	
Turkey	3.1	4.3	3.4	1.1	
U. K.	1.6	0.5	1.4	0.7	1.1
USA	1.8	0.5	0.4	2.1	

The estimated values agree quite well in magnitudes with the common idea of the errors made in calculation of the fossil fuel emission, believed to be of few percents. They also agree well with the estimates calculated for few countries, as presented in a review by Gugele at al. ([6], Tab. 6). A little bigger figures obtained in some of our calculations may be connected with some additional factors that might have influenced the calculated estimates, as year-to-year variations in the weather conditions or variations due to change in economic factors of the countries.

4 Empirical parametric models

4.1 Variable parameter model

In the previous section we noticed that the consecutive values in the emission sequence might be correlated. To better model this property, in this section we consider a set of values x_i forming a time series consisting of N elements. Then we motivate the choice of the model and finally present some results for fitting the model to the emission data for some countries.

As we assumed that x_i are positive we can define a new time series

$$g_i = \frac{x_{i+1}}{x_i} - 1 = \frac{x_{i+1} - x_i}{x_i}, \qquad i = 0, 1, \ldots, N-1$$

Each element g_i of a new time series can be interpreted as a relative difference of the two consecutive elements x_{i+1} and x_i.

From the latter relation we can now formulate the following difference equation

$$x_{i+1} - x_i = g_i x_i, \qquad x_0 = x(t_0) \tag{2}$$

As we have $y_i = (1 + u_i)x_i$, then (2) can be transformed to

$$y_{i+1} = (1 + g_i)\frac{1 + u_{i+1}}{1 + u_i} y_i$$

Dividing both sides by y_0 and taking logarithms yields

$$Y_{i+1} = \ln(1 + g_i) + \ln\frac{1 + u_{i+1}}{1 + u_i} + Y_i$$

or approximately

$$Y_{i+1} - Y_i \approx g_i + u_{i+1} - u_i$$

from where an estimator $\hat{g}_i$ can be designed as

$$\hat{g}_i = Y_{i+1} - Y_i \tag{3}$$

Under our assumption on u_i's we have

$$E(\hat{g}_i) = E(Y_{i+1} - Y_i + u_i - u_{i+1}) = X_{i+1} - X_i = \ln(1 + g_i) \approx g_i$$

Thus the estimator is unbiased (up to the approximation done). Its variance is

$$\mathrm{var}(\hat{g}_i) = E(Y_{i+1} - X_{i+1} - Y_i + X_i)^2 =$$

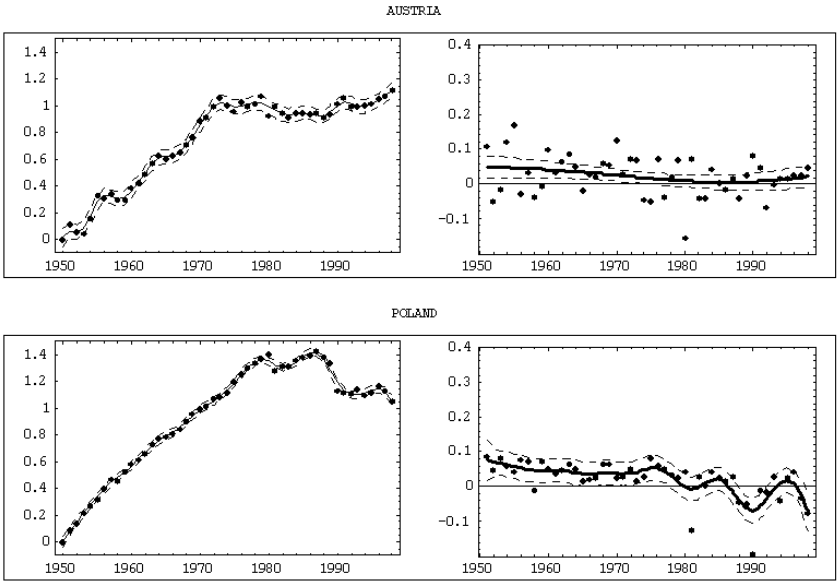

Figure 1: Results of smoothing and estimation of the function g for Austria and Poland in the years 1950-1998. Left panels: dots – observations, solid lines – smoothed emission function. Right panels: dots – estimates $\hat{g}_i$ from the formula (3), the bold solid lines – their smoothed continuous approximations, the normal thickness dashed lines – the 95% confidence intervals of these approximations.

$$= E(u_{i+1} - u_0 - u_i + u_0)^2 = E(u_{i+1} - u_i)^2 = \sigma_{i+1}^2 - 2\gamma_{i,i+1} + \sigma_i^2$$

The expression (3) was used to estimate the function g_i for few countries from the previously mentioned data of CO_2 emission from the fossil fuels published by Marland et al. [9]. Examples are presented in Fig. 1. The smoothing splines were used to smooth the points obtained from (3). For each country, in the left panel the observations (dots) and their smoothing spline approximations (solid lines) are depicted. The right panel shows the estimates of the function g_i. The dots represent the points calculated using the formula (3). The bold dashed line is obtained by smoothing these points. The normal thickness dashed lines on both panels show the 95% confidence intervals of the estimates.

Table 1 depicts the estimates of the standard deviation of the errors $u_{i+1} - u_i$. Comparison of the values presented in Table 1 shows that both estimates of the standard deviations are of the same order, although not always satisfactorily close to each other. Notice, however, that the results for smoothing show the standard deviations of the errors $u_i - u_0$, while that for the parametric model of $u_i - u_{i-1}$, which might cause the differences.

4.2 Piecewise exponential model

Although the estimated functions $\hat{g}(t)$ in the previous section vary in time, at least in some instances their patterns resembles the constant value lines. To better investigate this question let us start with examining few curves.

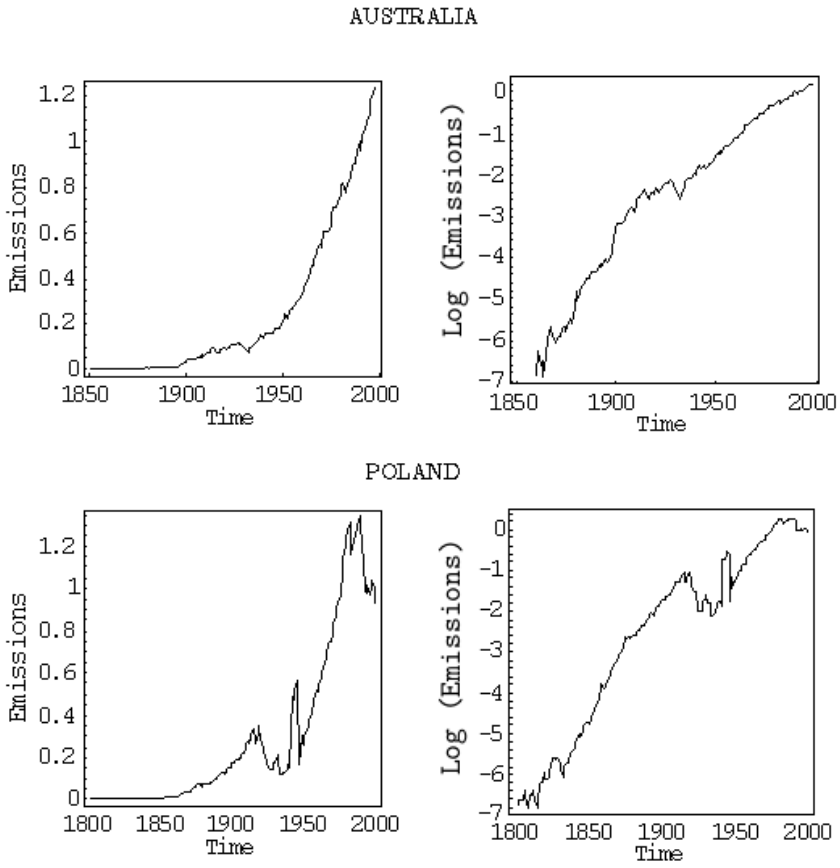

Figure 2: CO_2 total emissions and logarithms of emissions (1870 – 1998): Australia (upper panel) and Poland (lower panel)

Fig. 2 presents examples of emission curves $y(t_i)$ and logarithmic curves $Y(t_i) = \ln y(t_i)/y(t_0), t_0 = 1990$, for the emission data from Marland et al. [9] for Austria and Poland. It can be seen that the data evolve approximately along piecewise exponential curve, and the logarithmic curves are approximately linear.

However, looking at Fig. 2 we can easily notice periods where this simple constant evolution (and therefore the growth along the exponential curve) does not hold. This is particularly visible in the periods of the Great Crisis of 1930s, the 2nd World Word, and the collapse of the communist regime. Thus, the exponential growth models describe quite well development of data only in some intervals. These intervals seem to be the periods of constant development conditions.

The fit of this simple piecewise exponential model is quite good in the periods of growth or decay. In the period of steady growth it is almost perfect. In the decay periods the emission is often more volatile. War and transition periods, like those of 1970s in the West Europe or 1980s in Poland, are highly irregular and were skipped from fitting.

The results obtained are generally quite similar for both methods. The error variance estimates calculated by the regression method (parametric model) turn out to be usually small, although mostly greater than those calculated by the smoothing splines. This seems to be connected with too big simplicity of the exponential model used. The good fit of the piecewise exponential model seems to be an important observation. It means that up to now the emissions have followed approximately the exponential functions in defined longer periods. The jump from one such segment to another is mostly connected with a big political or economic change.

5 Geometric Brownian motion

Geometric Brownian Motion (GBM) is the most used stochastic process in financial economics theory, and in our case may be considered as an useful alternative from a practical point of view. In several cases this is not the better model, even being a reasonable mapping of probabilities in time.

For a signal $x(t)$ that follows a geometric Brownian motion, the stochastic equation for its variation in time t is

$$dx = gxdt + \sigma x dz \tag{4}$$

where $dz = \varepsilon dt^{1/2}$ is the Wiener increment, ε - standard normal distribution, g is the drift, and σ is the volatility of x.

In the above equation the first term on the right side is the expectation (trend) term and the second term is the variation term (deviation from the tendency or term of uncertainty), growing proportionally to the time interval.

The geometric Brownian motion is a log-normal diffusion process with the expected value of x at the time t (starting at $t_0 = 0$)

$$E[x(t)] = x(t_0)\exp(gt) \tag{5}$$

and the standard deviation (SD)

$$SD[x(t)] = x(t_0)\exp(gt)\big[\exp(\sigma^2 t) - 1\big]^{1/2} \tag{6}$$

This is illustrated in the Fig. 3.

Due to its simplicity, it is useful to work with the logarithmic diffusion equation. Letting $X = \ln(x)$, and using Itô's lemma we find that X follows the arithmetic (or ordinary) Brownian motion

$$dX = d\ln x = (g - \frac{1}{2}\sigma^2)dt + \sigma dz \tag{7}$$

so

$$dX = g'dt + \sigma dz$$

where $g' = g - \sigma^2/2$. The logarithm of x follows an arithmetic Brownian motion with the drift g' and volatility σ.

We should note here that although the volatility term here is the same in (7) as in in the geometric Brownian motion for x in equation (4), the element $d(\ln x)$ is different from dx/x due to the different drift expression (so called Itô's effect).

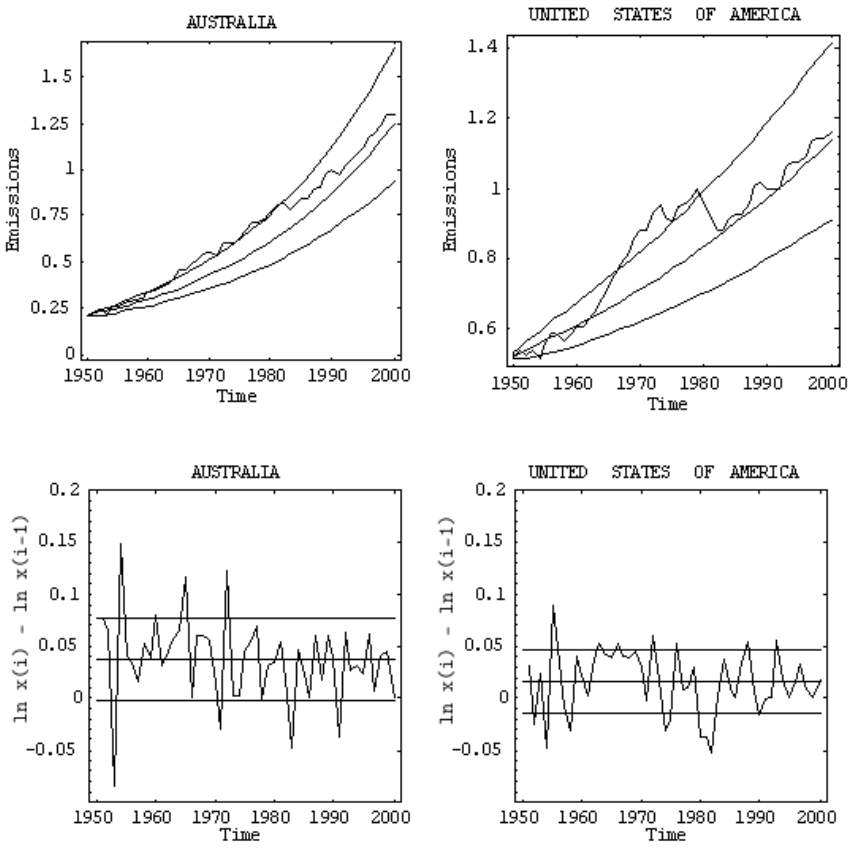

Figure 3: Illustrations of the considered stochastic process, showing sample paths, the 66% confidence intervals, and the forecasted expected values (exponential trend line) for two countries (upper panel) and calculation of $\hat{g}$ and its variance from the difference $\ln x_i - \ln x_{i-1}$ (lower panel).

Using the logarithm of x approach we can estimate drift parameter g as the average value of a set of differences of the logarithms $\ln x_i - \ln x_{i-1}$. With the same historical series we can get an estimation of the volatility σ by taking the standard deviation of $\ln x_i - \ln x_{i-1}$, as for the parametric model of Sec. 4. They can be inserted in equations (5) and (6) to obtain the characterization of the process in time.

The preliminary calculations, not presented here, show rather big estimates of the standard deviations, comparable to the piecewise exponential model of Sec. 4. These big values seem to be caused by constant value of g in the model.

6 Conclusions

Nonparametric and parametric methods for modelling the greenhouse gas emission phenomena and for estimating the parameters are proposed in the paper. They differ in degree of smoothing and precision of fitting the observations. Comparison of the methods made up to now reveals that the method of formula (3) gives more smooth curves in many instances, although it is more sensitive to the smoothing interval. The smoothing method of Sec. 3 is more accurate and better emphasizes the ripples in data. The parametric piecewise exponential model gives more rough but also simple description, showing general trends in evolution of emission data. The results from the Brownian motion model are too preliminary and it is perhaps too early to draw any finite conclusions on them.

The volatility of observations may be related not only to the observation errors but also to such factors as changing weather conditions and rapidly changing economic situation of the country. These phenomena may increase the estimated variance.

Under this reservations, the calculations performed for the fossil fuels indicate that the empirical approach gives reasonable estimates, comparable to those obtained from inventories. However, the present knowledge does not allow us to state definite conclusions as yet.

An interesting result connected with the relation between the piecewise exponential character of the emission curve and the economic development of the country may extend to some other components of the emission. An open question is whether this will concern removal of the greenhouse gases by sinks, also included in the full calculation of the greenhouse gas balance of countries. Evolution of this type of data in time will be possible to analyse when longer historical records will be available.

The proposed approach can be used to better estimate the real emissions, by filtering out the errors, and possibly for prognosis. The latter is still rather risky until more will be known on dependence of the emissions on economic and weather conditions.

Acknowledgment Partial financial support from the Polish State Scientific Research Committee within the grant 3PO4G12024 is gratefully acknowledged.

References

[1] Cox A.B., Bryson Jr. A.E.: Identification by a combined smoothing nonlinear programming algorithm. *Automatica*, 16, 689-694, 1980.

[2] Debnath L.: *Wavelet Transforms and Their Applications*. Birkhäuser, 2002.

[3] Report of the Conference of the Parties on Its Third Session, Held at Kyoto From 1 to 11 December 1997. Addendum. Document FCCC/CP/1997/7/Add.1. United Nations Framework Convention on Climate Change (FCCC), 1998. Available on the Internet: http://www.unfccc.de/ http://unfccc.int/index.html.

[4] Implementation of the Buenos Aires Plan of Action: Adoption of the Decisions Giving Effect to the Bonn Agreements. Draft Decisions Forwarded for Elaboration, Completion and Adoption. National Systems, Adjustments and Guidelines Under Articles 5, 7 and 8 of the Kyoto Protocol. Document FCCC/CP/2001/L.18. United Nations Framework Convention on Climate Change (FCCC), 2001. Available on the Internet: http://www.unfccc.de/.

[5] Gu Ch.: *Smoothing Spline ANOVA Models*. Springer, 2002.

[6] Gugele, B., K. Huttunen, Ritter M.: Annual European Community Greenhouse Gas Inventory 1990–2001 and Inventory Report 2003. Technical Report No. 95, European Environment Agency, Copenhagen, Denmark, 2003. Available on the Internet: http://reports.eea.eu.int/technical_report_2003_95/en.

[7] Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. J. Penman, D. Kruger, I. Galbally, T. Hiraishi, B. Nyenzi, S. Emmanuel, L. Buendia, R. Hoppaus, T. Martinsen, J. Meijer, K. Miwa and K. Tanabe (eds.), Intergovernmental Panel on Climate Change (IPCC) National Gas Inventories Program, Technical Support Unit, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan, 2000.

[8] Land Use, Land-Use Change, and Forestry. Special Report of the Intergovernmental Panel on Climate Change (IPCC). R.T. Watson, I.R. Noble, B. Bolin, N.H. Ravindranath, D.J. Verardo and D.J. Dokken (eds.), Cambridge University Press, Cambridge, United Kingdom, 2000.

[9] Marland G., R. J. Andres, T. A. Boden, C. A. Johnston, Brenkert A. L.: Global, regional, and national CO2 emission estimates from fossil fuel burning, cement production, and gas flaring: 1751-1996 (revised March 1999). Carbon Dioxide Information Analysis Center, 1999. Available on the Internet: http://cdiac.esd.ornl.gov/trends/emis/em_cont.htm.

[10] Nahorski Z., Jęda W., Jonas M.: Coping with uncertainty in verification of the Kyoto obligations. In: Studziński J., Drelichowski L., Hryniewicz O.: *Zastosowania informatyki i analizy systemowej w zarządzaniu*, IBS PAN, Warszawa 2003, 305-317.

[11] Nilsson, S., M. Jonas, M. Obersteiner, Victor D.: Verification: The Gorilla in the Struggle to Slow Global Warming. *The Forestry Chronicle*, 77, 475–478.

[12] Nilsson, S., M. Jonas, Obersteiner M.: COP 6: A Healing Shock. *Climatic Change*, 52: 25–28.

[13] Wahba G.: *Spline Models for Observational Data*. SIAM, 1990.

[14] Walter G.G.: *Wavelets and Other Orthogonal Systems With Applications*. CRC Press.