Estimation of Temporal Uncertainty Structure of Greenhouse Gas Inventories for Selected EU Countries

Jolanta Jarnicka
Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland;
and International Institute for Applied Systems Analysis
Schlossplatz 1, 2361 Laxenburg, Austria
Email: Jolanta.Jarnicka@ibspan.waw.pl

Zbigniew Nahorski
Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland;
and Warsaw School of Information Technology
Newelska 6, 01-447 Warszawa, Poland
Email: Zbigniew.Nahorski@ibspan.waw.pl

Abstract—The paper addresses the problem of uncertainty in the greenhouse gas emission inventories, by proposing an alternative method for assessing uncertainty and its evolution over time. To estimate the inventory accuracy, the revisions published in consecutive years are used. These revisions are considered nonstationary time series. We describe evolution by time-dependent models, used to analyze data from the National Inventory Reports published annually up to 2015, for selected EU countries. We present a parametric model and a procedure for estimating parameters, along with the results obtained.

I. INTRODUCTION

According to the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol, each of the cosignatories is obliged to provide annual data on greenhouse gas (GHG) inventory. These data are given in the National Inventory Reports (NIR), prepared either according to the 2000 IPCC report ‘Good Practice Guidance and Uncertainty Management in National Greenhouse Inventories’[1], or later to the 2006 ‘IPCC Guidelines for National Greenhouse Gas Inventories’[2], describing in detail how uncertainty analysis should be conducted. Each NIR report contains data from a given year and revisions of past data. Data for previous years are revised given more precise information. This means that, revisions made in different years use different knowledge, and hence uncertainties in different revisions change.

In general, uncertainty associated with GHG inventory can be classified as scientific uncertainty (when the actual emission and/or removal process is insufficiently understood) and estimation uncertainty (mostly structural, connected with activity data, emission factors, and other parameters), present whether GHG emissions are quantified. Three tier’s are described for categorizing both emissions factors and activity data. A tier represents a level of methodological complexity. Tier 1 is the basic method, while Tier’s 2 and 3 are each more demanding in terms of complexity and data requirements. Two of these approaches i.e. the error propagation (Tier 1) and the Monte Carlo approach (Tier 2) are recommended to assess uncertainty. The first one is much easier to calculate, while the second is considered more accurate. Since the use of the given approach is only suggested, most countries use only one of these approaches, or change the method of uncertainty assessment in consecutive reports, which makes it difficult to compare the estimates and its changes over time. In particular, it may happen that the alleged reduction in uncertainty is in fact connected with the different method of its assessment. The goal of this paper is to present an alternative, data-driven method of uncertainty assessment.

The problem of uncertainty analysis from the report data is not new, and has been dealt with for several years. Various databases were analyzed, including the IPCC data from the National Inventory Reports, but in most papers, all revision data were studied independently, e.g. in [3]. The question of how to analyze temporal evolution of the accuracy of emission inventories from several revisions was first formulated in [5], where uncertainty estimates were calculated for Austria using an algebraic approach, based on available data from different revisions year by year. Some conclusions from the results obtained there, were also presented in [4]. A similar year by year approach was presented in [8], using the Austrian NIR data as well.

This paper presents a different, revisions oriented analysis. We are interested in all consecutive yearly revisions, and differences between them, rather than in examining each of them separately. Intuitively this means that, by analyzing consecutive revisions and therefore errors and inaccuracies associated with them, which are different for each revision, we want to capture the structure of the uncertainty and its evolution over time. The method proposed combines a nonparametric regression technique using smoothing splines, as presented in [3], with a parametric model. This two-step semiparametric approach enables prior preparation of the data to which the model is fitted. Using the spline, is aimed at smoothing the data, i.e. at de-trending of the time series reported, and this in turn results in much better modelling.
This paper continues considerations outlined in [6] and carried out in [7], where we discussed some parametric model, applied to the NIR data published up to the year 2007. With the analysis carried out on longer samples, i.e. based on the NIR data published up to 2015, we managed to significantly improve the model, and get more representative results.

In Section II we present the idea of interpreting the data and propose a parametric model, that describes the uncertainty structure. Section III contains the results of fitting the model to the data on CO2 emission from the National Inventory Reports for selected EU countries, along with the uncertainty assessment. Conclusions are given in Section IV.

II. DATA AND MODEL

We analyze data from the National Inventory Reports for selected EU countries. To consider a model, the data must first be interpreted in a manner which allows the extraction of uncertainty.

A. Data interpretation

Let $E^n_{y_j,t}$ denote the inventory data for the country $i$, in the year $n$, $n = 1, \ldots, N_j$, revised in the year $y_j$, $j = 1, \ldots, J$, where $y_j$ is the last year, when the last revision was made. The index $j$ enumerates the revisions. For a given country $i$, all the inventory data form a table (Table I), in which each row contains revision data reported in the year $y_j$ i.e. a time series indexed by $n$, and each column contains emission inventory for the year $n$, recalculated in consecutive yearly revisions up to $n$, (a time series indexed by $y_j$). The analysis will be conducted for rows of that table, i.e. investigating emission inventories from consecutive yearly revisions.

For a given country $i$, we model any revision data to be composed of the ‘real’ emission, which we call the ‘deterministic’ fraction and the ‘stochastic’ fraction, related to our lack of knowledge and imprecision of observation of the real emission. We assume that the uncertainty is related to the stochastic part of the model.

For the most recently revised data, there is

$$E^n_{y_j,t} = D^n_{y_j,t} + S^n_{y_j,t}, \quad S^n_{y_j,t} \sim \mathcal{N}(0, \sigma^n_{y_j,t}),$$

where $E$ stands for the emission inventory, $D$ for its deterministic fraction, $S$ for the stochastic fraction, and $n$ is the year, for which the revised data were recalculated.

Now, the data revised in the year $y_j$, where $j = 1, \ldots, J - 1$ are modeled as having the same deterministic fraction. Thus they follow the same type of decomposition

$$E^n_{y_j,t} = D^n_{y_j,t} + S^n_{y_j,t}, \quad S^n_{y_j,t} \sim \mathcal{N}(0, \sigma^n_{y_j,t}), \quad (1)$$

where the standard deviations $\sigma^n_{y_j,t}$ are of the form

$$\sigma^n_{y_j,t} = \sqrt{\sigma^2_{y_j,t} + \alpha_{j,t}(y_j - y_j)^2}, \quad \alpha_{j,t} > 0. \quad (2)$$

Parameters $\alpha_{j,t}$, associated with the stochastic fraction $S^n_{y_j,t}$, can be estimated from the data together with $\sigma^n_{y_j,t}$. They describe a shift of the precision level and depend on the difference between the revision year $y_j$, $j = 1, \ldots, J - 1$ and the most recent revision year $y_J$, due to the learning. The deterministic fraction $D^n_{y_j,t}$ is found, using smoothing spline, as presented in [3]. Applying this nonparametric approach to the most recently revised data $E^n_{y_J,t}$, gives not only the estimate of the deterministic fraction, but also an estimate of the variance $\sigma^2_{y_J,t}$. Considering dependence of the obtained results on spline determination, all results given in the paper are conditioned on the splines, though it is not additionally stressed in the sequel.

B. Model and Parameters

Given the smoothing spline $S^n_{y_J,t}$, we consider it the estimate of $D^n_{y_J,t}$ and use it, along with the emission data $E^n_{y_i,t}$, for $j = 1, \ldots, J - 1$ to estimate uncertainty described by $S^n_{y_J,t}$. The method is based on analysis of the differences between the revisions $E^n_{y_i,t}$, and the smoothing spline $S^n_{y_J,t}$

$$v^n_{y_i,t} = E^n_{y_i,t} - S^n_{y_J,t},$$

where $j = 1, \ldots, J - 1, n = 1, \ldots, N_j$. Following (1) – (2), we assume that, for a fixed country $i$

$$v^n_{y_i,t} \sim \mathcal{N}(0, \sigma^n_{y_i,t}), \quad (3)$$

where differences $v^n_{y_i,t}$ are independent and

$$\sigma^n_{y_i,t} = \sqrt{\sigma^2_{y_i,t} + \alpha_{j,y}(y_i - y_i)^2} \quad (4)$$

Parameters $\alpha_{j}$ in (4) can be obtained as Maximum Likelihood estimators. Due to (3) the log-likelihood function, with parameter $\sigma^n_{y_i,t}$ is of the form

$$\ln L(\sigma^n_{y_i,t}) = -N_j \ln \sqrt{2\pi} - \frac{1}{2} N_j \ln(\sigma^n_{y_i,t}) - \frac{1}{2(\sigma^n_{y_i,t})^2} \sum(v^n_{y_i,t})^2$$

Substituting (4), gives

$$\ln L(\alpha_{j}) = -N_j \ln \sqrt{2\pi} - \frac{1}{2} N_j \ln((\sigma^n_{y_i,t})^2 + \alpha_{j}(y_i - y_i)^2)$$

Then

$$\frac{d \ln L(\alpha_{j})}{d\alpha_{j}} = -\frac{N_j (y_i - y_i)^2}{2((\sigma^n_{y_i,t})^2 + \alpha_{j}(y_i - y_i)^2)}$$

$$+ \frac{1}{2} \sum(v^n_{y_i,t})^2 ((y_i - y_i)^2 - \alpha_{j}(y_i - y_i)^2)^2$$
Applying the necessary condition of extreme, we get the ML estimator of $\alpha_j$, $j = 1, \ldots, J - 1$

$$\hat{\alpha}_j = \frac{1}{(y_j - y_j)^2} \left( \frac{1}{N_j} \sum (v^n_j)^2 - (\sigma^n_j)^2 \right).$$ \hspace{1cm} (5)

Having obtained (5), we take

$$\alpha_j = \beta(y_j - y_j)^\gamma$$, where $\alpha_j > 0$$ \hspace{1cm} (6)

which leads us to the following model

$$v^n_j \sim N(0, \sigma^n_j), \text{ where } \sigma^n_j = \sqrt{\sigma^2_{y_j} + \beta(y_j - y_j)^{\gamma + 2}}.$$ \hspace{1cm} (7)

Parameters $\beta$ and $\gamma$ in (7) are to be estimated by the Least Squares method, fitting (6) to the sequence $\hat{\alpha}_j$. We put $\alpha_j = \ln \hat{\alpha}_j$, and $\beta = \ln \beta$, which brings (6) to the following regression model

$$\hat{\alpha}_j = \tilde{\beta} + \gamma \ln(y_j - y_j).$$ \hspace{1cm} (8)

Parameters $\sigma^n_j$, $j = 1, \ldots, J - 1$ are now obtained from (7). Dividing $\sigma^n_j$ by the smoothing spline $Sp_{y_j}$ gives the relative uncertainty estimates of the form

$$\hat{\alpha}_j = \frac{\sigma^n_j}{Sp_{y_j}}, \text{ where } j = 1, \ldots, J - 1.$$ \hspace{1cm} (9)

The procedure for a fixed country $i$, consists of two steps. 

Procedure 2.1: Assessing uncertainty.

Step 1 For the most recently revised data $E^n_y$

- find the smoothing spline $Sp_{y_j}$
- estimate the variance $\sigma^2_{y_j}$
- calculate the differences $v^n_j = E^n_y - Sp_{y_j}, j = 1, \ldots, J - 1.$

Step 2 To fit the parameters in model (3) – (4)

- find $\hat{\alpha}_j$, $j = 1, \ldots, J - 1$, using (5)
- estimate $\beta$ and $\gamma$ in regression model (8)
- find $\sigma^n_{y_j}$, $j = 1, \ldots, J - 1$, due to (7)
- find relative uncertainty estimates $\hat{\alpha}_j$, $j = 1, \ldots, J - 1$, using (9).

III. UNCERTAINTY ASSESSMENT

We analyze UNFCCC data on CO₂ emission (in Gg) without land-use, land-use change and forestry (LULUCF), published yearly in the National Inventory Reports up to the year 2015 [9]. Calculation of emission estimates based on the measurements collected takes approximately two years, so the data reported in 2015 originate from the year 2013.

To illustrate various features of uncertainty structure we consider the data for six EU countries: Austria, Belgium, UK, Denmark, Ireland, and Finland. Each of them started to report data on GHG emission before the agreed year 2003 (on $y_j=2001$), conducting a test phase and providing data on emission since 1999 (Austria, UK, Ireland, and Finland) and since 2000 (Belgium and Denmark). This means we analyze the data on CO₂ emission excluding LULUCF for the year $y_j = 2013$, and all the earlier revisions, down to 1999 or 2000.

The smoothing splines $Sp_{y_j}$, built for the most recently revised data, i.e. a time series $E^n_{y_j}$, where $n = 1990, \ldots, 2013$, for each of the countries considered are depicted in Fig. 1. They clearly evidence sudden year-to-year changes in the inventories that are interpreted as results of errors with respect to the curves obtained by smoothing.

One can notice that, the spline fit is not the same for all countries. It seems to be much better for countries whose emissions are depicted in figures in the left-hand column of Fig. 1. Since the purpose of the spline was to extract data on the ‘real’ emission, this can be explained by different levels of uncertainty, reported by these countries for the year 2013. Total uncertainty, reported for 2013 in the Austrian NIR was equal 4.27%, the one for the UK 4%, and 3.45% for Ireland. In the case of Belgium and Denmark the spline obtained significantly smooths the data reported, which can be interpreted in terms of higher uncertainty – the total uncertainty reported for 2013 was equal 5.53% and 5.2% for Belgium and Denmark respectively.

Fig. 1(d) can be considered a good example illustrating the problem of uncertainty assessment. Finnish total uncertainty reported for 2013 is the highest of them, estimated for 6%, although the fit may suggest much lower value. This is connected with the fact, that the uncertainty assessment in the Finnish NIR was obtained using Tier 2 (Monte Carlo approach), while the remaining ones are calculated using Tier 1. Just like in this case, the results obtained through different approaches are often difficult to compare.

Fig. 1 demonstrates also some similarities in the monotonic behaviour of emissions for the countries analyzed. It’s easy to see a decreasing trend starting from 2005 (it is visible even in the case of oscillating Belgian and Danish emissions, heavily smoothed by the spline). More interesting, however, is a significant drop in emissions in 2009, associated with the economic crisis.

The estimates for variances $\sigma^2_{y_j}$, where $y_j = 2013$, calculated when building the smoothing spline are equal $2715127$, $8091277.78$, $121449538.0$, $1518214.8$, $650202.035$, and $5940914.85$ for Austria, Belgium, UK, Denmark, Ireland, and Finland respectively.

Having built the smoothing spline for $y_j = 2013$, we subtracted it from all the earlier revisions $E^n_{y_j}$, where $y_j = 1999, \ldots, 2012$ in the case of Austria, the UK, Ireland, and Finland, and $y_j = 2000, \ldots, 2012$ for Belgium and Denmark. In each case, the revision data represented a time series $E^n_{y_j}$, for $n = 1990, \ldots, y_j$.

The assumptions in model (7) were checked, by performing statistical tests. The differences obtained, were tested for normality using the Shapiro-Wilk test (considered the most reliable normality test). Since in some cases we had to deal with small samples, the results were confirmed by the Lilliefors test (modification of the Kolmogorov-Smirnov test with unknown parameters). Moreover the differences were tested for significance of true population mean, using two-sided t-test, and taking significance level 0.05.
In most cases, there was no statistical evidence against the null hypothesis on normality of the data analyzed, and hence the alternative hypothesis was clearly rejected. The test failed in the case of the most initial revisions, in particular those provided in the test phase, i.e. concerning emission data on the years 1999 – 2000, which can partly be explained by the fact that the method of calculating emission revisions was being developed at that time. The \( t \)-tests performed on these differences, for which the normality assumption was met, showed that in most cases true population mean, is statistically insignificant and can be assumed zero. This means that, the assumptions taken in the model considered, were reasonable.

Following procedure 2.1, we found coefficients \( \alpha_j \), as ML estimators (5) and fitted parameters in regression model (8). The model fitted has been verified. All the parameters estimated turned out to be significant – the null hypotheses considering them insignificant were rejected, since \( p \)-values in the case of \( \tilde{\beta} \) ranged from \( \exp(-7) \) to \( \exp(-13) \) and for \( \gamma \) from \( \exp(-4) \) to \( \exp(-8) \), the same as \( p \)-values in the \( F \)-test.

Coefficient of determination \( R^2 \), indicating the goodness of fit of the model considered, was equal 84% for Austria, 77% for Belgium, 86% for the UK, 94% for Denmark, 81% for Ireland, and 97% for Finland.

The estimates of \( \gamma + 2 \) and \( \tilde{\beta} \) are given in Table II, along
with the results of model (8) validation.

The main result of the paper – relative uncertainties $u_j$, calculated due to (9) based on $\sigma_{y_j}^n$, $j = 1, \ldots, J - 1$ in (7), for each of the six countries considered, are given in Table III and depicted in Fig. 3.

The temporal evolution of standard deviations $\sigma_{y_j}^n$, $j = 1, \ldots, J - 1$ in (7), for all countries analyzed is shown in Fig. 2. It can be observed that, they are decreasing rather slowly in time. In turn, the corresponding relative uncertainty estimates $\hat{u}_j$, $j = 1, \ldots, J - 1$, presented in Fig. 3, for some countries, like Austria, UK, and Ireland decrease, some others, like those for Denmark or Belgium increase slowly until 2004 – 2005. An exception is Finland, for which nonmonotonic and slightly oscillating values of $u_j$ are shown. However, it can be observed that the uncertainty estimates for all of the countries analyzed in the paper grow quite rapidly in later years (in particular those for Denmark and Ireland). This is connected with noticeable decrease of emissions for all countries in the years 2005 – 2012, much quicker than the slow decrease of standard deviations $\sigma_{y_j}^n$. Note that the values of $\sigma_{y_j}^n$ estimated for the UK, are much higher than for other countries, due to much higher emissions. This did not prevent us, however, in getting a good uncertainty assessment, also for that country.

It is worth stressing that, the uncertainty estimates obtained due to (9) agree quite well with the official uncertainty assessments provided in the National Inventory Reports (see Fig. 4). Uncertainty assessments have become part of the 2000 IPCC Good Practice Guidance [1]. Next to the emission data, parties are expected to provide assessment of total uncertainty level of reported emission and trend uncertainty, using Tier 1 analysis (error propagation), along with the uncertainty assessment for each greenhouse gas, and each of the key IPCC categories. It is also suggested to give the results at the Tier 2 level (Monte Carlo simulation), if available. The advantage of using Tier 2 methodology is that uncertainties are taken into account and the ranking shows where uncertainties can be reduced. In the 2006 IPCC guidelines [2] it is suggested that, good practice reporting should include analysis of both Tier 1 and Tier 2.

Since applying given approach is only suggested, most countries use only one of them, i.e. Tier 1, which is easier to calculate, although considered to be less accurate. Tier 2 approach is used by Austria (starting with 2005), for Finland (for the years 2001-2005 Tier 1 analysis was not conducted, the uncertainty assessments published in the National Inventory Reports was based on Tier 2), and the UK (2003), while Belgium provides the uncertainty assessments obtained using Monte Carlo approach only for Flanders.

We compared the resulting uncertainty estimates $\hat{u}_j$, with the reported Tier 1 trend uncertainty, and the uncertainty of CO$_2$ (Fig. 4). In the case of Finland, the trend uncertainty reported for 2001-2005 and 2011-2012 was calculated using Tier 2 approach, therefore in Fig. 4 both methods are considered. For the convenience of the reader, we set the same range on the vertical axis for all figures in Fig. 4, which allows for better comparison of the estimated uncertainties.

It can be seen that, the proposed uncertainty estimates correspond to those reported in the National Inventory Reports. In the case of Austria and the UK, they almost coincide with the official CO$_2$ uncertainty assessments, but after 2006, you may notice a slight difference in both ratings – the relative uncertainties $u_j$ are then slightly higher than the reported ones. For Belgium and Denmark the values of $u_j$ are pretty close to the uncertainty assessments for CO$_2$ and show similar monotonic behaviour. The estimates obtained for Ireland and Finland agree rather with the CO$_2$ uncertainty assessment, however the general monotonic behaviour is comparable with the trend uncertainty reported.

Assessments of trend uncertainty, reported since 2003 and 2004, are somewhat higher than the determined values of $u_j$, which may be partly explained by the fact that, they relate to total GHG emission.

The results obtained using the method proposed can therefore be considered independent confirmation of the official uncertainty estimates calculated according to [1] and [2].
TABLE II  
ESTIMATES OF \( \gamma + 2 \) AND \( \bar{\beta} \) IN (8).

<table>
<thead>
<tr>
<th>Model</th>
<th>Austria</th>
<th>Belgium</th>
<th>UK</th>
<th>Denmark</th>
<th>Ireland</th>
<th>Finland</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma + 2 )</td>
<td>0.60</td>
<td>0.91</td>
<td>0.59</td>
<td>0.62</td>
<td>0.48</td>
<td>1.02</td>
</tr>
<tr>
<td>( H_0 : \gamma = 0 ) against ( H_1 : \gamma \neq 0 )</td>
<td>8.48e-05</td>
<td>7.35e-05</td>
<td>6.33e-07</td>
<td>4.78e-08</td>
<td>7.75e-05</td>
<td>4.93e-11</td>
</tr>
<tr>
<td>( \beta )</td>
<td>7.63</td>
<td>7.50</td>
<td>8.08</td>
<td>8.40</td>
<td>6.74</td>
<td>7.67</td>
</tr>
<tr>
<td>( H_0 : \beta = 0 ) against ( H_1 : \beta \neq 0 )</td>
<td>1.53e-09</td>
<td>1.60e-10</td>
<td>0.00078</td>
<td>1.60e-13</td>
<td>2.24e-08</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>F-test</td>
<td>8.48e-05</td>
<td>7.35e-05</td>
<td>0.00078</td>
<td>4.78e-08</td>
<td>7.74e-05</td>
<td>4.92e-11</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.843</td>
<td>0.773</td>
<td>0.856</td>
<td>0.933</td>
<td>0.814</td>
<td>0.968</td>
</tr>
</tbody>
</table>

Fig. 4. Uncertainty assessment by means of relative values \( \hat{u}_j \) compared with the NIR uncertainty reported.
The method gives quite smooth temporary curves for uncertainty estimates evolution. This effect can be induced by a model structure which was designed to catch the main directions of uncertainty evolution rather than very accurate estimation of local changes. An advantage of such model is, that the estimates obtained from our statistical estimation resemble the uncertainty estimates reported by countries as to the similar smoothness of both kinds of curves. They are also close in values to the estimates reported by countries.

The method presented is general enough to be applied to inventory data provided by other countries. Being a statistical method, its accuracy depends a lot on the number of available data points. The six countries analyzed in this paper were chosen due to their relatively long inventory sequences. It is intended to extend the calculation for other countries. Studying common properties of these models hopefully enables obtaining reliable results also for countries with shorter reported samples.

**REFERENCES**


