Estimation of Human Impact in the Presence of Natural Fluctuations

Anders Grimvall

Abstract

Natural fluctuations in weather conditions can long conceal or distort important trends in the human impact on the environment. The EU-IST project IMPACT provides methods and tools that will facilitate extraction of anthropogenic signals from collected data. In this paper, we show how deterministic, process-oriented models can be incorporated into statistical procedures for the estimation of anthropogenic trends in air and water quality.

1. Introduction

During the past few years, scientists and policy makers have become increasingly aware of the fact that observations of the state of the environment rarely speak for themselves. Simple questions regarding the effect of specific policy measures or general temporal trends in environmental quality can not be answered unless natural fluctuations in measured data are identified and removed or suppressed. Hence, a number of normalisation and adjustment techniques have been developed. Air quality data are normalised by removing the effect of anomalies in wind direction, temperature and humidity at the sampling occasions (Thompson 2001). Riverine loads of nutrients and other substances are normalised to represent long-term average runoff values (Stålnacke/Grimvall 2001), and measurements in brackish waters are normalised to remove the impact of temporal changes in the mixing of seawater and freshwater.

Most of the currently used techniques to normalise environmental quality data are based on some kind of regression analysis. The response variable under consideration is regressed on weather data or other concomitant information representing natural variability, and thereafter the fitted regression model is used to adjust observed responses. The EU-IST project IMPACT has demonstrated that both conventional regression techniques and more advanced statistical procedures can greatly facilitate the extraction of anthropogenic signals from collected data. However, the

1 Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden
project has also demonstrated that purely statistical approaches can have significant
drawbacks. Complex non-linear relationships between the response variable and the
explanatory variables can be difficult to identify, and there is a considerable risk of
over-adjustment if the number of possible explanatory variables is large. In this pa-
per, we examine how the normalisation of environmental data can be made more
efficient by incorporating process-oriented or mechanistic models into statistical
procedures.

2. A simple example

Let us assume that a major intervention has been undertaken in a catchment and that
we would like to assess the impact of that intervention on the runoff. Let us further
assume that we have a time series $y$ of observed runoff values before and after the
intervention, and that we have fitted a physics-based, mechanistic model

$$\tilde{y}_0 = f_0(u_0)$$

to runoff ($y_0$) and rainfall ($u_0$) data collected before the intervention. We can then
keep the model parameters representing the baseline conditions, extend the run to
the entire time period, and compute the residuals

$$y - \tilde{y}_0 = y - f_0(u).$$

If the mechanistic model is adequate, this operation will remove a substantial part
of the natural variability of the observed runoff values, and hence facilitate the as-
assessment of intervention effects.

3. The need for alternative procedures

Let us now consider an environmental system that is influenced by a large number
of partly unknown or unobservable interventions, and let the function

$$\tilde{y} = f(u, v)$$

be a mechanistic model that describes how the natural forcings $u$ and the anthropo-
genic forcings $v$ influence the system under consideration. We can then try to re-
move the natural fluctuations of the measured state $y$ of the environment by first
computing model outputs

$$\tilde{y} = f(u, v_0)$$

for some constant level $v_0$ of the anthropogenic forcing and thereafter form residuals

$$y - \tilde{y}.$$ 

If the mechanistic model is correct and the model output can be written as a sum
\[ \bar{y} = f(u, v) = g(u) + h(v), \]

where \( u \) and \( v \) are statistically independent, the procedure just described can be justified (Grimvall 2001). However, in most cases it is not optimal.

First, it is rarely realistic to assume that a mechanistic model can mimic all dynamic properties of the system under consideration. In some cases, potentially important processes are omitted. In other cases, the model requires inputs that are not available in the desired temporal or spatial resolution. In yet other cases, several combinations of model parameters can produce practically identical outputs, because the model is over-parameterised.

Second, the anthropogenic and natural forcings of the studied system need not be independent. For example, the weather conditions can influence both the operation of power plants and the fate of the substances that are emitted during the production of electricity.

Finally, it is often overlooked that many mechanistic models of processes in the environment provide average outputs for grid cells, whereas empirical measurements typically refer to a specific point in the study area. Hence, all discrepancies between the modelled and observed data shown in Figure 1 need not be due to model or measurement errors. On the contrary, the model would be wrong if grid cell averages exhibited the same temporal variation as point data.

![Scatter-chart of observed and modelled nitrogen deposition vs. observed precipitation at Rörvik on the west-coast of Sweden, 1985-1996. The measured values represent a single point, whereas the modelled values represent grid cell averages obtained by the EMEP Lagrangian Acid Deposition Model.](image)
4. Procedures developed within IMPACT

The EU project IMPACT has identified three major roles of deterministic models in statistical normalisation procedures:

- Incorporation of model outputs as covariates
- Reduction of complex sets of covariates
- Derivation of reduced models for statistical normalisation

4.1 Model outputs as covariates

The simple method of forming differences
\[ y - \tilde{y} = y - f(u, v_0) \]

between observed values and model outputs corresponding to a fixed anthropogenic forcing \( v_0 \) can easily be generalised to a simple linear regression of the form

\[ y = \alpha + \beta \tilde{y} + e \]

where the intercept and slope parameters are determined by minimising the residual sum of squares. Similarly, the model outputs can be incorporated into arbitrary non-linear regression models

\[ y = k(\theta, \tilde{y}) + e \]

with parameters \( \theta \).

In a geostatistical context, outputs of mechanistic models can be used to introduce so-called external drift (Chilès/Delfiner 1999). If \( Z(x,t) \) and \( Z^*(x,t) \) respectively denote the actual and modelled state of the environment at location \( x \) and time \( t \), we may write

\[ Z(x,t) = b_0 + b_1 Z^*(x,t) + U(x,t) \]

where the residuals \( U(x,t) \) is a random function in space and time. Statistically this is regression model with correlated residuals. From a geostatistical perspective this equation implies that the drift of \( Z(x,t) \) is defined externally through an auxiliary variable \( Z^*(x,t) \) rather than from some smooth version of \( Z(x,t) \) itself. We may also regard the cited equation as a universal kriging decomposition into drift and fluctuation in which the basis drift functions are \( (1, Z^*(x,t)) \) instead of monomials in the geographic coordinates \( x \).

4.2 Reduction of complex sets of covariates

The most widespread techniques to normalise air quality data are based on regression analyses of contemporaneous observations of air quality and weather data from
a single station. Considering that local wind directions and other local weather observations provide imprecise information about the origin of the sampled air masses, this implies that much of the regional dynamics of air pollution is ignored in the predominant normalisation procedures. In principle, it would be possible to remove this weakness by extending the list of covariates to include both contemporaneous and time-lagged weather data from a whole network of stations. However, due to the large number of meteorological variables that may influence the response variable under consideration, such purely statistical procedures may remove important anthropogenic signals along with the natural fluctuations (Libiseller, 2002). This calls for procedures, in which the set of meteorological covariates is reduced prior to the statistical normalisation.

In an on-going study of normalisation techniques for background concentrations of surface ozone in Finland, a physics-based model (TRADOS) for the calculation of back-trajectories is used to reduce large sets of pressure field data. Different statistical characteristics of these back-trajectories are then used as covariates in regression-based normalisation models. Figure 2 illustrates that even if the information in the trajectories is reduced to a dummy variable for the crossing of the 70th degree of north latitude, it may still be relevant for the normalisation of ozone concentrations.

Fig. 2. Monthly mean values of ozone concentrations at Ähtäri, Finland, representing different source regions of the sampled air masses.
4.3 Derivation of reduced models

Even if a given mechanistic model explains a rather small fraction of the natural variability of the response variable under consideration, it may still be of great value in the search for efficient statistical normalisations. By feeding the model with different sets of inputs and regressing model outputs on model inputs, the inputs that have the strongest influence on the model outputs can be identified. Further analysis of relationships between model inputs and outputs may also reveal which resolution of the inputs that is needed and how long memory effects of previous inputs that should be taken into account in the normalisation. Moreover, the mechanistic model may draw attention to important nonlinearities of the system under consideration.

Figure 3 illustrates the results of a statistical analysis of nitrogen leaching and runoff data produced by feeding a soil nitrogen model (SOIL/SOILN; Johnson 1987) with a large set of artificially generated meteorological inputs. More precisely, ordinary least squares regression was used to regress annual values of the leaching of nitrogen from the root-zone on monthly runoff values, and the estimated regression coefficients represent impulse response weights for the 24 monthly runoff values during the current and previous year.

Fig. 3. Regression coefficients (impulse response weights) representing the response in annual SOILN-modelled nitrogen losses to monthly runoff values. All calculations refer to cultivation of barley on a loamy sand soil, and the three curves correspond to different depths in the soil.
Based on the results just described, it was concluded that annual riverine loads of nitrate in agricultural catchments may be normalised by using a sparsely parameterised regression model with two years of monthly runoff values as explanatory variables (Forsman, 2002). On a more general level, this example shows how a reduced model of a given mechanistic model can be utilised.

5. Conclusions

- Incorporation of mechanistic models into statistical normalisation procedures can greatly facilitate the extraction of anthropogenic signals from environmental data.
- Outputs of mechanistic models can play a key role as covariates in regression-based and geostatistical normalisation procedures.
- Even if the model under consideration can only explain a minor fraction of the variability of the response variable, it may play a key role in the search for reduced models that subsequently can be used for statistical normalisation.

Acknowledgement

The authors are grateful for financial support from the European Commission, grant IST-1999-11313.

Bibliography


