Sensitivity Analysis for Environmental Models

Alessandro Fassò,* University of Bergamo
Pier Francesco Perri, University of Milano-Bicocca

Gruppo di Ricerca per le Applicazioni della Statistica ai Problemi Ambientali

Working paper n.5 - July 4, 2000

Introduction

Generally speaking, sensitivity analysis (SA) concerns the mathematical model representation of a physical system, and attempts to assess the sensitivity of the model outputs to variations of model inputs given by variables or parameters and variations of model assumptions.

The environmental statistician is concerned with two classes of problems. $i)$ There is a usually complex mathematical or computer model (CM) able to simulate some environmental system by means of numerical input/output data. Suppose for a moment that it can be given in closed analytical form by

$$y = f(x_1, ..., x_k) = f(\mathbf{x}).$$

Hence the partial derivative $\frac{\partial f(\mathbf{x})}{\partial x_i}$ is the basic quantity for assessing the sensitivity of y to x_i near the point $\mathbf{x}$. From a practical point of view, f is unknown in a closed form and, because of its complexity, computing time is a matter of concern. Uncertainty about the inputs is considered known by means of a (probability) distribution $p(\mathbf{x})$; popular assumptions are the uniform or the lognormal distribution and independence among inputs. Of course, input uncertainty results in output uncertainty. Hence some synthetic

*Work partially supported by italian MURST-1998 grant.

sensitivity or influence measure is needed. Using simulated data, the first step in SA is "uncertainty analysis", which assesses the marginal output variability and distribution $p(y)$. While the second step is ranking input variables by their influence on y. This point is addressed in section "Sensitivity Analysis of Computer Models" below.
$ii)$ In the second setup, the environmental model is in fact a statistical model, e.g. multiple regression. Here SA is concerned with the particular model output given by the sample estimates, $\hat{\theta}$ say, and aims at assessing the influence of individual data, groups of data, variables and model assumptions to $\hat{\theta}$. This point is addressed in sections "Sensitivity Analysis of Statistical Models", "Sensitivity Analysis in Regression Models" and "Sensitivity Analysis in Time Series" below.

Sensitivity Analysis of Computer Models

SA of CMs has been developed over the last two decades together with massive computer simulations of engineering, chemical and environmental systems. As a matter of fact, CMs often need a lot of input parameters whose relevance is not a priori obvious and not easy to evaluate on simulated data. Through assessing the influence of each input to the model outputs, SA gives useful tools for understanding complex models, reducing analysis complexity and/or dimensionality by deleting unimportant inputs and, eventually, giving indirect evidence for validating a model with respect to the underlying real problem to be simulated.

A SA review of environmental CMs is given by Hamby [21]. Helton [25] discusses SA techniques with special reference to performance assessment for radioactive waste disposal, and reports several references to SA of CMs related to environmental issues. Among these are chemical kinetics, atmospheric chemistry, nuclear safety, ecosystem models, as well as hydrology and water pollution. A recent and comprehensive handbook on the subject of SA is given in Saltelli et al. [48].

The basic, introductory SA method has been discussed in atmospheric tritium dose estimates [1] and in environmental radiological assessment [29]. The inputs x_i are moved one-at-a-time (OAT) to their respective extreme values, $x_i^{\min}$ and $x_i^{\max}$, ceteris paribus, i.e. keeping all the other inputs x_j at their central or "base case" value $x_{0,j}$. In this way, using only $2k$ CM runs, we get a very preliminary ranking of input influence based on the following sensitivity index

$$SI_i = \frac{y_i^{\max} - y_i^{\min}}{y_i^{\max}},$$

where $y_i^{\max}$ is the maximum between $y\left(x_i^{\min}\right)$ and $y\left(x_i^{\max}\right)$, and $y(x_i)$ is $f(\mathbf{x})$ computed at the base case point $\mathbf{x}_0$ except ith input at x_i. In [14] AOT runs are proposed at $x_{0,i} \pm c\sigma_i$ which are appropriate for Gaussian input marginal distributions $p(x_i)$ with a standard deviation $\sigma_i = \sigma(X_i)$.

To get more informative SA, we can use one of two approaches, either the local SA or the global SA. The former is essentially a non-statistical method assessing output sensitivity to small input perturbations. It is based on differential analysis, and the model responses are obtained by the OAT approach.

The latter is inherently statistical since it uses output uncertainty analysis, which is based on output cdf and variability. It is primarily aimed at assessing input influence using some kind of output variance decomposition. These decomposition techniques with related sampling designs have increasing complexities and computing demands. The related models range from the simple linear model to the response surface with interaction model, nonparametric models together with Fourier and Monte Carlo approximations. The relevance of these models in SA will be discussed in the following subsections.

The model output considered so far is scalar. In various environmental CM applications, the model output is a possibly vector valued grid trajectory or surface. For example in [49] the SA of space-time air quality models is considered.

Local SA

Under suitable smoothing assumptions, a simple approach to the sensitivity of f can be given considering the second order Taylor expansion around the base case $\mathbf{x}_0$:

$$f(\mathbf{x}_0+\Delta) = f(\mathbf{x}_0) + \sum_{i=1}^{k} \frac{\partial f(\mathbf{x}_0)}{\partial x_i}\Delta_i + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_i \partial x_j}\Delta_i\Delta_j. \qquad (1)$$

For example, in chemical kinetics SA literature, e.g. [52] $\frac{\partial f}{\partial x_i}$ is called first order local sensitivity and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ second order local sensitivity. The exact computation of these quantities may be a complex task for large models requiring the solutions of large linear systems or specialized techniques like the Green function method. The OAT approach, with $k+1$ runs, may be used to get the finite difference approximation to first order local sensitivity, i.e.

$$\frac{\partial f}{\partial x_i} \approx \frac{y(x_{0,i}+\Delta_i) - y(x_{0,i})}{\Delta_i},$$

for small perturbation Δ_i, e.g. 1%.

If the first order truncation of equation (1) approximately holds over the full input range, i.e. the model is approximately linear, local sensitivities may then be used to assess output global uncertainty. For example, for uncorrelated and symmetric inputs, $\mathbf{X} = (X_1, ..., X_k)'$, with base case at the expectation vector, $\mathbf{x}_0 = E(\mathbf{X})$, we have

$$E(Y) = f(\mathbf{x}_0)$$

and

$$Var(Y) = \sum_{i=1}^{k} \left(\frac{\partial f(\mathbf{x}_0)}{\partial x_i}\right)^2 Var(X_i).$$

Hence, ratios $\left(\frac{\partial f(\mathbf{x}_0)}{\partial x_i}\right)^2 Var(X_i)/Var(Y)$ give importance measures.

A further step towards global SA is based on repeated OAT. According to this approach, we run the model on n random replicates of x_i from the marginal input distribution $p(x_i)$. We then compute importance measures based on output uncertainty due to x_i uncertainty, ceteris paribus. For example, we can use this data to compute a variance ratio

$$\frac{S(y(X_i))^2}{Var(X_i)}$$

where $S(y(X_i))^2$ is the sample variance of simulated ys due to X_i. Moreover, similarly to the coefficient of variation, we have an adimensional importance measure given by

$$\frac{S(y(X_i))}{\bar{y}(X_i)}$$

where $\bar{y}$ is the sample mean.

Then using some nk runs, these OAT statistics allow input ranking according to their importance without considering interactions among inputs. All are local in the sense that they use the base case as reference, and changing it may cause a change to conclusions.

Global SA

A (global) SA method usually aims at a SA index or measure for assessing and ranking the overall or average input influence under simultaneous input variations. This index is usually based on a statistical model simplifying the CM, and needs a certain amount of data or computer runs to be estimated.

The main stream of SA is based on importance measures resulting from Pearson's well known correlation ratio,

$$\eta_i^2 = \frac{Var(E(Y|X_i))}{Var(Y)}, \tag{2}$$

which gives the percentage output variation due to input X_i. In equation (2), the conditional expectation $m(x) = E(Y|X_i = x)$ is the regression function, possibly nonlinear, and $Var(m(X_i))$ is the variance of $m(X_i)$ using the marginal distribution $p(x_i)$. For high dimensional $\mathbf{X}_i$, $m(\mathbf{X}_i)$ is a regression surface and the problems arising in evaluating and estimating its variance will be discussed below. Robustness issues may arise for heavily tailed or skewed output distributions.

Unfortunately no reliable portmanteau method efficient for whatever CM is presently available and the choice depends greatly on the particular case under study. Good statistical modelling practice including diagnostic and validation should lead to the appropriate model and sampling solution. Because of these, global SA is closely related to sampling and Monte Carlo. These methods are based on a number n of CM runs giving simulated data $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)$ where $\mathbf{x}_m = (x_{m1}, ..., x_{mk})$, $m = 1, ..., n$, are random or quasi-random numbers from the distribution $p(\mathbf{x})$. Of course the distribution $p(\mathbf{x})$ defines the input dependence structure and its choice should be considered in an extended SA [16] as it may affect deeply output uncertainty.

Design of Experiment

A particular non-random sampling technique is based on the classical design of experiments (DOE) often coupled with response surface methodology. In physical, chemical and engineering experiments these techniques are well known, and references like [7] and [8] are classics.

In SA this approach has been used to obtain a surrogate model on the assumption that a linear plus interaction approximation to the CM is good enough. For example, in environmental SA, it has been used in the estimation of dose-equivalent to man from the concentration of radioactivity in air, water and food $[14]$.

The absence of the replication error of the classical model, $y = f(\mathbf{x}) + \varepsilon$, entails some differences between classical DOE and DOE for CMs $[46]$. For example, the least square method can be interpreted only as an interpolation method and residuals are systematic errors. Next, the classical DOE concepts of blocking, replicate and randomization become irrelevant.

Moreover, the stochastic nature of observed inputs is underscored in classical DOE, and high input dimensionality of CMs may prevent using standard DOE software. Special designs for CMs have therefore been developed for screening SA. For example in $[54]$, case studies with 20 input variables and $30 \div 50$ CM runs are considered.

Other developments use random and stratified sampling. Among these, Latin hypercube sampling gives good estimates of the output expectation, $\mu = E(f(\mathbf{X}))$, when f is monotonic in each coordinate. A univariate Latin hypercube sample of size n is a stratified sample in which the strata are the n equal probability intervals and the n sample values are randomly selected from each stratum using the distribution $p(x_i)$ conditional to the stratum. Multivariate definitions readily apply for independent vectors and appropriate techniques covering inputs with particular correlation structures have been applied in $[32]$ to the geological disposal of radioactive waste.

In the light of more powerful but more computation-demanding Monte Carlo SA below, simplified DOE and AOT sampling approaches may be useful in screening SA for a preliminary reduction of high dimensional inputs.

Correlation Methods

The simplest approximation to the general correlation ratio (2) is the square of Pearson's linear correlation coefficient

$$r(y, x_i) = \frac{\sum_m (y_m - \bar{y})(x_{mi} - \bar{x}_i)}{\sqrt{\sum_m (y_m - \bar{y})^2 \sum_m (x_{mi} - \bar{x}_i)^2}}.$$

The next step is to consider simultaneous input effects. The corresponding linear correlation coefficient is the partial correlation coefficient $[51]$ which, roughly speaking, is the correlation between y and x_i after eliminating the linear effect of the remaining inputs x_j, $j \neq i$, from both y and x_i.

Regression Methods

If two different inputs x_i and x_j have the same correlation with the model output, they may have different influences in the differential sense. Using a linear regression approximation,

$$y = \mathbf{x}'\beta + e, \tag{3}$$

we see that the first order local sensitivities of equation (1) can be estimated by the LS regression coefficients (see e.g. (17) below). Related measures, corrected for the uncertainty, are based on the standardized regression coefficients

$$\beta_i \frac{\sigma_i}{\sigma(Y)}.$$

Of course, these regression quantities are meaningful if model (3) has a good fitting, for example the multiple determination coefficient R^2 cannot be small.

In order to reduce the influence of non-normality and/or nonlinearity on these statistics some data transformation may be useful. For example, in environmental related risk analysis, the $\log$-transform [33] and rank-transform [32] have been used. The former is a popular transform in environmental statistics because of the log-normal distribution. The latter is distribution free for continuous distributions, and is capable of detecting input importance for nonlinear monotonic relations. Both are weak in detecting U-shaped or periodic relations, and both distort the original scale and curvature.

Graphical regression diagnostics are useful as complements to regression methods. Among these, in SA literature, bivariate scatterplots of inputs vs. output are widely used tools giving information on importance and linearity.

Nonlinear Sensitivity Analysis

The more complex and refined SA methods are based on nonparametric variance decompositions applied to nonlinear models. The idea of Sobol', as discussed in [30] and [9] is based on the following decomposition

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{k} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + ... + f_{1,2,...,k}(x_1, ..., x_k), \tag{4}$$

where $f_0 = E(f(\mathbf{X}))$, f_i are the first order factors, f_{ij} are the second order interactions and so on. These component functions are zero mean and uncorrelated, so that the corresponding variance decomposition in $2^k - 1$ summands holds, namely

$$Var(Y) = V = \sum V_i + \sum_{i<j} V_{ij} + ... \tag{5}$$

As in equation (2), the first order components,

$$V_i = E\left(f_i^2\right) = Var(E(Y|X_i)), \tag{6}$$

can be used to assess the first order nonlinear influence of X_i. Similarly V_{ij}s give higher order sensitivities by means of

$$S_{ij} = \frac{V_{ij}}{V}$$

and so on.

For estimating first order sensitivities, various sampling plans are available. To see these, consider input partitioning

$$\mathbf{x} = (x_i, \mathbf{x}_{(i)}) \tag{7}$$

where $\mathbf{x}_{(i)}$ is $\mathbf{x}$ without its ith element x_i.

The conceptually simplest, but computationally more demanding, sampling design requires estimating conditional means using r replicates for every x_{mi}, i.e.

$$\bar{y}_{mi} = \frac{1}{r} \sum_{t=1}^{r} f(x_{mi}, \mathbf{x}_{t(i)}).$$

This gives the $ANOVA$-like sensitivity estimate

$$\tilde{\eta}_i^2 = \frac{\tilde{V}_i}{\tilde{V}} = \frac{\frac{1}{n} \sum_{m=1}^{n} (\bar{y}_{mi} - \bar{y})^2}{\frac{1}{nr} \sum_{m=1}^{n} \sum_{t=1}^{r} f(x_{mi}, \mathbf{x}_{t(i)}) - \bar{y})^2},$$

where $\bar{y}$ is the grand mean. A way to reduce the number of runs nr for $\tilde{\eta}_i^2$ is to estimate directly $E(f_i^2)$ in equation (6) using two independent replicates, $\mathbf{x}_{m(i)}$ and $\mathbf{x}^*_{m(i)}$, for every x_{mi}. This gives

$$\hat{V}_i = \frac{1}{n} \sum_{m=1}^{n} f(x_{mi}, \mathbf{x}_{m(i)}) f(x_{mi}, \mathbf{x}^*_{m(i)}) - \hat{f}_0^2 \tag{8}$$

at the cost of $2n$ runs plus another m runs for $\hat{V}$ and $\hat{f}_0 = \frac{1}{n} \sum_m f(\mathbf{x}_m)$. Note that since both above V_i estimates have $\sqrt{n}$ rate of convergence using Monte Carlo sampling, $\hat{V}_i$ is consequently more efficient then $\tilde{V}_i$.

Second order terms are now estimated using

$$\hat{V}_{ij} = \frac{1}{n} \sum_{m=1}^{n} f(x_{mi}, x_{mj}, \mathbf{x}_{m(ij)}) f(x_{mi,} x_{mj}, \mathbf{x}^*_{m(ij)}) - \hat{V}_i - \hat{V}_j - \hat{f}_0^2$$

and corresponding formulae hold for higher order interaction estimates. Hence a full SA would require a number of runs of order $2n2^k$.

If we have $\sum_{i=1}^{k} V_i \geq V$, then the CM under study does not show relevant interactions. Unfortunately, this is not often the case in CM simulation. In order to avoid the above curse of dimensionality without losing too much information on interactions, a single total sensitivity index TSI_i has been considered. It takes into account all interaction-components in equation (4) which have the term x_i.

To see this, recalling input partition (7), rewrite equation (4) as

$$f(\mathbf{x}) = f_0 + f_i(x_i) + f_{(i)}\left(\mathbf{x}_{(i)}\right) + f_{i(i)}\left(x_i, \mathbf{x}_{(i)}\right)$$

and variance equation (5) as

$$V = V_i + V_{(i)} + V_{i(i)}.$$

Hence the total variance due to X_i is

$$V_i^{tot} = V_i + V_{i(i)} = V - V_{(i)}.$$

In the light of formula (8), $V_{(i)} + f_0^2$ is estimated using $2m$ runs as follows

$$\hat{V}_{(i)} = \frac{1}{m}\sum_{m=1}^{h} f\left(x_{mi}^*, \mathbf{x}_{m(i)}\right) f\left(x_{mi}, \mathbf{x}_{m(i)}\right) - \hat{f}_0^2$$

and the total sensitivity index is computed by

$$TSI_i = \frac{\hat{V} - \hat{V}_{(i)}}{\hat{V}}. \tag{9}$$

The accuracy of these estimates can be assessed by their asymptotic Monte Carlo standard deviations or confidence intervals. This is easily done, noting that $\hat{f}_0$ and $\hat{V}_i + \hat{f}_0^2$ statistics are essentially sample means. Hence, using standard sampling results, we have the following estimates without additional CM runs:

$$Var\left[\hat{f}_0\right] \approx \frac{1}{m}\left[\frac{1}{m}\sum_{m=1}^{h} f\left(x_{mi}^*, \mathbf{x}_{m(i)}\right)^2 - \hat{f}_0^2\right]$$

and

$$Var\left[\hat{V}_i\right] \approx \frac{1}{m}\left[\frac{1}{m}\sum_{m=1}^{h} f\left(x_{mi}, \mathbf{x}_{m(i)}\right) f\left(x_{mi}^*, \mathbf{x}_{m(i)}\right) - \left[\hat{V}_i + \hat{f}_0^2\right]^2\right].$$

FAST Method

The above method with Monte Carlo sampling is quite general and little model-assumption demanding. As a matter of fact, it gives model-free importance measures with estimated standard deviations, output uncertainty and cdf without using surrogate models. The main drawback may be the large number of CM runs required.

An approach giving similar importance measures and requiring a smaller number of CM runs is based on the Fourier amplitude sensitivity test $(FAST)$ [47]. This technique is based on the following Fourier approximation

$$E(Y) \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} F(s)\, ds \tag{10}$$

where

$$F(s) = f(g_1(s), ..., g_k(s))$$

and

$$x_i = g_i(s) = G_i(\sin \omega_i s)$$

for appropriate functions G_i and frequencies ω_i. Approximation (10) and its companion $Var(Y) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} F(s)^2 ds - E(Y)^2$, are interesting because they reduce a multiple integral into a one-dimensional integral.

Moreover, using the Fourier series representation for $f(s)$, we get

$$Var(Y) \approx \sum_{h=1}^{\infty} \left(A_h^2 + B_h^2 \right)$$

and

$$V_i \approx \sum_{h=1}^{\infty} \left(A_{h\omega_i}^2 + B_{h\omega_i}^2 \right) \tag{11}$$

where A_h and B_h are the Fourier coefficients given by

$$A_h = \frac{1}{\pi} \int_{-\pi}^{\pi} F(s) \cos(js)\, ds$$

and

$$B_h = \frac{1}{\pi} \int_{-\pi}^{\pi} F(s) \sin(js)\, ds.$$

Hence, $FAST$ method is essentially based on computing the quantities A_h and B_h via appropriate numerical and/or Monte Carlo one-dimensional integration based on the same, N say, CM runs.

$FAST$ method does not give practical methods for second and higher order interactions, but extended $FAST$ [47] allows computing total sensitivity indexes (9) without additional CM runs. To do this, extended $FAST$ assigns a frequency ω_i to x_i and a common frequency $\omega_{(i)} \neq \omega_i$ to all other inputs. Hence, $V_{(i)}$ is estimated via equation (11) filled with frequency $\omega_{(i)}$.

Summing up, $FAST$ is an efficient tool for global SA but its proper use and understanding require a level of mathematics which may be difficult to explain.

Sensitivity Analysis of Statistical Models

In statistics, sensitivity analysis (SA) can be considered a set of tools aiming at ascertaining the degree of dependence of the results to the hypotheses underlying the particular techniques employed (e.g. normality in the regression model) or to nature of the data

(presence of atypical, unusual observations). A lack of sensitivity to such aspects is known as robustness. In other words, SA tries to evaluate how the outcome of the (statistical) analysis changes when perturbations are introduced in the problem formulation.

Considering the features in the data, SA looks for observations mostly affecting some or all of the aspects of the analysis. In this context, a subset of data is said to be influential if its omission causes substantial changes in the estimates, confidence regions, tests, etc. Usually, influential observations appear to be outside the pattern defined by the bulk of the data because gross errors can occur in recording or processing the data or because the underlying distribution that generates the data is different from the hypothesized one.

In model building, finding influential observations can be useful for the analyst in at least two ways. In the first place, influential cases may highlight gross measurements and recording or processing errors. Secondly, finding many influential observations could suggest that the adopted model is inadequate for describing the real problem under study.

Influential observations may become noninfluential if the model is modified (e.g. adding or trasforming the variables, selecting a different model) or if alternative fitting methods are used (robust or bounded-influence).

Influence Curves

A general tool for measuring the influence of an observation is represented by the influence curve (or function) introduced by Hampel ([22][23]), further aspects of which are investigated in [24][28] [31].

The main idea of influence curve is intuitively easy. Let $x_1, x_2, .., x_n$ be a sample of independent observations from the cdf F. Let $T_n = T(x_1, x_2, .., x_n)$ be a (Fisher) consistent estimator of the unknown (scalar or vectorial) parameter $\theta = T(F)$ expressed as a functional of F (e.g. the population mean can be written $T(F) = E(X) = \int xdF$). Analogously, let $T_n = T(F_n)$ be the estimator expressed as a functional of the empirical cdf F_n (e.g. the sample mean $n^{-1} \sum_{i=1}^{n} x_i$ can be written $T(F_n) = \int xdF_n$).

In order to evaluate the influence on T_n of a new observation x added to a large sample (i.e. $n \to \infty$), the influence curve of the estimator T_n is defined, under suitable conditions, as:

$$IC(x; T, F) = \lim_{\epsilon \to 0} \frac{T\left[(1 - \epsilon) F + \epsilon \delta_x\right] - T(F)}{\epsilon}, \tag{12}$$

where δ_x is the cdf of a degenerate variable which gives mass 1 to x. In other words, the influence curve assesses the variation in T caused by a small contamination of F at the value x.

The influence curve indicates the sensitivity of the estimator to an individual observation and describes its local behavior. Moreover, the influence curve provides a tool for evaluating the robustness of an estimator with respect to atypical data: estimators with a bounded influence curve can be considered almost insensitive to outliers.

The previous definition (12) is entirely asymptotic because it focuses on the infinitesimal influence that an observation has on the asymptotic value of the estimator. Nevertheless, for practical purposes, various finite sample versions of the influence curve are available, for example in [24]. The simplest finite approximation is the so-called empirical

influence curve, obtained by adding x to the sample $(x_1, ..., x_{n-1})$ and plotting the values of the estimator $T_n(x_1, ..., x_{n-1}, x)$ against different values of x.

Here we focus on a more useful version given by the following sensitivity curve (SC):

$$SC_n(x) = n\left[T_n(x_1, ..., x_{n-1}, x) - T_{n-1}(x_1, ..., x_{n-1})\right]. \quad (13)$$

If the estimator is expressed as functional of the empirical cdf, equation (13) becomes:

$$SC_n(x) = n\left[T\left((1-\frac{1}{n})F_{n-1} + \frac{1}{n}\delta_x\right) - T(F_{n-1})\right]. \quad (14)$$

Note that this last expression can be also obtained from equation (12) replacing ϵ with n^{-1} and omitting the limit.

A slight variant of expression (13), widely used in SA aiming to evaluate the influence on T_n of the observation x, when it is deleted from the sample, is the sample influence curve (SIC), defined as:

$$SIC(x) = (n-1)\left[T_n(x_1, ..., x_{n-1}, x) - T_{n-1}(x_1, ..., x_{n-1})\right]. \quad (15)$$

As an example, consider a symmetric distribution with its centre of symmetry given by θ. In this case, we may use the sample mean or the sample median, widely used in environmental statistics, as estimators of θ. The SCs of the two estimators are represented in figure 1 and figure 2, respectively.

Observe that the SC for the mean is unbounded when $x \to \pm\infty$, that is the mean is extremely sensitive to a single large outlier. On the contrary, the SC for the median is bounded. Thus, the median appears to be more robust than the mean with respect to the outliers, but it is still sensitive to rounding and grouping errors around central values. In order to overcome this drawback, the broadened median [28] may be considered.

The sample median may be more efficient than the mean when dealing with distributions characterized by tails heavier than the normal distribution. A natural compromise between the mean and the median is represented by both the trimmed and winsorized means [28]. The first estimator does not consider the extreme observations, while the second gives them different weights.

Sensitivity Analysis in Regression Models

The elements defining the regression model are the underlying hypotheses, variables and observations. The SA (or influence study) for regression models consists in monitoring the changes in the final results when the elements of the model are perturbed.

The main issues in designing methods for regression SA are the choice of perturbation scheme, the aspects of the model needing checking and the way to assess and measure any influence.

If we focus on the data, an interesting method for perturbation is to delete observations, individually or in groups. This approach is known as case-deletion and it aims to evaluate the influence of an observation, or observation subset, on some aspects of the regression analysis. The observations that, either individually or in groups, have a

demonstrable impact on some or all aspects of the analysis, are called in‡uential. This de…nition tends to be subjective: it does not provide a way to measure the in‡uence of an observation and neither does it clarify which aspects of the in‡uence should be investigated. In fact, an observation may in‡uence the vector of parameters estimated, a linear combination of these parameters, the estimation of the variance, …tted values, con…dence intervals as well as the goodness-of-…t statistics, though it may well have a di¤erent e¤ect on each of these. Thus, the question "in‡uence on what?" plays a central role in designing in‡uence diagnostics. Many measures of in‡uence are proposed. Chatterjee & Hadi [10] suggest two approaches and discuss di¤erent methods for each of them.

In the regression model the study of in‡uence produces information about the reliability of the conclusions. The detection of in‡uential observations could suggest an inadequacy of the model or a violation of the underlying hypothesis ([40][53]).

We consider the usual multiple regression model:

$$\mathbf{Y} = \mathbf{X}\bar{} + ² \tag{16}$$

where $\mathbf{Y}$ is a n £ 1 vector of values of the response (dependent) variable; $\mathbf{X}$ is a n £ p full-column rank matrix of known regressors (or explanatory variables, factors and carriers) usually including a constant column of ones; $\bar{}$ is a p £ 1 vector of unknown parameters (or coe¢cients) to be estimated; ² is a n £ 1 vector of errors, independent and normally distributed with zero mean and variance σ^2. The ith observation will be indicated with $(\mathbf{x}_i', y_i)$, where $\mathbf{x}_i'$ represents the ith row of $\mathbf{X}$ and y_i is the ith element of $\mathbf{Y}$. Readers interested in regression models can consult [15][37][50][53].

Let

$$\hat{\bar{}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{17}$$

be the least squares (LS) estimator. We call residual the quantity $r_i = y_i - \hat{y}_i$, where $\hat{y}_i = \mathbf{x}_i'\hat{\bar{}}$ is the ith estimated value of the dependent variable. In the context of regression diagnostics, observations with high residuals are considered vertical outliers or outliers on the response variable. These observations may generally in‡uence the intercept of the model. Nevertheless, in order to …nd the outliers, di¤erent forms of residual standardization may be used to obtain more suitable diagnostics. The most common are: standardized residuals $r_i^* = r_i/\hat{\sigma}$, where $\hat{\sigma}$ is the standard error of regression; internally studentized residuals $t_i = r_i/(\hat{\sigma}\sqrt{1 - h_{ii}})$, where h_{ii} is the ith diagonal element of the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, or externally studentized residuals $t_{(i)} = r_i/(\hat{\sigma}_{(i)}\sqrt{1 - h_{ii}})$, where $\hat{\sigma}_{(i)}$ is the standard error of regression computed without the ith observation. As a rule of thumb, an observation with residual, in absolute value, larger than 2.5 can be considered an outlier on the response variable.

In the study of in‡uence, the elements h_{ij} of the hat matrix play a central role. They have a double interpretation. First of all, since $\hat{y}_i = \sum_{j=1}^{n} h_{ij}y_j$, it follows that $\dfrac{\partial \hat{y}_i}{\partial y_j} = h_{ij}$, and thus h_{ij} can be considered as the amount of leverage each observed value y_j has in determining the …tted $\hat{y}_i$. We refer to a point $(\mathbf{x}_i', y_i)$ with a high h_{ii} as leverage point, and generally we call h_{ii} the (potential) leverage of the data point $(\mathbf{x}_i', y_i)$.

From the properties of the hat matrix widely discussed in [10] it follows that in the regression model without intercept, $0 \leq h_{ii} \leq 1$. Then, if $h_{ii} = 1$, $\hat{y}_i$ is determined only

by y_i and the regression hyperplane passes through y_i while, if $h_{ii} = 0$, then y_i does not have any influence on $\hat{y}_i$ which takes zero value. Moreover, for the regression model with intercept, we find that $n^{-1} \leq h_{ii} \leq 1$.

The elements h_{ii} also have an interesting geometric interpretation. In the regression model with intercept term ([42][53]), the leverage is related to squared Mahalanobis distance (MD) of the nonconstant regressors by the relation $h_{ii} = \frac{1}{n-1} MD_i^2 + \frac{1}{n}$. It then represents in the $(p-1)$ dimensional space a remoteness measure of the point $\mathbf{x}_i'$, without the constant term 1, from the centroid of the nonconstant regressors. A high value of h_{ii} indicates that $\mathbf{x}_i'$ lies far from the bulk of other points. The leverage is then used as a diagnostic measure in order to identify (horizontal) outliers in the factor space.

Observations which appear outlying in the factor space are said good leverage points if the point $(\mathbf{x}_i', y_i)$ is in the main data trend. Otherwise these atypical observations are said bad leverage points, and usually have a strong influence on several aspects of regression analysis. Rousseeuw & Leroy [42] call the bad leverage points and vertical outliers regression outliers. A typical cut-off value for the leverage is $2p/n$.

If we focus on the influence that a single observation has on the LS estimator, we may use the influence curve or one of its sample versions, discussed in depth in [10][13]. A widely used finite sample version of influence curve representing the influence of the ith observation on the LS estimator is the finite sample influence curve (SIC), already defined by equation (15). In the regression model it becomes:

$$\begin{aligned} \mathbf{SIC}_i &= (n-1)(\hat{\beta} - \hat{\beta}_{(i)}) \\ &= \frac{(n-1)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i r_i}{1-h_{ii}}, \qquad i = 1, 2, \ldots n, \end{aligned} \tag{18}$$

where $\hat{\beta}_{(i)}$ is the LS estimated vectors of coefficients obtained after deleting the ith case $(\mathbf{x}_i', y_i)$.

Since the $\mathbf{SIC}_i$ is a vector of size p, a direct comparison between $\hat{\beta}$ and $\hat{\beta}_{(i)}$ is not possible. Then, in order to characterize the influence and rank the observations according to their influence, it may be useful to use a scalar version of the SIC. A well known scalar measure of influence is Cook's distance ([11] [12]), defined as:

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' \mathbf{X}'\mathbf{X} (\hat{\beta} - \hat{\beta}_{(i)})}{p\hat{\sigma}^2}. \tag{19}$$

It expresses the impact of the ith observation on the LS estimated coefficients vector. An interesting expression of Cook's distance equivalent to the above one but taking into account the sensitivity of the fitting and prediction capability model, is:

$$D_i = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})' (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})}{p\hat{\sigma}^2}. \tag{20}$$

Here $\hat{\mathbf{Y}}_{(i)}$ is the $n \times 1$ estimated values vector of the dependent variable obtained without the ith case. In this form D_i can be interpreted (except for a scale parameter) as the squared Euclidean distance between the two fitted vectors $\hat{\mathbf{Y}}_{(i)}$ and $\hat{\mathbf{Y}}$.

Another expression of equation (19) computationally more convenient and rich in information, because it puts in evidence the link with the regression outliers, is:

$$D_i = \frac{1}{p} t_i^2 \frac{h_{ii}}{1 - h_{ii}}. \tag{21}$$

In this form D_i is a function of the internally studentized residual t_i, which reflects how well the model fits the y_i, and of h_{ii} which measures the outlyingness of the point $\mathbf{x}_i'$ in the factor space. This statistic provides, therefore, a global measure of the influence of the ith observation. For example, in hydrological assessment, Becciu et al. [2] use Cook's distance to model the minimum in-stream flow in drainage basins of the Central Alps. Cook ([11][12]) suggests considering as heuristic threshold value the percentage point of a central Fisher distribution with p and $n - p$ degrees of freedom. More simply, Cook & Weisberg [13] retained a cut-off of 1.

We can find some other well known diagnostics based on the single-case-deletion approach, such as $DFFITS_i$, $DFBETAS_i$, in [3][10][13]. These methods have been widely accepted because of their computational simplicity, intuitive aspects and for being available in the most common statistical packages.

Other different approaches may be developed modifying the perturbation scheme and/or using different measures from the influence curves. For classical diagnostics, interested readers are referred to [10][13]. In estimating the sources of airborne particulate matter, Kim & Henry [34] propose some original nondeletion regression diagnostics for identifying influential species in the chemical mass balance receptor model.

The methods investigated so far are useful for assessing the influence of a single observation. In principle, the assessment of the influence of multiple observations can be obtained by generalizing the influence diagnostics previously introduced. For instance, if $\mathbf{I} = (i_1, i_2, .., i_k)'$ with $1 \cdot i_j \cdot n$ is a vector of the indices of selected cases, a natural extension of Cook's distance is:

$$D_{\mathbf{I}} = \frac{(\bar{} - \hat{\bar{}}_{(\mathbf{I})})'\mathbf{X}'\mathbf{X}(\bar{} - \hat{\bar{}}_{(\mathbf{I})})}{p\hat{\sigma}^2}. \tag{22}$$

If we do not have any prior knowledge about the number k of influential observations in the data set, all possible subsets of observations should be investigated. When the size of the data set is large, this procedure may become unfeasible. Moreover, more complex situations may occur for which a deeper study is needed. For example, an observation not individually influential becomes so if considered in groups. On the other hand, observations which are individually influential may not be if considered together with others. This situation is represented in figure 3.

A way to obtain some information on the possible number of influential observations is to look for regression outliers appearing in the data and then fix k equal to the number of outliers discovered. If a data set contains a single outlier or influential observation the problem of identifying such an observation is relatively simple: traditional regression diagnostics of remoteness or influence, based on the LS residuals and on the elements h_{ii}, work well. Nevertheless, detecting multiple outliers by means of the traditional LS diagnostics could become more difficult (see e.g. [42]). This is the consequence of masking

and swamping effects. Masking occurs when a subset of outliers is not correctly detected because it is closed to another group of observations. Swamping occurs when typical observations are incorrectly identified as outliers because of the presence of another outlying subset of observations. These two effects occur because the LS method is sensitive both to high leverage points and vertical outliers.

To overcome the problem of correctly identifying outlying observations, other measures not based on the traditional LS quantity are needed. The idea is to use robust estimation methods not influenced by outliers and, on the basis of the results obtained, building new diagnostics.

The most common robust estimation methods fall into three classes: M-estimators, L-estimators and R-estimators and, for each of them, many estimators are proposed. For the robust regression methods the reader is referred to [4][35][42] while for robust methodology in general to [24][28][31][55].

The degree of robustness of an estimator is measured by the breakdown point, which represents, roughly speaking, the largest proportion of outliers that can occur in the data without making the value of the estimator arbitrarily unbounded. Because of the high sensitivity to regression outliers, the LS estimator has a breakdown point equal to 0%. The M, L and R estimators are instead more robust with regard to vertical outliers, but are still sensitive to leverage points. In any case, their breakdown point is less than 30%. The maximum possible breakdown point is 50%. This value is achieved by least median of squares (LMS) and least trimmed squares (LTS) methods of estimation. For a complete description of these methods the reader is referred to [41] and [42] while for further developments to [43][44][45]. A computer program, PROGRESS, allows the calculation of the LMS or LTS coefficients for (robust) regression model and a robust version of the residuals. Moreover, Rousseeuw & Van Zomeran [43] propose a classification rule for the observations, based on the plot of the LMS robust residuals against a robust version of the Mahalanobis distance, which can be obtained by the computer program MINVOL. The two mentioned programs are available and they have been incorporated for example in S-Plus and SAS/IML software packages.

Even if the LMS and LTS methods have some drawbacks ([17][20][26][42]) many examples ([40][42][43][44]) show that these methods could be used as exploratory tool for detecting and classifying atypical data.

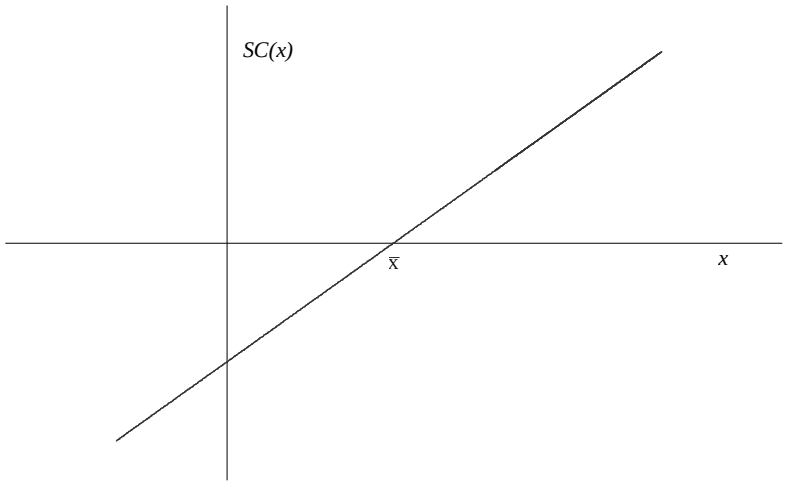

Figure 1: Unbounded sensitivity curve for the sample mean.

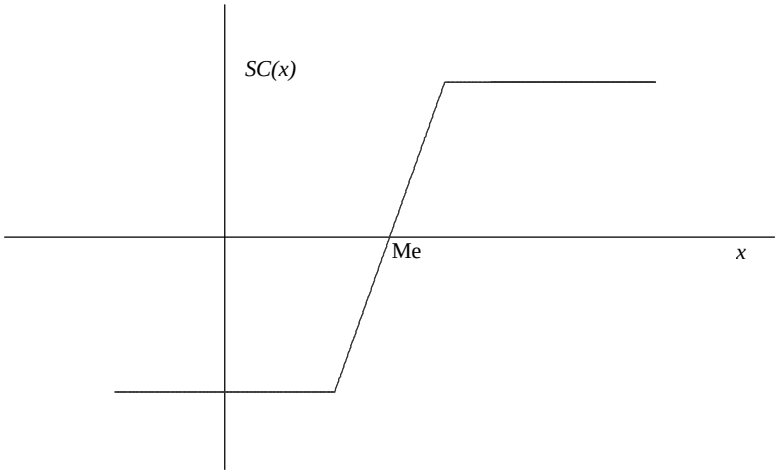

Figure 2: Bounded sensitivity curve for the sample median.

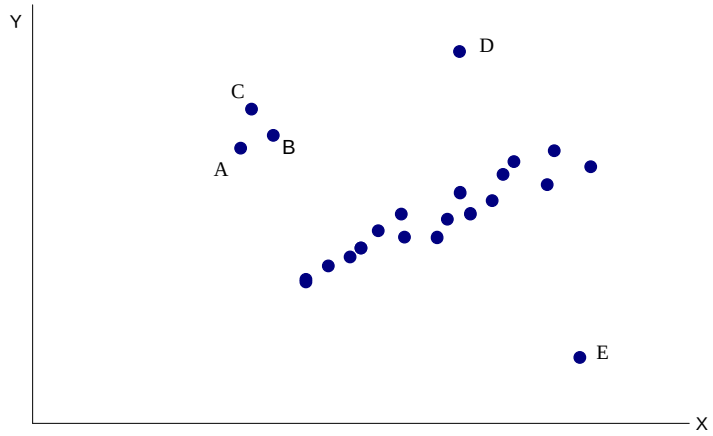

Figure 3: Observations A, B and C are influential in group but not individually. Observations D and E are individually influential, but not together.

Sensitivity Analysis in Time Series

Time series analysis plays a central role in environmental statistics since data collected over time are widely used and seasonality, trend, forecasting and monitoring are often relevant topics.

SA for environmental time series is important because if some data are very influential for the model at hand, we find ourselves facing one of the following two points: $i)$ data quality is low and the estimated model is not robust to outliers; $ii)$ the model is inadequate for the data under study.

In this section, we will consider SA methods based on influence curves and Cook's distance. As long as time series models are related to regression models, the corresponding SA methods show similarities and differences. In particular, the difference among horizontal and vertical outliers is not immediate as in the regression setup.

Nonparametric Influence Measures

In this section, we follow Lefrançois [36] and consider an influence measure for the autocorrelation function allowing indication for influential observations at an early stage of model building. Strictly speaking, the outlier definition is model dependent. Nevertheless, the nonparametric approach of this section is useful because outliers are often influential for the entire class of models under study. Moreover, searching for influential observations in this way is computationally cheaper than the parametric approach of next section where, using the single-case-deletion approach the model have to be re-estimated n times.

Let Y_t, $t = 1, 2, ...$ be a regular stationary Gaussian process [27] with autocorrelations ρ_h and sample autocovariances and autocorrelations given by c_h and r_h, respectively. Considering the influence for the single parameter ρ_h due to the ith observation, a statistic equivalent for large samples to the sample influence curve (15) is

$$SIC_{i,h} = Z_i \left(Z_{i+h} + Z_{i-h} - r_h Z_i \right),$$

where $Z_t = \dfrac{Y_t - \bar{Y}}{c_0}$. The influence for the autocorrelation function truncated at lag L is then collected by the vector

$$\mathbf{SIC}_i = (SIC_{i,1}, ..., SIC_{i,L})'$$

and summarized by the quadratic form

$$QIC_i = (\mathbf{SIC}_i - Z_i \hat{\tau}_i)' \, \hat{\Psi}_i^{-1} \, (\mathbf{SIC}_i - Z_i \hat{\tau}_i),$$

where $\hat{\tau}_i$ and $\hat{\Psi}_i$ are matrix functions of the sample autocorrelations $r_1, ..., r_L$. Using the known asymptotic distribution of QIC, we can then test for influential observations and for the most influential observation at an early stage of model building.

Influence Measures for ARIMA Models

Let Y_t, $t = 0, \pm 1, \pm 2, ...$ be an $ARIMA(p, d, q)$ standard process [27] given by the polynomial equation

$$a(B)\ \nabla^d Y_t = c(B)\, A_t$$

or the equivalent difference equation

$$\nabla^d Y_t = \sum_{i=1}^{p} a_i \nabla^d Y_{t-i} + \sum_{i=1}^{q} c_i A_t.$$

In these equations, $a(B) = 1 - a_1 B - ... - a_p B^p$ and $c(B) = 1 - c_1 B - ... - c_q B^q$ are polynomial operators with roots outside the unit circle. The backshift operator B is such that $Y_{t-1} = BY_t$ and $\nabla^d = (1 - B)^d$. Moreover $\nabla^d Y_t$ is a zero mean stationary process and A_t is an independent $N(0, \sigma^2)$ sequence.

Consider the model of additive outlier at time i for which Y_t^* is the true $ARIMA$ process but we observe

$$Y_t = \begin{cases} Y_t^* & t \neq i \\ Y_t^* + w_i & t = i. \end{cases} \tag{23}$$

This model is useful e.g. in environmental high frequency time series analysis as environmental data are often affected by measurement errors [5]. Together with the innovational outlier model, the above additive outlier model is a particular case of the general intervention analysis model which Box and Tiao [6] used to assess the impact of the new law of gasoline composition on tropospheric ozone in Los Angeles City in the 1966.

Additive and innovational outlier models may be handled by iterative detection-estimation techniques, [27]. On-line statistical detection of measurement error outliers is applied by [5] to hourly ozone data in downtown Bologna, Italy. Here we consider off-line or retrospective influence measures for single outliers.

In analogy with influence measures of previous sections, Peña [38] introduced a Cook-type distance. To see this, let

$$Y_t^* = \sum_{j=1}^{h} \pi_j Y_{t-j}^* + A_t \tag{24}$$

be the AR approximation of the above $ARIMA(p, d, q)$ process for appropriate h. Considering the regression interpretation of equation (24), under the non-outlier hypothesis, $w_i = 0$, we have $Y_t = Y_t^*$ and $\hat{\pi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the standard conditional maximum likelihood estimate of $\pi = (\pi_1, ..., \pi_h)'$ where $\mathbf{X}$ is the regressor-type $(n - h) \times h$ dimensional matrix with (g, j) th element given by Y_{h+g-j} and $\mathbf{Y}$ is the response vector with j th element given by Y_{h+j}. Moreover, let $\left[\hat{\pi}_{(i)}, \hat{w}_i\right]$ be the corresponding unconstrained conditional maximum likelihood estimate, where $\hat{\pi}_{(i)}$ corresponds to the single-case-deletion approach of previous section. As a matter of fact, using Kalman filtering techniques it can be shown that, [39] $\hat{\pi}_{(i)}$ is the same as the conditional maximum likelihood estimate

when Y_i is missing. Then as influence statistic, Peña suggests the following standardized Mahalanobis distance

$$D_i = \frac{\left(\hat{\gamma} - \hat{\gamma}_{(i)}\right)' \left(\mathbf{X}'\mathbf{X}\right) \left(\hat{\gamma} - \hat{\gamma}_{(i)}\right)}{(p+q)\,\hat{\sigma}^2}, \tag{25}$$

where $\hat{\sigma}^2$ is the usual one step prediction error variance estimate.

Like Cook's distance in standard regression (19), Peña's distance is based on similar SIC quadratic form and has similar interpretation for assessing the observation influence to the forecasts. In particular, the one step forecasts using the model with and without outliers and the full regressor data set are given by $\hat{\mathbf{Y}} = \mathbf{X}\hat{\gamma}$ and $\hat{\mathbf{Y}}_{(i)} = \mathbf{X}\hat{\gamma}_{(i)}$. Note that $\mathbf{X}$ is the same in both equations and, strictly speaking, $\hat{\mathbf{Y}}_{(i)}$ is not the forecast from an additive outlier model where the Y_i elements in $\mathbf{X}$ would be corrected for $\hat{w}_i$. Now, the Peña statistic is proportional to the squared Euclidean distance between the above prediction vectors, namely

$$D_i = \frac{\left(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)}\right)' \left(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)}\right)}{(p+q)\,\hat{\sigma}^2}. \tag{26}$$

Moreover, a decomposition (21) can be defined which is conceptually similar to but less simple than (21).

Plotting this statistic against i gives a simple tool for detecting influential outliers. Unfortunately, no direct reference to a model independent hat matrix is available and, unlike in the standard regression case (21), here, for each $i = 1, ..., n$, we need to re-estimate $\hat{\gamma}_{(i)}$ or the related $ARIMA$ parameters $\left(\hat{\mathbf{a}}_{(i)}, \hat{\mathbf{c}}_{(i)}\right)$ and the computation time for the above mentioned D_i plot may be relevant.

Note that, the corresponding Cook's statistic for innovational outliers was shown by [38] to give an unreliable procedure.

Influence Measures for Dynamic Regression Models

In environmental time series analysis, often covariates and regression with correlated errors have to be used. Examples are water pollution monitoring in Niagara River [18], and tropospheric ozone concentration model with meteorological variables and precursors [19].

An environmental protection problem is concerned in [39] where, considering a gas furnace, the relation among the output carbon dioxide concentration and gas-feed rate (U_t) is assessed through the classical transfer function model [27]. This can be written as

$$Y_t^* = \nu(B)\,U_t + \frac{1}{\pi(B)} A_t, \tag{27}$$

with rational operators $\nu(B)$ and $\pi(B)$ of order (m, a) and (p, q) respectively. Paralleling the $ARIMA$ case (24), equation (27) can be written as a lagged regression, namely

$$Y_t^* = \sum_j \pi_j Y_{t-j}^* + \pi(B) \sum_j \nu_j U_{t-j} + A_t.$$

Then, the forecast distance (26) for the full parameter $\mu = (¼, °)$ can be decomposed into an influence measure for ¼, an influence measure for ° and an interaction term, namely

$$D_i(\theta) = c_1 D_i(¼) + c_2 D_i(° \mid ¼) + INT_i(¼, °).$$

Here, $D_i(\pi)$ is the same as in equation (25). Moreover, using maximum likelihood estimates $\hat{v}$ for the full data set and $\hat{v}_{(i)}$ for the one-case-deletion, we have

$$D_i(° \mid ¼) = \frac{\left(\hat{°} - \hat{°}_{(i)}\right)' \hat{\Sigma}_°^{-1} \left(\hat{°} - \hat{°}_{(i)}\right)}{(m+a)}, \tag{28}$$

where $\hat{\Sigma}_°$ is the estimated variance-covariance matrix of °. This quantity gives the change in the transfer function given the noise structure ¼. Finally, the interaction term INT_i is given by cross-product type computation as in [39].

Using decomposition (28), we can then assess the separated influences to the dynamic component and the transfer function. Moreover, $D_i(\theta)$ generalizes both the $ARIMA$ and the standard regression cases of previous sections giving them as particular cases when $° = 0$ and $¼ = 0$ respectively.

References

[1] Bauer, L.R., Hamby, D.M. (1991) Relative sensitivities of existing and novel model parameters in atmospheric tritium dose estimates. Radiology and Prot. Dosimetry. 37, 253-260.

[2] Becciu, G., Bianchi, A., Fassò, C.A., Larcan, E., Fassò, A. (2000) Quick calculation of minimum in-stream flow in drainage basins of Central Alps. In Proceeding of New trends in water environmental engineering for safety and life: eco-compatible solutions for aquatic environments. Capri, July 3-7, 2000. (CD-ROM). A.A. Balkema, Rotterdam.

[3] Belsley, D.A., Kue, E., Welsch, R.E. (1980) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York.

[4] Birkes, D., Dodge, Y. (1993), Alternative methods of regression. Wiley, New York.

[5] Bordignon, S., Scagliarini, M. (2000) Monitoring algorithms for detecting changes in the ozone concentrations. Environmetrics. 11, 2, 125-137.

[6] Box, G.E.P. and Tiao, G.C. (1975) Intervention analysis with applications to economics and environmental problems. Journal of the American Statistical Assosciation, 70, 349, 70-79.

[7] Box, G.E.P., Hunter, W.G., Hunter J.S. (1978) Statistics for experimenters: an introduction to design, data analysis and model building. Wiley, New York.

[8] Box, G.E.P., Draper (1987) Empirical model-building and response surfaces. Wiley, New York.

[9] Chan, K., Tarantola S., Saltelli, A., Sobol', I. (2000) Variance-based methods. In Sensitivity Analysis, Saltelli, A., Chan, K., Scott, M. eds. Wiley, New York.

[10] Chatterjee, S., Hadi, A.S. (1988) Sensitivity analysis in linear regression. Wiley, New York.

[11] Cook, R.D. (1977) Detection of in‡uential observations in linear regression. Technometrics, 19, 1, 15-18.

[12] Cook, R.D. (1979) In‡uential observations in linear regression. Journal of the American Statistical Association, 74, 365, 169-174.

[13] Cook, R.D., Weisberg, S. (1982) Residuals and in‡uence in regression. Chapman and Hall, London.

[14] Downing, D.J., Gardner, R.H., Hoffman, F.O. (1985) An examination of response-surface methodologies for uncertainty analysis in assessment models. Technometrics. 27, 2, 151-163.

[15] Draper, N.R., Smith, H. (1981) Applied regression analysis. 2nd edition, Wiley, New York.

[16] Easterling, R.G. (1986) Discussion of Downing, Gardner and Hoffman (1985). Technometrics, 28, 1, 91-93.

[17] Ellis, S.P. (1998) Instability of least squares, least absolute deviation and least median of squares linear regression. Statistical Science, 13, 4, 337-350.

[18] El-Shaarawi, A.H. (1993) Environmental monitoring, assessment and prediction of change. Environmetrics, 4, 4, 381-398.

[19] Galbally, I.E., Miller, A.J., Hoy, R.D., Ahmet, S., Joynt, R.C., Attwood, D. (1986) Surface ozone at rural sites in the Latrobe valley and Cape Grim, Australia. Atmospheric Environment, 20, 2403-2422.

[20] Hadi, A.S., Simonoff, J.S. (1993) Procedure for the identi…cation of multiple outliers in linear models. Journal of the American Statistical Association, 88, 424, 1264-1272.

[21] Hamby, D.M. (1994) A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment, 32, 135-154.

[22] Hampel, F.R. (1968) Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley.

[23] Hampel, F.R. (1974) The in‡uence curve and its role in robust estimation. Journal of the American Statistical Association, 62, 383-393

[24] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A. (1986) Robust statistics: the approach based on influence functions. Wiley, New York.

[25] Helton, J.C. (1993) Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering and System Safety, 42, 327-367.

[26] Hettmansperger, T.P., Sheather, S.J. (1992) A cautionary note on the method of least median squares. The American Statistician, 46, 2, 79-83.

[27] Hipel, K.W., Mcleod, A.I. (1994) Time Series Modelling of water resources and Environmental Systems. Elsevier.

[28] Hoaglin, D.C., Mosteller, F., Tukey, J.W. (1983) Understanding robust and exploratory data analysis. Wiley, New York.

[29] Hoffmann, F.O., Gardner, R.H. (1983) Evaluation of uncertainties in environmental radiological assessment models. In Radiological assessment: a Textbook on environmental dose assessment. Till, J.E., Meyer, H.R. eds.. Report n. NUREG/CR-3332. U.S. nuclear regulatory commission, Washington, DC.

[30] Homma, T., Saltelli A. (1996) Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Safety, 52, 1-17.

[31] Huber, P. (1981) Robust statistics. Wiley, New York.

[32] Iman, R.L., Conover W.J. (1982) A distribution-free approach to inducing rank correlation among input variables. Communications in Statististics - Simulation and Computation, 11, 3, 311-334.

[33] Iman, R.L., Hora, S.C. (1990) A robust measure of uncertainty importance for use in fault tree system analysis. Risk Analysis, 10, 401-406.

[34] Kim, B.M., Henry, R.C. (1999) Diagnostics for determining influential species in the chemical mass balance receptor model. Journal of the Air & Waste Management Association, 49, 1449-1455.

[35] Lawrence, K.D., Arthur, J.L. (1990) Robust regression analysis and application. Marcel Dekker, Inc.

[36] Lefrançois, B. (1991) Detecting over-influential observations in time series. Biometrika, 78, 1, 91-99.

[37] Montgomery, D.C., Peck, A.E. (1982) Introduction to linear regression analysis. Wiley, New York.

[38] Peña, D. (1990) Influential observations in time series. Journal of Business & Economic Statistics, 8, 2, 235-241.

[39] Peña, D. (1991) Measuring influence in dynamic models. Technometrics, 33, 1, 93-101.

[40] Ritschard, G., Antille, G. (1992) A robust look at the use of regression diagnostics. The Statistician, 41, 41-53

[41] Rousseeuw, P.J. (1984) Least median of squares regression. Journal of the American Statistical Association, 79, 388, 871-880.

[42] Rousseeuw, P.J., Leroy, A.M. (1987) Robust regression and outliers detection. Wiley, New York.

[43] Rousseeuw, P.J., Van Zomeren, B.C. (1990) Unmasking multivariate outliers and leverage points, (with discussion). Journal of the American Statistical Association, 85, 411, 633-651.

[44] Rousseeuw, P.J., Van Zomeren, B.C. (1992) A comparison of some quick algorithms for robust regression. Computational Statistics & Data Analysis, 14, 107-116

[45] Rousseeuw, P.J., Wagner, J.M. (1994) Robust regression with a distributed intercept using least median of squares. Computational Statistics & Data Analysis, 17, 65-75.

[46] Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989) Design and analysis of computer experiments. Statistical Science, 4, 409-435.

[47] Saltelli, A., Tarantola, S., Chan, K. (1999) A quantitative model-independent method for global sensitivity analysis of model output. Journal of the American Statistical Association, 41, 1, 39-55.

[48] Saltelli, A., Chan, K., Scott, M. (2000) Sensitivity Analysis, Wiley, New York.

[49] Sampson, P.D., Guttorp, P. (1999) Operational evaluation of air quality models. In Environmental Statistics: Analysing data for environmental policy. Novartis Foundation Symposium 220. 33-51. Wiley, New York.

[50] Searle, S.R. (1971) Linear Models. Wiley, New York.

[51] Stuart, A., Ord, J.K. (1991) Kendall's advanced theory of statistics, Vol. 2. Hodder & Stoughton, London.

[52] Turanyi, T., Rabitz, H. (2000) Local Methods, in Sensitivity Analysis, Saltelli, A., Chan, K., Scott, M. eds. Wiley, New York.

[53] Weisberg, S. (1985) Applied linear regression. 2nd edition, Wiley, New York.

[54] Welch, W.J., Buck, R.J., Sacks J., Wynn H.P., Mitchell T.J., Morris, M.D. (1992) Screening, predicting and computer experiments. Technometrics, 34, 1, 15-25.

[55] Wilcox, R.R. (1997) Introduction to robust estimation and hypothesis testing. Academic Press, San Diego.