Environmental data modeling

Hartmut Hebbel

Fachbereich Wirtschafts- und Organisationswissenschaften,
Universität der Bundeswehr Hamburg,
Holstenhofweg 85, D-22043 Hamburg, Germany

In the statistical analysis of environmental data, space and time are often disregarded by the use of classical methods such as hydrological analysis of frequencies or factor analysis. But these methods, based on the assumptions of independent identically distributed observations, cannot be efficient. This article discusses more appropriate approaches regarding the space and time influences, and surveys some important proposals of modeling environmental data. Three examples show the workability of the presented theory. Within the first example, a system to detect abnormal occurrences in water quality as early as possible depending in quasi-continuous data is developed. A second example decomposes a water quality time series into three unobservable components. Finally, it is shown how the factor model can be extended to time series data.

Keywords: Data analysis, water pollution control.
Subject classification: JEL C32, Q25.

1. Introduction: General environmental problems, questions and objectives

Relevant environmental problems are becoming of interest in many areas. For example:

- whether there have been changes in climate;
- how changes in climate will take effect;
- where one can find the main causes of air pollution (regionally, nationally, globally);
- if there are direct interrelations of environmental loads and climate;
- the influence variables of forest decline;
- how to examine the water quality of a river or lake;
- the degree of soil deterioration in certain regions.

For all studies, it is common that numerous variables are observed in space and time to explain the development, and to discover new results of the effects and mechanisms of the distinct influence variables (control process).

The statistician is dependent on the assistance of a specific scientist (expert, specialist) who knows the real dependences and interrelations to use this information for
evaluation techniques. To detect the causal structures of coherence and dependence by only using statistical methods cannot be done without the advice of a practitioner. It requires a great deal of cooperative effort on the part of the scientist and the statistician to complete the necessary knowledge of facts and effective statistical methods.

The most important means of application of statistical methods are:

- an exact question,
- a precise objective.

In many cases of data registration, there are neither precise questions nor objectives. In some cases, there are very general formulations of the objective. Nevertheless, statistical procedures were used, sometimes very elementary in nature, without formulating the real problem. We will return to this point in section 3.

2. Data as a sample in space and time

Hardly any environmental quality variables are continuous in space and time, hence, in other words, cannot in principle be determined at every:

- place or position \( z \) and
- time point \( t \).

Variables are also called parameters by practitioners, not to be confused with characteristic values to be estimated in statistics. Only in a few exceptional cases (e.g. a water-level print out) can data be collected continuously. An example is given in figure 1. In all other cases, data can only be observed discretely, at some:

- positions \( z_1, z_2, \ldots \) and
- time points \( t_1, t_2, \ldots \).

![Figure 1. Water-level curve with tide influence continuously observed.](image-url)