Multidimensional trends: The example of temperature

T.T. Huong Hoang\textsuperscript{1,2,a}, S. Parey\textsuperscript{2,b}, and D. Dacunha-Castelle\textsuperscript{1}

\textsuperscript{1} Laboratoire de Mathématiques, Université Paris 11, Orsay, France
\textsuperscript{2} EDF/R&D, 6 quai Watier, 78401 Chatou Cedex, France

Abstract. Our goal is to discuss the notion of multidimensional trend in the climate change context. Temperature series over Europe are used to derive non parametric trends in mean, variance and extremes. A criterion is proposed to compare the different trends, and a general methodology is applied to test if the trends in extremes are only due to the trends in the central field, as mean and variance ones. Examples are given which show that the temporal evolution in mean and variance of the temperature time series are linked, and that generally these evolutions explain a large part of the temporal evolution of the extremes.

1 Introduction

Climate change is generally presented and discussed in term of trends in the mean of climate variables over different spatial scales, from individual observation stations to spatial means over geographical areas or the entire globe. The study of temperature variability, confidence intervals and trends in extremes are also worth of interest. Generally, trends in mean are derived using ordinary least square regression methods \cite{1,2}. On the other hand, trends are usually studied separately: papers are devoted either to trends in mean, or to trends in extreme events. The trends in extreme events are analysed using linear least squares fits on the series of so called extreme indices \cite{3,4}. Mudelsee \cite{5} uses kernel fitting to study flood risk in a non parametrical way.

In this paper, we would like to address the non parametric derivation of trends, as well as the links between trends in different quantities such as mean, variance and extremes, and then illustrate the point by considering temperature series in Europe \cite{6}.

Generally, a trend is computed (not defined) as a slow and thus regular component of a time series, superimposed on a series of quite stationary and less variable residuals. In other words, computing a trend consists in extracting some deterministic signal from noisy data. In most studies, as stated before, this is done using ordinary least squares fitting, which will be referred to as classical trends.

The point and main topic considered here can be formulated as follows: once the trend in mean has been estimated, are there other trends which could describe the temporal evolution of the series, and how many are significant? Are they independent in some sense? If we leave out of this study, for simplicity, the usual seasonality and all periodic phenomena, our goal is then to look for multidimensional deterministic trends, which, when removed, leave an as stationary as possible residual series. In practice, as we will illustrate on temperature series, the search for stationarity has to be stopped when the physical meaning of the trends becomes too weak or when the trends are too strongly linked.

\textsuperscript{a} e-mail: huong.hoang@math.u-psud.fr
\textsuperscript{b} e-mail: sylvie.parey@edf.fr
As our goal is to derive non-parametric trends in the most objective and general possible way, we first mention the general qualitative properties, requirements and assumptions made to identify a trend. They can be listed as follows:

- **Choice of the time scale**: When observations are given over a time period $T$ one has to decide which time scale (less than $T$) is of interest to study the signal variation; to simplify we will call it the window length in this work. For example, if the topic of interest is the possible anthropogenic effect on climate change, if one has to consider data over the industrial period, the window length is more or less arbitrarily chosen as 30 years if two to three centuries of observations are available. Over the most recent period, the decade is often chosen as window length if the period length of observations is at most fifty years. The window length for a given observation period is indeed often chosen in a quite arbitrary way and this choice is very important for instance in the case of the increasing mean temperature. We will show in the paper that such a choice based on informal or subjective ideas (heuristics) can be improved when using more intrinsic (mathematical) considerations.

- **Almost invariance by data extension and localisation**: Once the length $T$ of the observations has been fixed, if we add new observations whose period length is a significant fraction of $T$ ($T/2$ for example), then the modifications on the previously computed trend on any subinterval of the initial dataset have to be small. Of course, if we consider a trend estimated on a 50-year period of observations, and then include this period in a larger period of say 300 years, the trend is obviously modified and the intrinsic smoothness will change. Nevertheless, intrinsic choices allow to give an objective interpretation of this modification and can help to define changes or breaks in the trend behaviour.

- **Monotonicity**: In order to easily interpret the physical phenomena, the monotonicity can be imposed to a trend. For instance if trends are used in order to define an extrapolation necessary to compute return levels, monotonicity is required in general. Variants are in constraints on the number of intervals where the trend is monotone or convex. There are general tools to obtain constrained estimates as isotonic regression (see [7] for example).

Starting from these considerations, the paper is devoted to the identification of trends in mean, variance and extremes of a time series, together with their possible links. It is organised as follows: first, the statistical framework is given in section 2 before coming to the results for trends in mean, variance and extremes of temperature series in Europe in section 3. The link between the trends in mean and variance and the trends in extremes is discussed in section 4 and a methodology is proposed to test if the trends in extremes are due to trends in mean and variance, before concluding in section 5.

### 2 Statistical framework

#### 2.1 Trend in mean

We first discuss the trend in mean, with the supposition that there is no seasonal trend (or any trend with an obvious shorter window, this point can be checked for example using a wavelet analysis).

The basic approach uses a moving average procedure, with a moving window length $L$. In this simplest case as for the more complicated forthcoming, we define the model, here:

$$X(t) = m(t) + e(t)$$  \hspace{1cm} (1)

where $X(t)$ is the observation series with unknown mean $m(t)$ and $e(t)$ is a centred process, which is expected in general to be stationary and uncorrelated.

This “signal + noise” approach is first improved in using non-parametric statistics, and then often followed by the choice of a regression model in parametric statistics.

Non-parametric methods satisfy the previously given requirements, but now the window length of interest is chosen in an intrinsic way and has a precise definition. The main lines are