Trends in environmental data: when uncertainty matters

Krista Jaarsma [a,b,c], Hans Visser [a,*] and Ernst C. Wit[b]

[a] PBL Netherlands Environmental Assessment Agency, P.O. Box 303, 3720 AH, The Netherlands.

[b] Johann Bernouilli Institute, RUG University of Groningen, Nijenborgh 9, 9747 AG, Groningen, The Netherlands

[c] Quantitative Analysis PwC, T.R. Malthusstraat 5, Amsterdam

Contact hans.visser@pbl.nl

Abstract

Many models exist in the environmental literature on trend estimation. In fact, there is no 'best' or 'true' trend model. The choice of practitioners will depend on their goals or 'interests'. In this article we present a well-known trend model based on the theory of structural time series models and the Kalman filter: the Integrated Random Walk (IRW) model. This trend model can be seen as a natural extension of the linear OLS trend: the trend estimate can be a straight line or show more flexibility while retaining full uncertainty information (uncertainties in trends, trend differences and probabilities for crossing thresholds). Here, we add a new feature to this modeling approach expanding uncertainty information further by introducing the concept of *trend ensembles*, constructed directly from the Kalman filter formulae. We show that these trend ensembles allow one to estimate simultaneous confidence bands in addition to pointwise confidence bands. The advantage is that simultaneous confidence limits better resemble what users think these bands mean, namely that the true trend or trend difference will lie between the upper and lower bands. We give three applications from climate change research, based on the TrendSpotteR software, available in R.

Keywords climate change - integrated random walk - discrete Kalman filter - simultaneous uncertainty bands - TrendSpotteR

1. Introduction

Netherlands Environmental Assessment Agency (PBL) is a national institute for strategic policy analysis in the fields of environment, nature and spatial planning. As part of their work a website is maintained containing over 800 indices covering all aspects of the environment. Many of these indices have a time-series character. An example is a time series on sea level rise along the Dutch coast, based on six century long tide-gauge series (1890-2012): http://www.compendiumvoordeleefomgeving.nl/indicatoren/nl0229-Zeespiegelstand-Nederland.html?i=9-54 . Questions we have for data such as these, are: what is the long-term trend and are changes in trends statistical significant?

But how could trends in data be estimated, along with uncertainties? The environmental literature on trends shows an overwhelming variety of trend models. We name moving averages with a pre-defined window, ordinary least squares (OLS) linear trend models, polynomial trend fits, splines, exponential smoothing, trends based on Bayesian principles, trends in rare events based on logistic

regression, smooth transition models, LOESS estimators, binomial filters, trends based on extreme value theory (time-varying GEV distributions), GARCH trends, Friedman super smoothers, wavelets, structural time-series trend models, robust regression trends (MM or LTS regression), Sen-Theil trend estimators for linear trends, and Box-Jenkins ARIMA models and variations (SARIMA, GARMA, ARFIMA). Next to these trend models a number of statistical tests for trends exist: Student's t test for sub-periods of time, Mann-Kendall tests for monotonic trends (with or without correction for serial correlation), and trend tests against long-memory time series.

However, the choice becomes limited if we are interested in stochastic trend models where uncertainty bands are given, both for a trends μ_t and any difference thereof, thus $[\mu_t - \mu_s]$, with times 't' and 's' within the sample period. We name OLS linear trends and higher order polynomial trend fits and models from the class of structural time series models (STMs) (Visser, 2004; Chandler and Scott, 2011; Visser and Petersen, 2012). Some other models are stochastic in nature but do not give uncertainties for trend differences, such as trends based on extreme value theory (Coles, 2001) and trends based on nonparametric regression (Chandler and Scott, 2011 - Ch. 4).

In this article we describe in detail a trend estimation approach which can answer these questions, namely the integrated random walk (IRW) trend model, which is a sub-model from the class of STMs. The IRW trend model is a well-known stochastic trend model which has a number of advantages for practitioners (Kitagawa, 1981; Visser, 2004; Young *et al*., 1991). First, the model can be seen as a natural extension of the OLS linear trend: both linear and more flexible trends can be estimated, while full uncertainty information is retained. Furthermore, the trend is given over the full sample period and predictions can be generated along with uncertainty bands. The flexibility of the trend can be chosen by maximum likelihood optimization (based on minimization of one-step-ahead prediction errors). All estimates are gained by the discrete Kalman filter yielding minimum mean square estimators (MMSEs). Finally, the IRW trend can be extended to models incorporating cycles and explanatory variables, in the following additive way:

$$y_t = trend_t + cycle_t + influence\ of\ explanatory\ variables\ x_t + noise_t \quad (1a)$$

where y_t denotes a measurement at time t. This reads in mathematical terms,

$$y_t = \mu_t + \gamma_t + \alpha_{1,t}\chi_{1,t} + \alpha_{2,t}\chi_{2,t} + \cdots + \varepsilon_t \quad (1b)$$

References for these extended models are Harvey (1984, 1989), Visser (1994), Commandeur and Koopman (2007), Mills (2010) and Young (2011).

In this article we expand the IRW trend model by introducing the method of estimating *trend ensembles*. A trend ensemble is a set of, say, 1000 IRW trend realizations which might occur with equal chance, based on the same IRW trend model. We will show how these trend realizations can be derived from the Kalman filter equations and how these ensembles can be applied to extend the possibility of estimating uncertainty bands. One such extension is the estimation of *simultaneous confidence bands*, in addition to pointwise confidence bands. The advantage of having simultaneous confidence bands is that these bands appeal to the intuition of people. In our experience many environmental scientists interpret pointwise uncertainty bands around a trend means as that the **true** trend will lie between the upper and lower limits, with a chance of 95%. However, that interpretation is only true for simultaneous confidence bands (cf. Wynn and Bloomfield, 1971). Thus, presenting simultaneous confidence bands rather than pointwise confidence limits, has *communicational* advantages.

Furthermore, we will show how trend ensembles allow one to construct confidence bands for the probability of crossing thresholds. Uncertainties for these probabilities or their reverses, i.e. the average return periods, will be derived. These uncertainties cannot be gained from standard theory (cf. Visser and Petersen, 2012, Section 7).

In this article we will present the theory of IRW trends along with the Kalman filter in Section 2

where details are given in Appendix A. Next to that we will show how the Kalman filter formulae can be used to generate trend realizations, leading to trend ensembles (Appendices B and C). We will show how these trend ensembles can be used to obtain uncertainty bands, including the construction of simultaneous confidence bands. Then, three applications from the field of climate change are given in Section 3. Here, estimates are gained from an R implementation of the approach given in Section 2. These applications are the analysis of spring temperatures in the Netherlands, 1906-2012, the analysis of global economic losses due to weather related disasters (1980-2010), and the analysis of annual maximum temperatures in the Netherlands (1951-2012). Strong and weak sides of the approach are discussed in Section 4 while conclusions are given in Section 5.

2. IRW trends, trend ensembles and the Kalman filter

In this Section we describe how the IRW trend model can be used to estimate trends along with full uncertainty information. As stated in the Introduction, this information comprises uncertainty bands for trends μ_t and any difference thereof, thus $[\mu_t - \mu_s]$, and the chance of crossing predefined thresholds. A new aspect is the estimation of simultaneous confidence limits, based on trend ensembles.

2.1. IRW trends and the Kalman filter

The IRW trend model has the following form:

$$y_t = \mu_t + \varepsilon_t \quad \text{and} \quad \mu_t = 2\mu_{t-1} - \mu_{t-2} + \eta_t, \tag{2}$$

where y_t denotes a measurement at time t; η_t and ε_t are independent, normally distributed, white noise processes with zero mean and variances σ_η^2 and σ_ε^2, respectively. For estimating trends from this model by use of the Kalman filter, model (2) needs to be rewritten in the state-space form:

$$\begin{pmatrix}\mu_{t+1}\\ \lambda_{t+1}\end{pmatrix} = \begin{bmatrix}2 & -1\\ 1 & 0\end{bmatrix}\begin{pmatrix}\mu_t\\ \lambda_t\end{pmatrix} + \begin{pmatrix}\eta_t\\ 0\end{pmatrix} \quad \text{and} \quad y_t = (1\ 0)\begin{pmatrix}\mu_t\\ \lambda_t\end{pmatrix} + \varepsilon_t, \tag{3}$$

where the term λ_t equals μ_{t-1} , and $\eta_t \sim N(0, \sigma_\eta^2)$ and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Under these assumptions of normality, the Kalman filter provides optimal estimates $\hat{\mu}_t$ for the trend μ_t: the filter yields the minimum mean square estimator (MMSE) for the vector $(\mu_t, \lambda_t)'$, based on observations up to and including time t. If the noise processes are not normally distributed, the filter generates the minimum mean square linear estimator (MMSLE). This still yields optimal estimates but the filter is less powerful. For more information on the Kalman filter please refer to Appendix A and to Harvey (1984, 1989), Durbin and Koopman (2001), and Chandler and Scott (2011- Section 5.5).

The IRW trend model yields both linear and flexible trends, depending on the noise variance σ_η^2. If this variance is set to zero, the IRW trend equals the OLS linear trend. On the other hand, when σ_η^2 is set to a large number, the trend will be extremely flexible. Since the value of this noise variance σ_η^2 steers the flexibility of the trend, σ_η^2 is also known as the 'smoothing parameter'. The optimal value for σ_η^2 can be obtained using maximum likelihood estimation (Harvey, 1984, 1989). This implies a minimization of the sum of squared one-step-ahead prediction errors. In this way the flexibility is

'adapted' to the data.

Due to the statistical nature of the IRW trend model one can compute predictions $\tilde{y}_t$, given a trend model μ_t. From Eq. (2) it follows that these predictions have a mean $\hat{\mu}_t$ and variance $\hat{\sigma}^2_{\mu,t} + \hat{\sigma}^2_{\varepsilon}$, or $\tilde{y}_t \sim N\left(\hat{\mu}_t, \hat{\sigma}^2_{\mu,t} + \hat{\sigma}^2_{\varepsilon}\right)$. These parameters directly follow from the Kalman filter equations. Now, the chance that a prediction $\tilde{y}_t$ will cross a threshold L, simply equals to $P_t^L \equiv P(\tilde{y}_t > L)$.

2.2. Trend ensembles and uncertainties

Using the IRW trend model and the Kalman filter for a suitable time series, optimal estimates for the trend μ_t can be found given the measurements $y_1, ..., y_N$. Now, the Kalman filter generates estimates of the trend values $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)$ that follow a multivariate normally distribution if $\eta_t \sim N(0, \sigma^2_{\eta})$ and $\varepsilon_t \sim N(0, \sigma^2_{\varepsilon})$ for all times t. We show in Appendix A how the vector $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\Sigma}$ of this multivariate normal distribution can be deduced from the general Kalman filter theory.

Once $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are known, we can generate a series of M trends based on this multivariate normal distribution. Two approaches for drawing random vectors from a multivariate normal distribution exist. One approach uses conditional distributions and the other applies eigenvalue decomposition of the covariance matrix. For more information about both approaches see Appendix B. The set of M trends constructed in the way outlined above, will be denoted here as a *trend ensemble*.

The advantage of having trend ensembles lies in the fact that it makes the computation of confidence bands easy. Only simple statistical concepts are needed. E.g., the trend estimate $\hat{\mu}_t$ and its corresponding variance $\hat{\sigma}^2_{\mu,t}$ follow directly from the M simulations $\hat{\mu}_t^{(1)}, ..., \hat{\mu}_t^{(M)}$ by taking the mean and variance of these M simulated trend estimates at time t. For large M this mean and variance will converge to the mean and variance which follow directly from the Kalman filter formulae. In the same way trend differences and uncertainty bands are calculated for any trend difference $[\mu_t - \mu_s]$, with time t and time s lying in the sample period and $t \geq s$. Since we have M estimates $\hat{\mu}_t^{(1)}, ..., \hat{\mu}_t^{(M)}$ and M estimates $\hat{\mu}_s^{(1)}, ..., \hat{\mu}_s^{(M)}$ we also have M estimates for $[\hat{\mu}_t - \hat{\mu}_s]$. Thus, we can calculate means, variances or any percentile on these M values. E.g., 95% pointwise confidence bands follow from the 2.5th percentile and the 97.5th percentile.

In fact, any possible trend function with corresponding confidence interval can be generated. One such trend function is the calculation of trend *accelerations*. In climate change research it is of interest if trends in global mean sea level rise show accelerations or in other words, is the second derivative of μ_t positive? These second derivatives read as $[\mu_t - 2\mu_{t-1} + \mu_{t-2}]$ and analogous to the trend differences $[\mu_t - \mu_{t-1}]$ we can compute M differences $[\mu_t - 2\mu_{t-1} + \mu_{t-2}]$ along with any confidence band of interest.

Confidence limits for the probability of crossing a threshold L (denoted by P_t^L) can also be derived from trend ensembles (Visser and Petersen, 2012 - Figs 1 and 6). For any simulated trend k we can calculate the variance σ_k^2 of the residuals $[y_1 - \hat{\mu}_1^{(k)}], ..., [y_N - \hat{\mu}_N^{(k)}]$ and we can compute $P_t^{L,k}$, t= 1, ..., N, assuming $\tilde{y}_t^{\,k}$ to be normally distributed: $\tilde{y}_t^k \sim N\left(\hat{\mu}_t^{(k)}, \hat{\sigma}_k^2\right)$. In this way we can generate M probabilities $P_t^{L,1}, ..., P_t^{L,M}$. Uncertainty bands for P_t^L follow by computing variances or percentile values of interest based on these M values $P_t^{L,k}$. Also uncertainties for any probability difference $[P_t^L - P_s^L]$ can be computed.

A final application of trend ensembles is the calculation of simultaneous confidence bands, in addition to pointwise confidence bands. Simultaneous confidence bands give upper and lower limits for the trend *as a whole*. In the context of trend ensembles simultaneous 95% confidence limits can be estimated by finding a tight envelop around the estimates $\hat{\boldsymbol{\mu}}_1, ..., \hat{\boldsymbol{\mu}}_N$ such that exactly 0.95M simulated trends fall within this envelop, or equivalently, the choice of the 0.05M extreme trends which fall *outside* the envelop. This choice is not unique since the term 'extreme' is not unique. In Appendix C we propose the procedure we have followed in the R software, described in the next

Section. We note that simultaneous confidence limits are rarely found in the literature. See Wynn and Bloomfield (1971), Montgomery and Peck (1982 - Section 9.6) and Liu *et al.* (2008).

2.3. Software implementation in R

We have implemented the IRW ensemble approach from Sections 2.1, 2.2. and Appendices A, B in a software package TrendSpotteR, implemented in R. Once this package is coupled to an R session, an option screen is shown (Figure 1). Options are:

- a log transformation to the data. If this option is chosen, the IRW model is estimated on log-transformed data and estimation results will be back transformed to the original scale (except for innovations) .
- one can choose for maximum likelihood optimization or set the smoothing parameter 'by hand'.
- if the data file contains missing data, the missing data code can be given.
- if threshold exceedances are of interest, a threshold can be given.
- if predictions for future trend values are of interest, the number of predictions can be specified.

The software generates an IRW trend, trend differences and the chance for crossing a threshold, along with pointwise and simultaneous confidence limits. Furthermore, one can check the validity of a number of assumptions underlying IRW trends. These checks are performed on the residuals or, more precisely, the standardized innovation series. A first check is the whiteness of residuals using an autocorrelation function (ACF). The homoscedasticity of residuals can be checked by visual inspection of the innovations or by using an F test. Additionally, the normality of innovations should be checked. This normality is a prerequisite for using trend ensembles (and thus for estimating simultaneous confidence bands). We note that the pointwise confidence bands for trend and trend differences in TrendSpotteR do not depend on the normality of residuals. However, 95% confidence limits should be read as 2-σ confidence limits if the assumption of normality is violated.

TrendSpotteR creates three graphs for checking model assumptions: the QQ-plot, the ACF plot and the plot of residuals. For more detailed information on checking residuals we refer to Harvey (1989, Ch. 5), and Chandler and Scott (2011, Section 3.3).

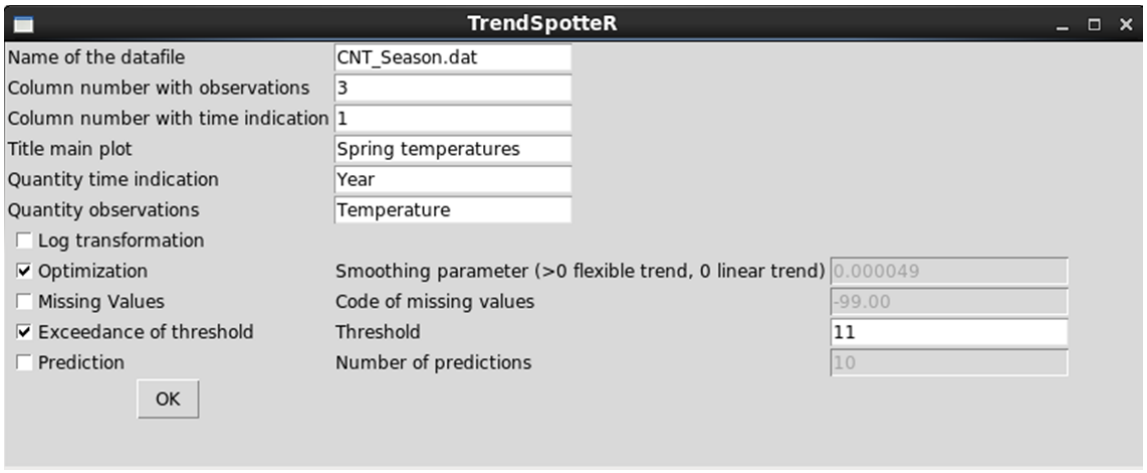

Fig. 1 Input options screen for the analysis of spring temperatures in the Netherlands (cf. Section 3.1)

3. Applications

We give three applications based on the IRW approach given in Section 2 and Appendices A, B and C. These applications are from the field of climate change and all estimates are gained with the TrendSpotteR software. The first example concerns a series of century-long spring temperatures in the Netherlands. The second example is on global economic damage due to weather-related disasters and the third example deals with a series of extreme temperatures in the Netherlands.

3.1. Spring temperatures, 1906-2012

The analysis of century long seasonal temperature data is important since climate change may have serious ecological and societal impacts. An example of the role of rising spring temperatures in the Netherlands has been given by Both and Visser (2001). They showed that increasing spring temperatures led to phase shifts in laying dates of long-distance migrant birds and earlier peaks in insect abundance. In this context we estimated an IRW trend in spring temperatures over the period 1906-2012.

These data are available in the Central Netherlands Temperature (CNT) database (Schrier et al., 2011). Seasonal temperatures from this database are given in the ASCII file 'CNT_Season.dat', being part of the TrendSpotteR software. The data file consists of six columns: year, temperatures in winter, spring, summer and autumn. The final column consists of annual averages. All temperatures are expressed in °C. Here, we choose to estimate an IRW trend model for spring temperatures, along with probabilities for crossing the threshold of 11.0 °C. The corresponding input screen is given in Figure 1.

The TrendSpotteR run generated six graphs with trends, trends differences, probabilities of crossing the threshold of 11.0 °C and differences thereof. See Figures 2a through f. Figure 2a shows the estimated trend $\hat{\mu}_t$, along with its corresponding confidence intervals (both simultaneous and pointwise). Figure 2b shows all simulated 1000 trends which have been used to compute these simultaneous confidence bands. Figure 2c shows the first differences $[\hat{\mu}_t - \hat{\mu}_{t-1}]$, along with confidence bands. The graph shows that these differences are statistical significant ($\alpha = 0.05$) over a short time interval: 1983-2005 (the lower limit of simultaneous confidence bands above zero). In the same way Figure 2d shows that the trend differences $[\hat{\mu}_{2012} - \hat{\mu}_t]$ are statistically significant positive for all years in the period 1906-2008.

Figure 2e shows the probability of crossing the threshold of 11.0 °C. The graph shows that the probabilities are near zero over the period 1906-1980, with a rapid acceleration afterwards. The probability in 2012 has risen to 0.22, with a 95% pointwise confidence limit of [0.078, 0.45]. Since the lower limit is above zero, the $P_{2012}^{11.0}$ value is statistically significant. The corresponding average return period is calculated from the inverse probability: once in five years with a confidence band of [2, 13] years. Finally, Figure 2f shows that the probability $P_{2012}^{11.0}$ is significant higher than all other probabilities $P_t^{11.0}$.

It is interesting to compare the widths of pointwise and simultaneous confidence bands. Figures 2a, c, d, e, and f show that these bands are pretty close: simultaneous confidence bands are 10 to 20% wider.

TrendSpotteR generated three graphs for checking model assumptions, see Figure 3. Both ACF (Figure 3a) and the innovation plot (Figure 3b) show that the innovations are in good agreement to be white noise processes. All 22 autocorrelations are within the 2-σ confidence limits and show no periodic behavior. We have checked the homoscedasticity of residuals by visual inspection of Figure 3b. Figure 3c shows the normality plot for the innovations. The graph shows that these innovations are reasonable centered around a straight line. Thus, no log-transformation or other Box-Cox transformations are needed here. Our judgment from these three graphs is that the innovations are reasonable normally-distributed white noise.

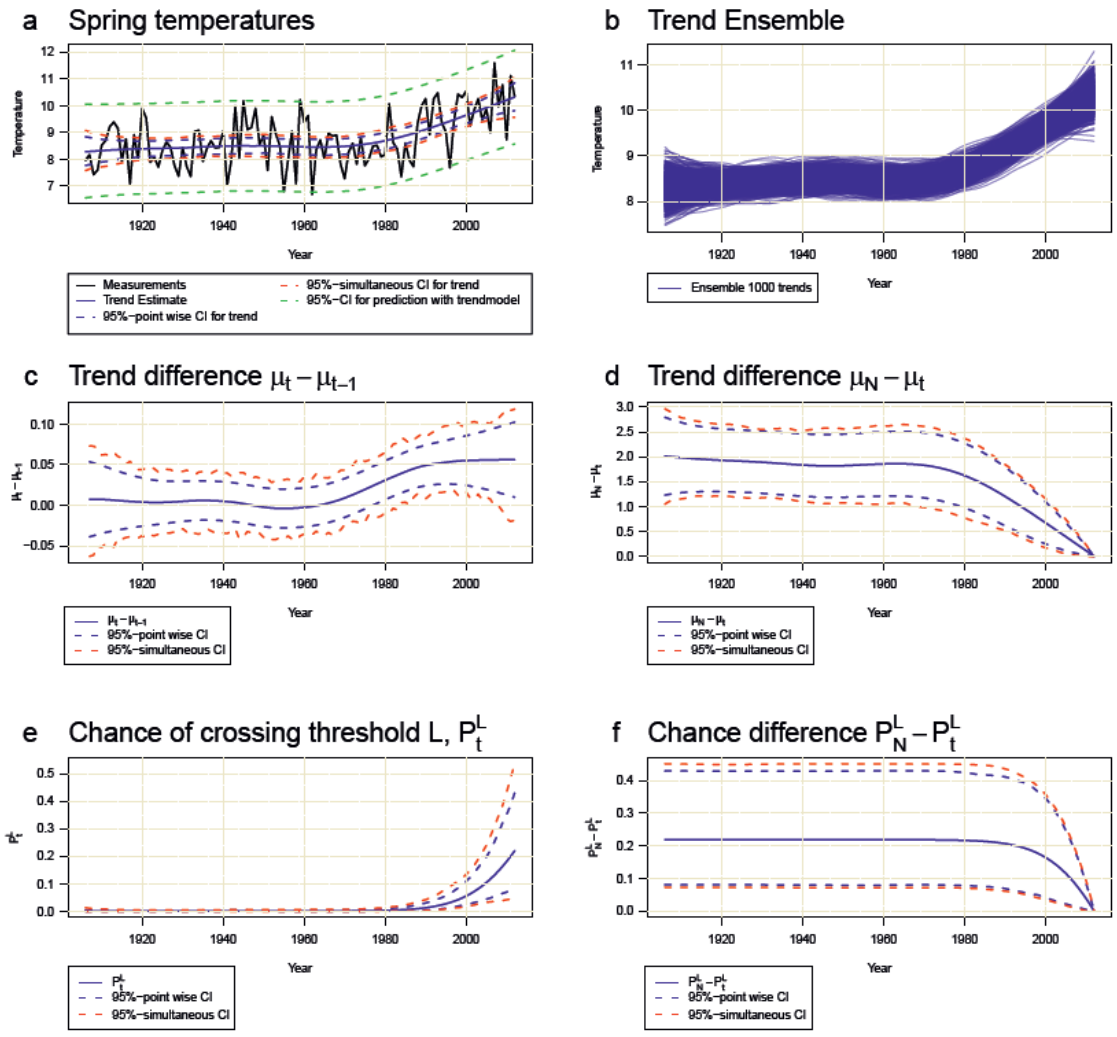

Fig. 2 Graphical output for spring temperatures in the Netherlands, 1906-2012. (a) Data and estimated trend, along with its corresponding pointwise and simultaneous 95% confidence bands. All temperature data are expressed in °C. (b) Trend ensemble consisting of 1000 trends. (c) and (d) Differences in trend values along with their corresponding confidence bands. (e) and (f) Probability of crossing the threshold of 11.0 ºC and differences in these probabilities, along with confidence bands

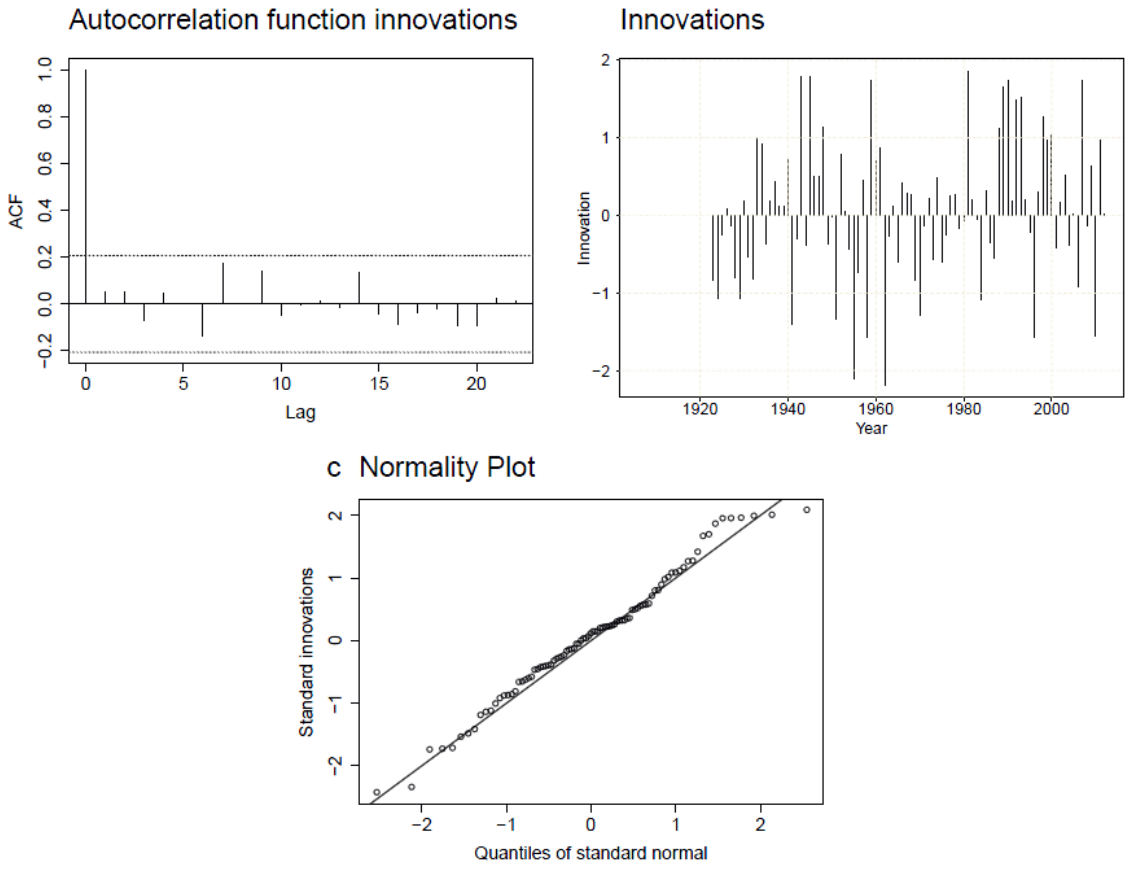

Fig. 3 Graphical output additional to those shown in Figure 2. Graphs are for checking assumptions on standardized innovations (= one-step-ahead prediction errors). (a) Autocorrelation function for lags 1 to 22. (b) Innovation series. Note that these innovations are shown from 1922 onwards (innovations have to stabilize first). (c) Normality plot for standardized innovations

3.2. Global economic losses from weather-related disasters, 1980-2010

Disasters such as floods, storms, heat waves and droughts, may have enormous implications in terms of health, the environment, and economic development of countries. Historic examples of severe disaster impacts are the drought in Ethiopia and Sudan which led to over 400,000 people killed through famine and Hurricane Katrina and subsequent flooding in the U.S. which led to economic damages valued at 140 billion US$. Recent publications on the (economic) impacts of weather-related disasters are Bouwer (2011), IPCC-SREX (2012),Visser et al. (2012) and Visser et al. (2013).

Here, we will give an example taken from Visser et al. (2012): estimating an IRW trend for global economic losses due to weather-related disasters. These data have been extracted from the CRED database on natural disasters and are analyzed over the period 1980-2010. Weather-related disasters are disasters due to hurricanes, extra-tropical storms, local storms, heat waves and other temperature extremes, droughts, coastal floods an fluvial floods. Economic losses are expressed in billion $US\$_{2010}$ (all losses expressed U.S. dollar currency for the year 2010). For this example we are interested in probabilities of crossing the threshold of 110 billion $US\$_{2010}$. Due to the large outlier in the year 2005 (largely due to Hurricane Katrina) we have decided to choose the log-transformation option. Global losses are given in the ASCII file 'GlobalDamage.dat', being part of the TrendSpotteR software.

Trend estimates are shown in Figure 4. Figure 4a shows the data, the estimated trend and

corresponding 95% confidence bands. The trend pattern shows a rise over the period 1980-1995 and a stabilization thereafter. The trend ensemble, consisting of 1000 trends, is shown in Figure 4b. It is interesting to note the variability is largest at the beginning and ending of the series. This variability skewed to higher values which can be explained from the log transformation. Figure 4c shows the first differences as the ratio $[\hat{\mu}_t \ / \hat{\mu}_{t-1}]$, along with confidence bands. The graph shows that these ratios are statistical significant ($\alpha = 0.05$) over a short time interval: the lower limit of simultaneous confidence bands lies above 1.0 over the period 1985-1991. In the same way Figure 4d shows that the trend ratios $[\hat{\mu}_{2010} \ - \hat{\mu}_t]$ are statistically significant higher than 1.0 for the years before 1980-1990.

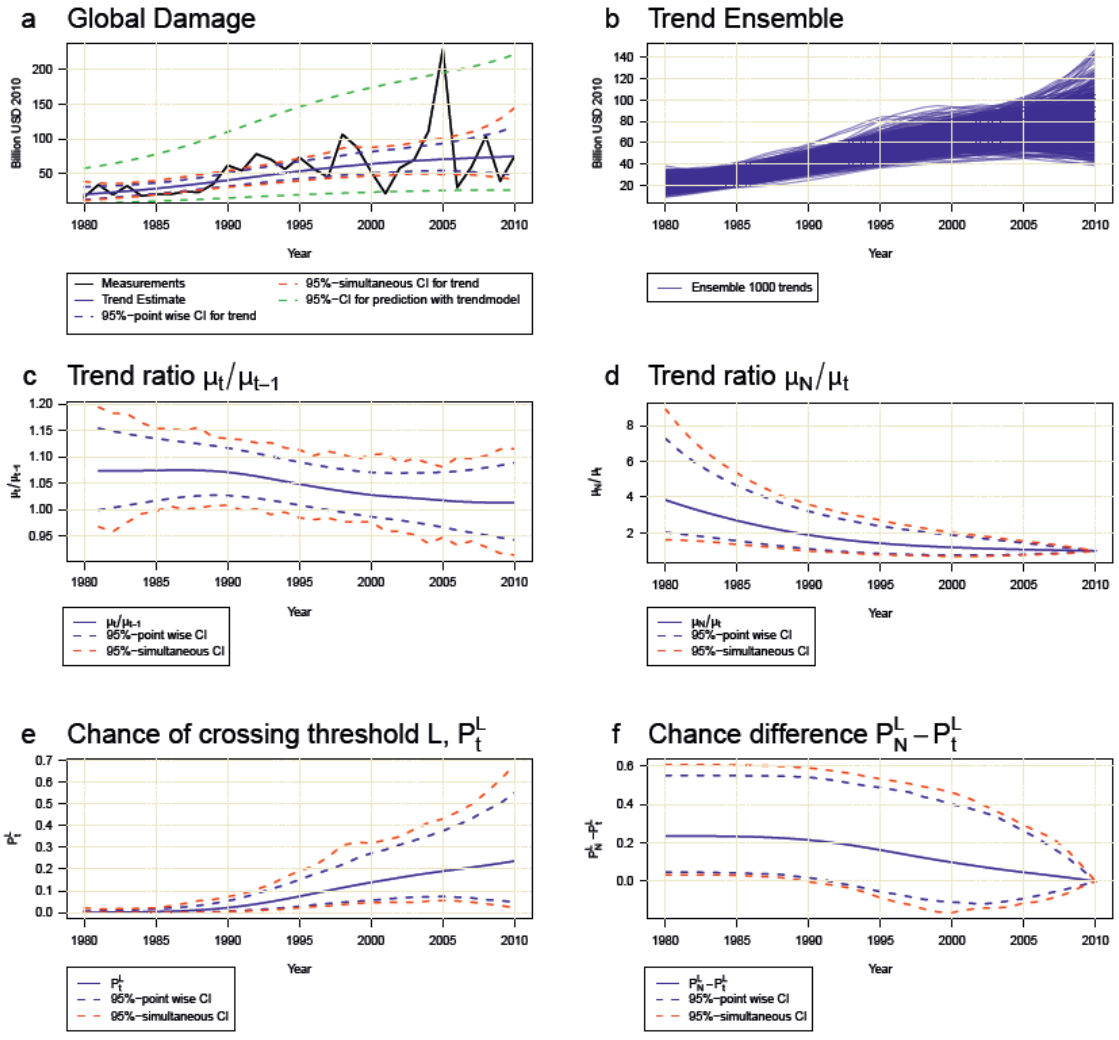

Fig. 4 Graphical output for global economic losses, 1980-2010. Losses are expressed in billion dollars US$$_{2010}$$. (a) Data and estimated trend, along with its corresponding pointwise and simultaneous 95% confidence bands. (b) Trend ensemble consisting of 1000 trends. (c) and (d) Differences in trend values along with their corresponding confidence bands. (e) and (f) Probability of crossing the threshold of 110 billion US$$_{2010}$$, and differences in these probabilities, along with confidence bands

Figure 4e illustrates the probability of crossing the threshold of 110 billion US$$_{2010}$. The graph shows that the probabilities are near zero over the period 1980-1990, with a rapid acceleration over 1991-2010. The probability in 2012 has risen to 0.24, with a 95% pointwise confidence limit of [0.050, 0.55]. The corresponding average return period is calculated from the inverse probability: once in four years with a confidence band of [2, 20] years. Finally, Figure 4f shows that the probability P_{2010}^{110} is significant higher than all other probabilities $P_t^{11.0}$ in the interval 1980-1990.

The residuals of the IRW trend model showed some heteroscedasticity, with larger variability at the end of the series. We have tested the homoscedasticity of residuals by dividing the sample period in two parts and by comparing the respective variances using an F test. The p value of the test is 0.052 which is on the edge of accepting or rejection the H_0 hypothesis of equal variances. As for normality the innovations are not perfectly centered around a straight line, but deviations are not too large. This result is typical for a number practical applications: residuals suffice 'more or less' to assumptions underlying the model chosen. We return to this point in the next section.

3.3. Extreme temperatures in the Netherlands, 1951-2012

Europe experienced an extreme heat wave in the year 2003 and countries such as France were hit hard. According to the French National Institute of Health there were over 14,000 heat-related deaths (mostly among the elderly). For the Netherlands this number was estimated to lie around 1500 people. Here, the highest temperature measured was 37.8 °C. In this example we estimate an IRW trend through annual maximum temperatures ('the hottest moment in a year'). The question is: are such extreme temperatures on the rise or was '2003' an incident?

Annual maxima, in jargon denoted by 'TXX_t', were taken from station De Bilt, being the main observatory of the Netherlands. Due to a large discontinuity in the data around 1950 (a change in the type of measurement screen) we analyzed the maxima over the period 1951-2012 (in stead of the full 1901-2012). This example is also illustrative since we analyzed the same data with the software package TrendSpotter (Visser and Petersen, 2012 - Figures 1, 4 and 6). As in Visser and Petersen (2012) we choose a threshold of 35.0 ºC to show how probabilities of crossing a high threshold changed over the sample period.

The TrendSpotteR graphs are shown in Figures 5a through f. Figure 5a shows the estimated trend $\hat{\mu}_t$, along with its corresponding confidence intervals (both simultaneous and pointwise). The pattern is almost linear (but not prescribed to be linear!). Note that the TXX value for 2003 is not the highest in the record. Figure 5b shows all simulated 1000 trends which have been used to compute these simultaneous confidence bands. Figure 5c shows the first differences $[\hat{\mu}_t - \hat{\mu}_{t-1}]$ which show a stabilized pattern: 0.060 ± 0.030 ºC/year around 1975 (95% pointwise confidence limits). The lower pointwise uncertainty band shows that all differences are statistical significant ($\alpha = 0.05$). The lower simultaneous confidence band lies above zero for all years except 1951-1959. Figure 5d shows that the trend differences $[\hat{\mu}_{2012} - \hat{\mu}_t]$ are statistically significant positive for all available years. The increase over the whole sample period 1951-2012 is considerate: $[\hat{\mu}_{2012} - \hat{\mu}_{1951}] = 3.6 \pm 1.6$ ºC (95% pointwise confidence limits).

Figure 5e shows the probability of crossing the threshold of 35.0 °C. The graph shows that the probabilities are near zero over the period 1951-1980, with a rapid acceleration afterwards. The probability in 2012 has risen to 0.18, with a 95% pointwise confidence limit of [0.07, 0.32]. Since the lower limit is above zero, the $P_{2012}^{35.0}$ value is statistically significant. The corresponding average return period is calculated from the inverse probability: once in five years with a confidence band of [3, 14] years. Finally, Figure 5f shows that the probability $P_{2012}^{35.0}$ is significant higher than all other probabilities $P_t^{35.0}$. Tests on the innovations yielded satisfactory results, both for whiteness and normality.

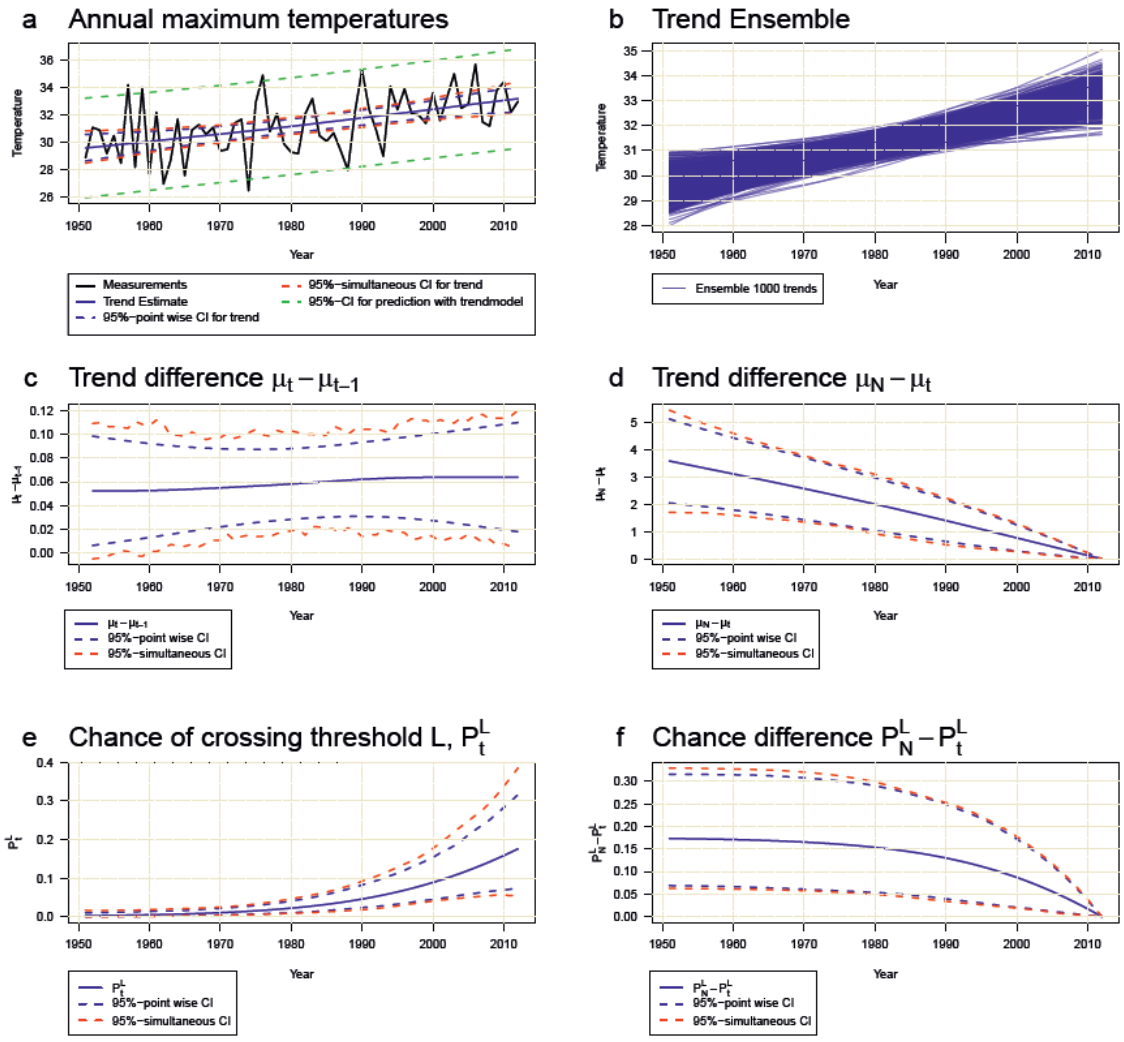

Fig. 5 Graphical output for annual maximum temperatures in the Netherlands, 1951-2012. (a) Data and estimated trend, along with its corresponding pointwise and simultaneous 95% confidence bands. All temperature data are expressed in °C. (b) Trend ensemble consisting of 1000 trends. (c) and (d): Differences in trend values along with their corresponding confidence bands. (e) and (f) Probability of crossing the threshold of 35.0 ºC and differences in these probabilities, along with confidence bands

4. Limitations of the IRW approach

The IRW approach described in Section 2 will not suffice for all data in practice. An example is given in Section 3.2. Here, the innovation series showed a more or less heteroscedastic pattern. The F test was on the edge of significance, or in other words, a doubtful case. However, severe deviations from whiteness or normality may occur in practice. We name three ways how to proceed in these cases.

1. If the assumption of whiteness of innovations is violated, one could correct pointwise confidence bands by applying effective numbers. In this case the sample size N is replaced by a smaller number of data, denoted as 'N_{eff}'. Please see Zieba (2010) and Chandler and Scott (2011, Section 3.3.3).
2. If innovations are heteroscedastic or non-normal or both, a transformation of original measurements will work in many cases. one way to do that is the application of Box-Cox transformations. Box-Cox transformations have the following form:

$$Y_t = \begin{cases} \frac{(X_t^{\lambda}-1)}{\lambda} & , \quad \lambda \neq 0 \\ \log(X_t) & , \quad \lambda = 0 \end{cases} \tag{4}$$

   where λ denotes the power of the transformation. The case where $\lambda = 0$ (log transformation) is included in the TrendSpotteR software. However if $\lambda \neq 0$, one may want to back transform the trend results to the original scales of X_t, or $X_t = (\lambda Y_t + 1)^{\frac{1}{\lambda}}$.
3. One can extend the trend model by adding explanatory variables and/or cycles, as given in Eq. (1). This extension might improve the pattern of residuals considerately. Significant first order autocorrelations or cyclic patterns in residuals might well 'disappear'. Software packages where such extensions are modeled, are TrendSpotter[1)] (Visser, 1994; Visser et al., 2010), the toolbox CAPTAIN (Young, 2011) and the commercial package STAMP. A practical example of extending the trend model has been given by Visser and Brandes (2013) in the field of sea-level rise research.

[1)] TrendSpotteR has the estimation of IRW trends in common with the older software package TrendSpotter (programmed in Fortran). TrendSpotter estimates IRW trends which could extended with seasonality and the influence of explanatory variables as given in Eq. (1). However, the extension to trend *ensembles* is only available in TrendSpotteR. For TrendSpotter examples please see Visser (2004), Visser et al. (2010), and Visser and Petersen (2009, 2012).

5. Summary and conclusions

A wide variety of trend models are available to estimate trends in (environmental) data. In this article we propose the IRW trend model for a number of reasons:

- both linear and flexible trends can be estimated. The linear version is identical to the well-known OLS linear trend model.
- the trend is available over the full sample period. Its flexibility is chosen by maximum likelihood optimization.
- predictions can be generated.
- the probability for crossing pre-defined thresholds can be computed.
- the model, in combination with the Kalman filter and trend ensembles, is able to supply full uncertainty information. Both pointwise and simultaneous confidence bands can be deduced from the Kalman filter formulae. These confidence bands can be applied to trends (μ_t), trend differences ($\mu_t - \mu_s$) , predictions, probabilities of crossing a threshold L (P_t^L), and differences thereof ($P_t^L - P_s^L$). Simultaneous confidence bands have the advantage of resembling the interpretation most users have, namely that the *true* curve will lie between the lower and upper limits. That interpretation does not hold for pointwise intervals.

The approach given in this article, is available in a software implemention in R, called TrendSpotteR. We have given three applications in the field of climate change using this software: the analysis of spring temperatures, global economic losses by weather-related disasters and the analysis of extreme temperatures (Figures 2 through 5).

There are also some drawbacks of the IRW approach. First, the Kalman filter equations are complex to understand which might 'scare off' practitioners. Second, IRW trends do not fit all data. For these cases we have given a number of ways one can proceed such as the application of Box-Cox transformations, recalculation of uncertainty bands based on effective numbers, or the extension of the trend model by adding explanatory variables or cycles, given in Eq. (1). If none of these hints lead to acceptable models, one could choose any other trend model (e.g., trend models given in Chandler and Scott, 2011).

Acknowledgements

The authors wish to thank the beta testers of the TrendSpotteR software: Fentaw Abegaz (RUG - University of Groningen), Arthur Beuse (PBL - Netherlands Environmental Assessment Agency), Leo Soldaat (CBS - Statistics Netherlands) and Guus Velders (RIVM - National Institute for Public Health and the Environment).

References

Both, C. and Visser, M.E. (2001) Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. Nature, 411, 296-298.

Bouwer, L.M. (2011) Have disaster losses increased due to anthropogenic climate change? Bulletin of the American Meteorological Society, 92(1), 39–45.

Chandler, R. and Scott, M. (2011) Statistical methods for trend detection and analysis in the environmental sciences. Chichester: Wiley & Sons Ltd.

Coles, S. (2001) An introduction to statistical modeling of extreme values. London: Springer Verlag.

Commandeur, J.J.F. and Koopman, S.J. (2007) An introduction to state space time series analysis. New York: Oxford University Press.

De Jong, P. and Mackinnon, M.J. (1988) Covariances for smoothed estimates in state space models. *Biometrika*, **75(3)**, 601-602.

Durbin, J. and Koopman, S.J. (2001) *Time series analysis by state space methods*. Oxford: Oxford Statistical Science Series 24.

Harvey, A.C. (1984) A unified view of statistical forecasting procedures. *J. of Forecasting*, **3**, 245-275.

Harvey, A.C. (1989) *Forecasting, structural time series models and the Kalman filter*. New York: Cambridge University Press.

IPCC-SREX (Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.-K., Allen, S.K., Tignor, M. and Midgley, P.M., eds.) (2012) *Managing the risks of extreme events and disasters to advance climate change adaptation. IPCC special report.* Cambridge: Cambridge University Press.
See website: http://ipcc-wg2.gov/SREX.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008) *Modern actuarial risk theory - using R*. Berlin: Springer-Verlag.

Kitagawa, G. (1981) A nonstationary time series model and its fitting by a recursive filter. *J. of Time Series Analysis*, **2(2)**, 103-116.

Liu, W., Lin, S. and Piegorsch, W. (2008) Construction of exact simultaneous confidence bands for a simple linear regression model. *Int. Statistical Review*, **76**, 39-57.

Mills, T.C. (2010) "Skinning a cat": alternative models of representing temperature trends. An editorial comment. *Climatic Change*, **101**, 415-426.

Montgomery, D.C. and Peck, E.A. (1982) *Introduction to linear regression analysis*. Chichester: Wiley & Sons Ltd.

Schrier, G. van der, Van Ulden, A. and Van Oldenborgh, G.J. (2011) The construction of a Central Netherlands temperature. *Climate of the Past*, **7**, 527-542.

Venables, W.N. and Ripley, B.D. (2002) *Modern applied statistics with S*. New York: Springer Verlag.

Visser, H. (1994) *Regression models with time-varying parameters: applications in the environmental sciences*. PhD thesis, University of Amsterdam.

Visser, H. (2004) Estimation and detection of flexible trends. *Atm. Environment*, **38**, 4135-4145.

Visser, H. and Petersen, A.C. (2009) The likelihood of holding outdoor skating marathons in the Netherlands as a policy-relevant indicator of climate change. *Climatic Change*, **93**, 39-54.

Visser, H., Bűntgen, U., D'Arrigo, R. and Petersen, A.C. (2010) Detecting instabilities in tree-ring proxy calibration. *Climate of the Past*, **6**, 367-377.

Visser, H. and Petersen, A.C. (2012) Inferences on weather extremes and weather-related disasters: a review of statistical methods. *Climate of the Past*, **8**, 265-286.

Visser, H., Bouwman, A., Petersen, A.C. and Ligtvoet, W. (2012) *A statistical study of weather-related disasters*. PBL report E555076. To be downloaded from: http://www.pbl.nl/sites/default/files/cms/publicaties/PBL_2012_Weather%20Disasters_555076001_0.pdf

Visser, H. and Brandes, L. (2013) Advanced time-series analysis of selected 19th century tide-gauge data. Journal of Coastal Research, submitted.

Visser, H., Petersen, A.C. and Ligtvoet, W. (2013) On the relation between weather-related disaster impacts and climate change. Climatic Change, submitted.

Wynn, H.P. and Bloomfield, P. (1971) Simultaneous confidence bands in regression analysis. *J. of the Royal Statistical Society. Series B*, **33(2)**, 202-217.

Young, P.C. (2011) *Recursive estimation and time-series analysis. An introduction for the student and practitioner*. Berlin: Springer-Verlag.

Young, P.C., Lane, K., Ng, C.N. and Palmer, D. (1991) Recursive forecasting, smoothing and seasonal adjustment of nonstationary environmental data. *J. of Forecasting*, **10**, 57-89.

Zieba, A. (2010) Effective number of observations and unbiased estimators of variance for autocorrelated data - an overview. *Metrology and Measurement Systems*, **17(1)**, 3-16.

Appendix A Kalman filter - smoothing equations

In this appendix we show how smoothed trend estimates $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, ..., \hat{\mu}_t, ..., \hat{\mu}_N)$ are gained, along with the corresponding NxN covariance matrix $\hat{\Sigma}$. Also the formulae for construction of pointwise confidence bands for trend differences $[\mu_t - \mu_s]$ are given. The Kalman filter equations and definitions will not be presented here in full. The reader is referred to Harvey (1984, 1989) or Visser (1994, Appendix A). We only present the smoothing equations in this Appendix.

We denote the state-space vector $(\mu_t , \lambda_t)'$ from Eq. 3 by α_t . Now, the smoothed estimate for $\boldsymbol{\alpha_t}$ is denoted by $\mathbf{a_{t/N}}$: $\mathbf{a_{t/N}} = E\{\boldsymbol{\alpha_t} \mid y_1, ..., y_N\}$. The matrix $\hat{\sigma}^2_{ML}\, P_{t|N}$ stands for the covariance matrix of $[\mathbf{a_{t/N}} - \boldsymbol{\alpha_t}]$. Here, $\hat{\sigma}^2_{ML}$ stands for the ML estimate of the measurement noise variance. We apply the fixed interval smoother. The equations of this smoother are calculated by a backward recursion, starting at $\boldsymbol{a_{N|N}}$ and $P_{N|N}$, both gained from the final iteration of the Kalman filter equations. The equations read as follows:

$$\begin{aligned} \boldsymbol{a_{t|N}} &= \boldsymbol{a_t} + P_t^*(\boldsymbol{a_{t+1|N}} - T\boldsymbol{a_{t|t}}), \\ P_{t|N} &= P_t + P_t^*\left(P_{t+1|N} - P_{t+1|t}\right)P_t^*, \\ P_t^* &= P_t T' P_{t+1|t}^{-1}, \end{aligned} \tag{A.1}$$

where $P_{t|N}$ is the variance of $\boldsymbol{a_{t|N}}$ at time t, given all measurements $y_1, \dots, y_t, \dots, y_N$. Recursion (A.1) operates backwards in time, up to $t = 1$.

De Jong and Mackinnon (1988) derived the following equation for the covariance matrix $P_{s,t|N}$ between the two state vectors $\boldsymbol{a_{s|N}}$ and $\boldsymbol{a_{t|N}}$ $(1 \le s \le t \le N)$:

$$P_{s,t|N} = P_s^* P_{s+1,t|N}. \tag{A.2}$$

As a boundary condition we have $P_{t,t|N} \equiv P_{t|N}$, for all $1 \le t \le N$. The matrix P_s^* is given in (A.1). Uncertainties in state differences $[\boldsymbol{\alpha_t} - \boldsymbol{\alpha_s}]$ are calculated as follows:

$$Var\left([\boldsymbol{a_{t|N}} - \boldsymbol{a_{s|N}}] - [\boldsymbol{\alpha_t} - \boldsymbol{\alpha_s}]\right) = \hat{\sigma}^2_{ML}\,[P_{t|N} - 2P_{s,t|N} + P_{s|N}] \tag{A.3}$$

with $P_{t|N}$ and $P_{s|N}$ given in (A.1), and $P_{s,t|N}$ given in (A.2).

The smoothed trend estimate $\hat{\mu}$ follows from the first element of the state vector $\mathbf{a_{t/N}}$, given in (A.1), and its variance from the (1,1) element of matrix $\hat{\sigma}^2_{ML}\, P_{t/N}$. The variance of trend differences $[\mu_t - \mu_s]$ directly follows from (A.3) since it is given by the (1,1) element of the right-hand side matrix. If noise processes follow normal distributions, all trend values μ_t will follow normal distributions either. Thus, trend $\boldsymbol{\mu}$ follows a multivariate normal distribution or $\boldsymbol{\mu} \sim N\left(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\right)$.

To generate a trend ensemble of size M, one has to sample from an N-dimensional multivariate normal distribution. The mean vector $\hat{\mu}$ and covariance matrix $\hat{\Sigma}$ are computed as follows:

$\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, .., \hat{\mu}_t .., \hat{\mu}_N)$ with $\hat{\mu}_t = (1 \quad 0)\boldsymbol{a_{t|N}}$ and

$$\hat{\Sigma} = Var(\hat{\mu} - \mu) = \hat{\sigma}_{ML}^2 \begin{pmatrix} P_{1,1|N}^{1,1} & \cdots & P_{1,N|N}^{1,1} \\ \vdots & \ddots & \vdots \\ P_{N,1|N}^{1,1} & \cdots & P_{N,N|N}^{1,1} \end{pmatrix} \quad \text{(A.4)}$$

where $P_{s,t|N}$ is given in Eq. (A.2).

Appendix B Drawing samples from multivariate normal distributions

There are two general approaches to sampling $\boldsymbol{\mu}^{(k)}$ from the N dimensional multivariate normal distribution $N(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$. The first approach is based on conditional distributions (Ripley, 1987). The second approach is followed in the R function mvrnorm(), found in the library 'MASS'. This function is implemented in the TrendSpotteR software. Library 'MASS' supports functions and datasets described in Venables and Ripley (2002). Function mvrnorm() makes use of the fact that, if $V_1, \ldots, V_N$ are iid $N(0,1)$ random variables and A is an NxN matrix, then $\boldsymbol{\mu}^{(k)} = \hat{\boldsymbol{\mu}} + AV$ is multivariate normal with mean $\hat{\boldsymbol{\mu}}$ and covariances $E[\mu_i^{(k)} - \hat{\mu}_i)(\mu_j^{(k)} - \hat{\mu}_j)]$ being the (i,j) element of the covariance matrix. Because of the linearity of the expectation this covariance matrix is equal to

$$E\left[\left(\boldsymbol{\mu}^{(k)} - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}^{(k)} - \hat{\boldsymbol{\mu}}\right)\right] = E[(AV)(AV)^T] = E[AVV^TA^T] = AE[VV^T]A^T = AIA^T = AA^T \quad \text{(B.1)}$$

Thus, to generate random vectors $\boldsymbol{\mu}^{(k)}$, we have to find A, the 'square root' of $\hat{\Sigma}$, or $\hat{\Sigma} = AA^T$. To find matrix A the R function mvrnorm() uses the method of eigenvalue decomposition. For further details please see Kaas et al. (2008).

Appendix C Simultaneous confidence bands

We are trying to find upper and lower confidence limits $\mathbf{w}^{\mathbf{U}} = (w_1^U, \ldots, w_t^U, \ldots w_N^U)$ and $\mathbf{w}^{\mathbf{L}} = (w_1^L, \ldots, w_t^L, \ldots w_N^L)$ such that the *true* trend $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_1, \ldots, \mu_N)$ will lie between $\hat{\boldsymbol{\mu}} - \mathbf{w}^{\mathbf{L}}$ and $\hat{\boldsymbol{\mu}} + \mathbf{w}^{\mathbf{U}}$, with a probability of 1-α. In other words, we have to derive vectors $\mathbf{w}^{\mathbf{U}}$ and $\mathbf{w}^{\mathbf{L}}$ such that:

$$P(\hat{\boldsymbol{\mu}} - \mathbf{w}^{\mathbf{L}} \leq \boldsymbol{\mu} \leq \hat{\boldsymbol{\mu}} + \mathbf{w}^{\mathbf{U}}) = 1 - \alpha \quad \text{(C.1)}$$

The uncertainty band $[\mathbf{w}^{\mathbf{L}}, \mathbf{w}^{\mathbf{U}}]$ is denoted as simultaneous 95% confidence band (α = 0.05). Such bands are difficult to derive analytically. Moreover, the bands are not unique and some extra shape constraints have to be defined. E.g., one can choose $\mathbf{w} \equiv \mathbf{w}^{\mathbf{L}} = \mathbf{w}^{\mathbf{U}}$ and the curves $\hat{\boldsymbol{\mu}} \pm \mathbf{w}$ to be parallel to the curve $\hat{\boldsymbol{\mu}}$. See Wynn and Bloomfield (1971), Montgomery and Peck (1982), Section 9.6) and Liu *et al*. (2008) for more information.

In the context of trend ensembles simultaneous confidence bands are easy to derive. Without loss of generality we will choose M = 1000 and α = 0.05 in the following (these are the settings chosen in TrendSpotteR). As described in Section 2.2 we have to construct an envelope around $\hat{\boldsymbol{\mu}}$ such that the

most extreme α M = 50 simulated trends fall outside this envelop. Once we have chosen these 50 trends from the set of 1000 trends, the lower and upper values w_t^L and w_t^U simply follow from the minimum and maximum of the remaining 950 trend values $\mu_t^{(1)}, \ldots, \mu_t^{(950)}$. Note that this procedure guarantees that exactly 950 trends will lie between $\hat{\boldsymbol{\mu}} - \mathbf{w}^{\mathbf{L}}$ and $\hat{\boldsymbol{\mu}} + \mathbf{w}^{\mathbf{U}}$.

We propose to choose the 50 most extreme trends using the following algorithm:

1. Call the matrix with 1000 simulated trends $T_{m,t}$ with t = 1, ..., N and m = 1, ..., 1000.
2. Rank the 1000 values for each time t from smallest to largest and place these ranks in a new matrix $R_{m,t}$.
3. Compute the relative ranks for each t and place these values in a matrix $R'_{m,t}$ as follows: $R'_{m,t} = \left| R_{m,t} - \frac{1001}{2} \right|$ with m = 1, ..., 1000.
4. For each row in $R'_{m,t}$ add the 5 maximum values and put these values in a 1xN vector $\mathbf{S} = (S_1, \ldots, S_m, \ldots, S_{1000})$, with S_m defined by $S_m = \sum_{5\ maxima\ \forall\ t} [R'_{m,t}]$.
5. The 50 trends that should be removed from matrix $T_{m,t}$ are those having the highest 50 values in vector **S**. Now, define a new matrix by $T_{m,t}'$, with t = 1, ..., N and m = 1, ..., 950.
6. The final step is simply to compute the uncertainty bands $[w_t^L, w_t^U]$ by taking the minimum and maximum of $T_{m,t}'$, m = 1, ..., 950.

The rationale behind our algorithm is that we use ranks in stead of absolute values, and that the number of extreme trends which lie above (or beneath) the estimated trend $\hat{\boldsymbol{\mu}}$ need not be exactly 25. Furthermore, a simulated trend is chosen to be 'extreme' if it is extreme at least in 5 time steps.