The dependence structure of daily UK hydrological time series

[1] University of BRISTOL

Nicholas Howden[1], Tim Burt[1], Fred Worrall[2], François Birgand[1], Sebastian Gnann[1], Ross Woods[1]

[2] Durham University

Introduction

The presence of dependence structure in hydrological time series is both problematic (it prevents the use of standard statistical tools for analysis) and useful (as it indicates a pattern in the data that may be predicted by a model. The dependence structure determines several key characteristics of the time series all of which relate to the way in which the structure contributes to the variance as a function of the autocorrelation function:

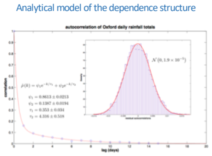
$$\mathrm{Var}\left[\sum_{i=1}^{n} X\right] = n\sigma^2\left\{1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho[k]\right\}$$

The dependence structure tends to inflate the variance, as individual measurements don't provide unique information. This is represented by the **variance inflation factor**:

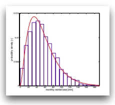
$$\mathcal{F}[n] = 1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho[k]$$

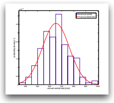
Example: Oxford (UK) Daily Precipitation 1827 - 2016

Analytical model of the dependence structure



Project monthly and annual distributions of P total





The variance inflation factor calculated from:

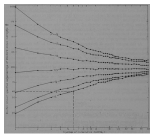
$$\mathcal{F}[n] = 1 + \sum_{i=1}^{2}\left\{\psi_i \vartheta_i[n,\lambda_i]\right\}$$

$$\vartheta_i[n,\lambda_i] = \vartheta_m[\lambda_i]\,\mathcal{C}[n,\lambda_i]$$

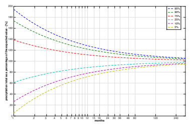
where:

$$\vartheta_m[\lambda_i] = \frac{2}{e^{\lambda_i}-1} \quad \text{and} \quad \mathcal{C}[n,\lambda_i] = 1 - \frac{1}{n}\left(\frac{e^{\lambda_i}}{e^{\lambda_i}-1}\right)\left(1-e^{-n\lambda_i}\right)$$

Forecast long-term P total distributions



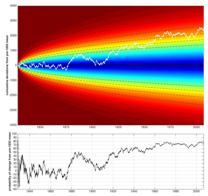
Empirical probability distribution of precipitation total in n consecutive months based on the period 1888 to 1966 and on precipitation amounts assessed over the Oxford area. The curves show the probability of not reaching the amounts specified by the ordinates.



Theoretical probability distribution of precipitation total in n consecutive months based on the dependence structure model for precipitation amounts assessed over the Oxford area.

Test for non-stationarity (trend) in the daily record

We can plot the cumulative deviations from mean daily rainfall (pre-1850), and look at the probability that this has changed:



$$\frac{\mathrm{CI}_{\%}\left[\bar{X}\right]}{\mu} \approx \Phi^{-1}[1-\alpha]\left\{\frac{\hat{\sigma}}{\hat{\mu}}\right\}\sqrt{\frac{\mathcal{F}[n]}{n}}$$

Results show that, although there has been change, there is only an 80% probability of this being significant, less than required by standard statistical tests for change.

We use an analytical model to represent the dependence structure (non-randomness) in hydrological time series from the UK. We can then identify the variance of these series to mainly comprise four distinct components: quick, slow, seasonal and random noise.

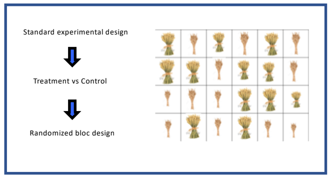
This analytical model allows us to:

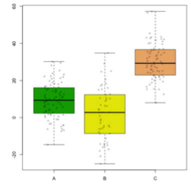
1. **Project the distribution of totals at larger timescales**
   - **forecast stochastic averages**
2. **Identify change-points to a level of probability**
   - **sensitive trend detection and attribution**
3. **Identify the information content of observations**
   - **define a time-domain alternative to the power spectrum**
4. **Demonstrate a robust flow decomposition technique**

The value of integrated hydrological and biogeochemical indicators

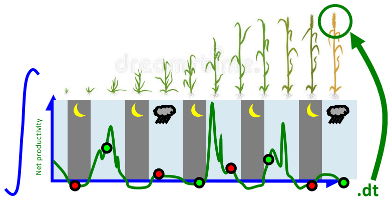
Experiments are designed to find which treatments lead to statistically significant differences in outcomes

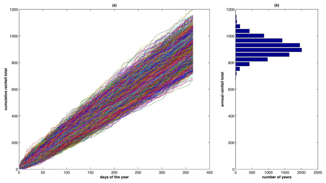
Example: Wheat Yield





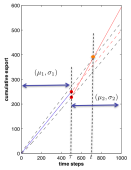
These experiments are monitored using integrated indicators of responses to the treatments:





Integrated indicators tend to smooth out variability and amplify the mean. In the presence of short memory, they also lead to the variance in the integrated indicator being normally-distributed which makes classical statistics more readily applicable.

Simple numerical tests can show how hydrological processes can be considered as integrated indicators, which may enable more sensitive measures of trend detection.



Can determine escape point as:

$$(\mu_1\tau - c\sigma_1\sqrt{\tau}) + \mu_2(t-\tau) = \mu_1 t_e + c\sigma_1\sqrt{t_e}$$

Introduce dimensionless factors:

$$\alpha = \frac{\tau}{t_e} \quad \text{and} \quad \lambda_e = \frac{\mu_2}{\mu_1}$$

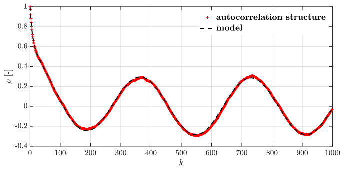
Then: $\lambda_e = \left[\frac{c}{(1-\alpha)\sqrt{t_e}}\right]\left[\frac{\sigma_1}{\mu_1}\right]\left[\sqrt{\alpha}+1\right]$

And if: $\gamma = \frac{\sigma_2}{\sigma_1}$ then:

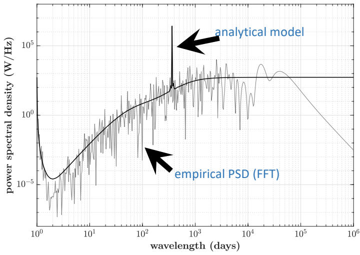
$$\lambda = \left[\frac{c}{(1-\alpha)\sqrt{t}}\right]\left[\frac{\sigma_1}{\mu_1}\right]\left[\sqrt{\alpha} + \gamma\sqrt{1-\alpha} + 1\right] + 1$$

Example: River Thames Streamflow 1883 to 2015

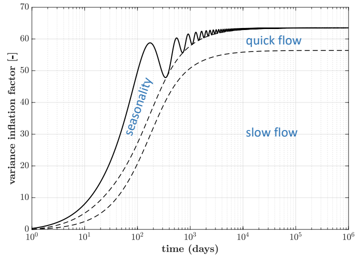
Analytical model fitted to the autocorrelation of the Thames streamflow data:



The contribution of different component frequencies to the variance is usually determined using the power spectral density. Here we can use our data to produce this, and produce a version predicted by our analytical model:



And we can do the equivalent in the time domain using our new approach:



Application for flow decomposition

The three flow components in UK river systems…