Stochastic models for hydro catchment inflows; an exploratory analysis

Peter Thomson
Statistics Research Associates

Outline

1. Background

2. Data

3. Review of streamflow models

4. Exploratory data analysis

5. Evaluation

Joint work with David Harte (SRA) and Mark Pickup (NZ Electricity Commission).

1. Background

The security of New Zealand's electricity supply is largely dependent on future annual patterns of water inflows into New Zealand's major hydro catchments.

The NZ Electricity Commission oversees New Zealand's electricity industry and markets. It needs to estimate the risk of extreme annual sequences of weekly inflows so that it can take steps to mitigate the effect of dry years.

In particular, the Commission is responsible for managing the electricity sector so that demand can be met in a 1-in-60 dry year without the need for emergency conservation campaigns.

The Commission contracted SRA to evaluate the feasibility of building a predictive model for weekly catchment inflows that

- captures the stochastic properties of historic inflow sequences sufficiently accurately

to be suitable for

- risk forecasting, particular of extremes;
- simulating realistic forward sample paths;

over seasonal to multi-year timescales.

Risk forecasting rather than point forecasting implies that

- construction of predictive distributions important since these provide the required probability estimates of risk

and so need to

- accurately specify suitable (marginal) inflow distributions; heavy–tails important to model if present; [Statics]

- build in sufficient persistence or clustering of inflows to generate suitable extremes within phases of seasonal and longer term cycles. [Dynamics]

The dynamic specification also needs to address the issues of timescale.

2. Data

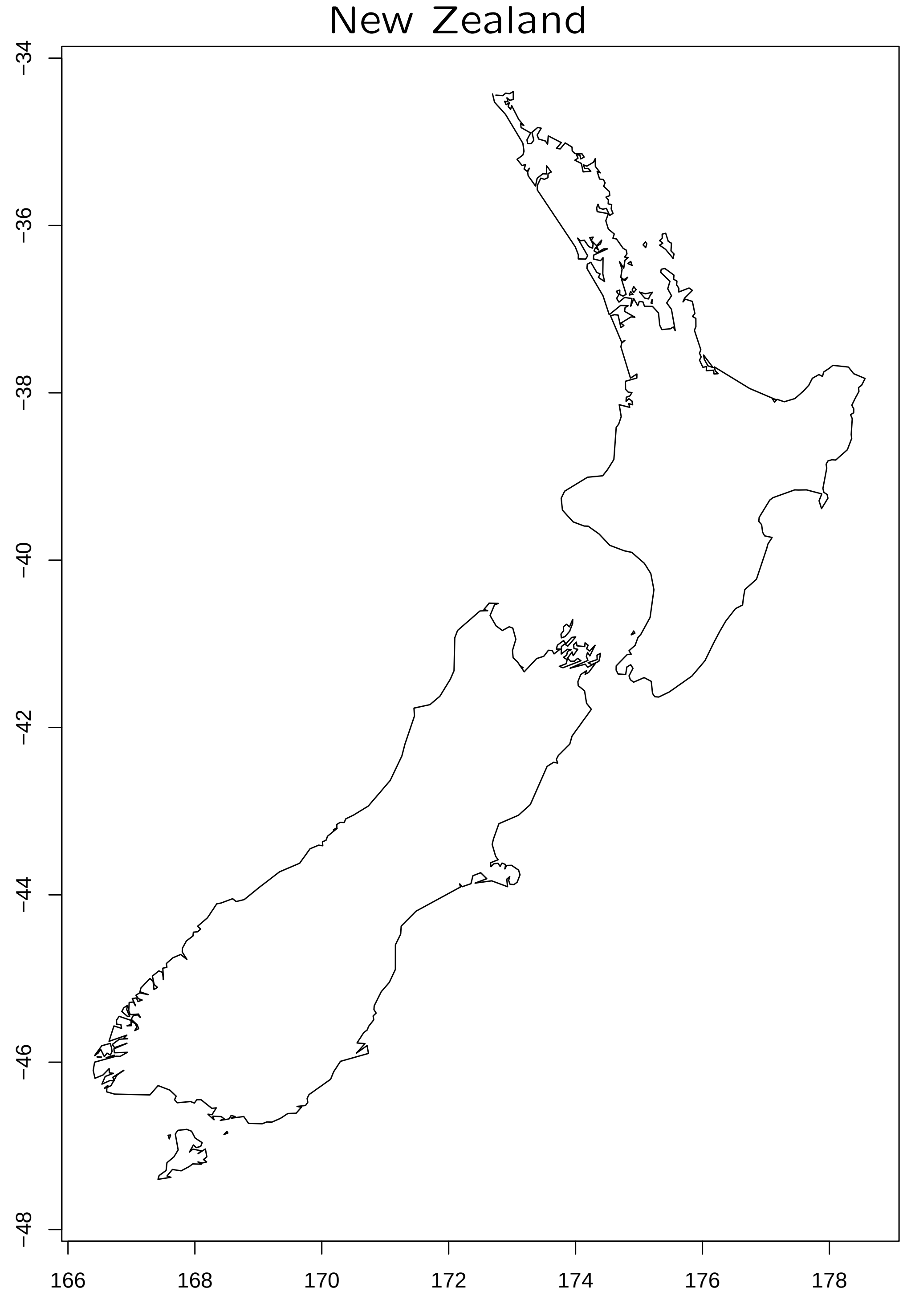
- Weekly inflows (cumecs) and their equivalent power generation potential (GWh) from 1931 to 1995.

- Lakes considered were Arapuni, Benmore, Cobb, Coleridge, Hawea, Karapiro, Manapouri, Mangahao, Matahina, Ohau, Pukaki, Rangipo, Taupo, TeAnau, Tekapo, Tokaanu and Waikaremoana.

- Flow rates calculated assuming present electricity generation system (with its dams, canals, etc) has been in place since 1931.

- Caveats on data quality; data anomalies and competing data sources led to recommendation that data construction, collection and measurement be reviewed.

- Seasonality a general feature with different patterns between North and South Islands, and within the South Island.


New Zealand
−34
−36
−38
−40
−42
−44
−46
−48
166
168
170
172
174
176
178

3. Review of streamflow models

Many attempts to model streamflows with mixed results.

- Extensive literature; see Hipel and McLeod (1994), Salas (1993) and a more recent summary in Srinivas and Srinivasan (2004).

- Two major strands: time series (simulation) models and application of statistical theory of extremes. The latter is not considered since it doesn't deal with sequences of inflows, a key requirement.

- No single candidate model clearly dominates. Perhaps to be expected; streamflow models for large continental rivers may well be quite different from those for New Zealand's rivers and more frequent rainfall.

- The requirement to model higher frequency weekly inflows is an important consideration. Most of the published analyses of streamflows are for monthly flows.

- Stochastic streamflow model adopted should ideally be posited within a sound statistical framework. Some simulation models are difficult to calibrate.

Modelling considerations

Streamflow time series typically exhibit:

- seasonal variation in mean levels, standard deviations and autocorrelation structure.

- highly skewed distributions that can vary with season (data often transformed to Gaussianity using shifted log transform

$$\log(Y_t - \theta_t)$$

  where $\theta_t = \theta_{t-52}$ or $\theta_t = \theta$ or $\theta_t = 0$);

- longer term variation, often associated with climate forcing variables such as El Niño (ENSO) and the interdecadal Pacific oscillation (IPO) etc.

All are present in the NZ data, as we shall see.

3.1 Parametric linear models

Linear Gaussian time series models have been adapted for use in streamflow modelling, in almost all cases to transformed data.

Periodic autoregressive (PAR) models and more general PARMA models are often used to model hydrological time series. For example, a weekly time series Y_t follows a PAR(1) model if

$$Y_t = \mu_t + \rho_t(Y_{t-1} - \mu_{t-1}) + \sigma_t\epsilon_t$$

where μ_t, σ_t, ρ_t are periodic with

$$\mu_t = \mu_{t-52}, \qquad \sigma_t = \sigma_{t-52}, \qquad \rho_t = \rho_{t-52}$$

and ϵ_t is white noise. PAR(p) and PARMA(p,q) models are defined analogously.

Thomas and Fiering (1962) among the first to use a PAR(1) model within hydrology. However model dates back further to Hannan (1955) who used it to model Sydney rainfall.

PARMA models:

- have a fixed mean seasonal pattern plus noise where the latter has fixed seasonal standard deviations and short–memory seasonal periodic autocorrelation;

- are not stationary although annual vectors of weekly streamflows are (VARMA);

- are reasonably well understood, although stochastic properties and estimation remain active research topics;

- potentially involve a large number of parameters, especially the higher frequency weekly inflows, leading to constrained PPAR and PPARMA models;

- include ARMA models as a special case.

Multisite parametric linear models

PARMA models readily generalise to multisite applications, but number of parameters increases dramatically.

Leads to restricted contemporaneous vector PARMA models (Salas, 1993; Hipel and McLeod, 1994) which are single site models driven by contemporaneously correlated white noise.

Other multisite models include:

- parametric covariance model (Koutsoyiannis, 2000) which can model major short and long–memory autocorrelation structure;
- disaggregation models which preserve second–order properties of multisite streamflows over short and long timescales

Note the need for models that can properly account for annual as well as weekly time scales.

3.2 Nonlinear and nonparametric models

Parametric linear models have met with mixed success, partly because they cannot easily reproduce any nonlinear behaviour in streamflows.

This has led to a number of nonlinear and nonparametric models including:

- Srinivas and Srinivasan (2004) who fit PAR(1) models at each site, randomly resample blocks of PAR(1) residuals (block bootstrap) and then reconstitute synthetic sequences of streamflows;

- Lewis and Ray (2002) who extend the PAR(p) models to incorporate nonlinear behaviour using thresholds.

An example of the latter is the regime switching threshold autoregressive (TAR) model

$$Y_t = \begin{cases} \mu_t + \rho_t(Y_{t-1} - \mu_{t-1}) + \sigma_t\epsilon_t & (Y_{t-1} > \alpha_t) \\ \tilde{\mu}_t + \tilde{\rho}_t(Y_{t-1} - \tilde{\mu}_{t-1}) + \tilde{\sigma}\epsilon_t & (Y_{t-1} \leq \alpha_t) \end{cases}$$

where μ_t, $\tilde{\mu}_t$, ρ_t, $\tilde{\rho}_t$, σ_t, $\tilde{\sigma}_t$ and α_t are periodic.

Such nonlinear models typically provide more realistic inflow sequences than linear models.

However, like the linear models, most of the nonlinear models proposed have strictly periodic means and standard deviations. As a consequence, they model departures from fixed seasonal patterns.

Point process models have also been proposed for streamflows. These are reviewed in Salas (1993) with an example of more recent work given in Abi–Zeid, Parent and Bobée (2004).

Hidden Markov models (HMM) have been widely used to model rainfall. They are nonlinear regime switching models which can handle hierarchical timescales (Sansom and Thomson, 2001) and can be conditioned to climate variables (Bellone, Hughes and Guttorp, 2000).

Despite their simple structure and utility, parametric nonlinear HMM models do not yet appear to be widely used for streamflow modelling.

4. Exploratory data analysis

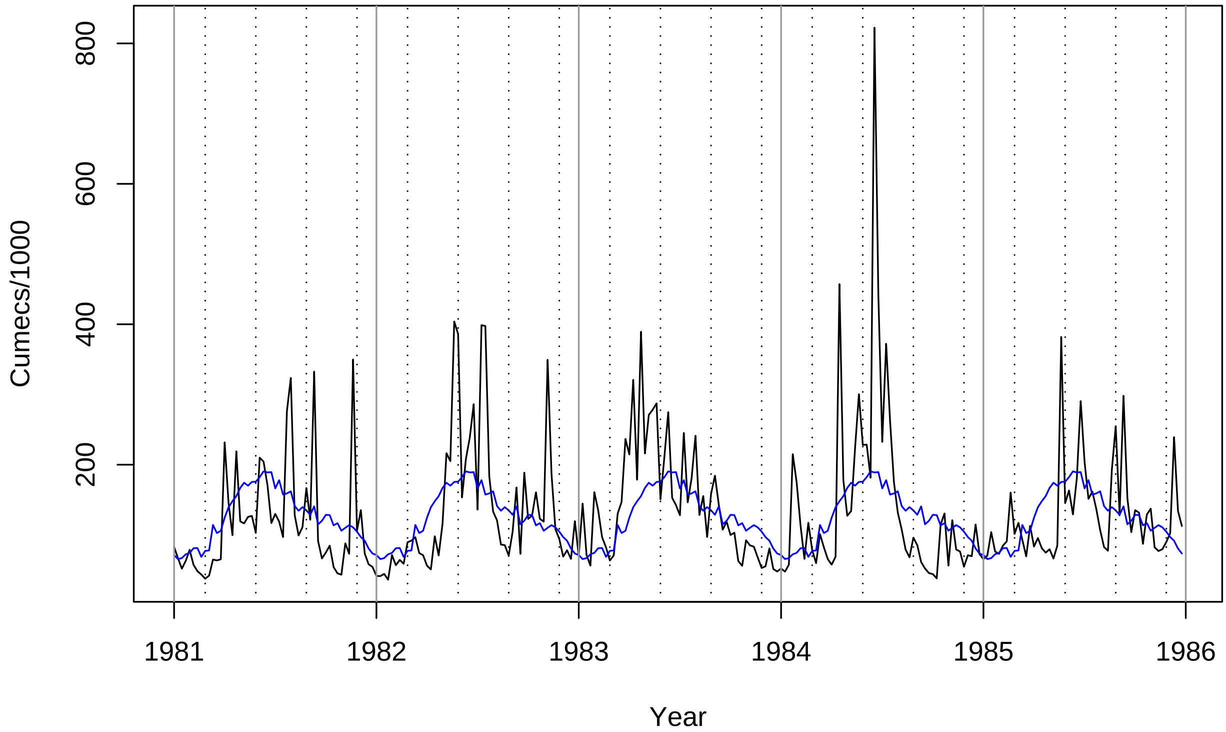
Benmore considered since its inflows exemplify other NZ inflow series.

Weekly data available from 1931 to 1995, a total of 65 years.

A first look at the data

Benmore inflows:

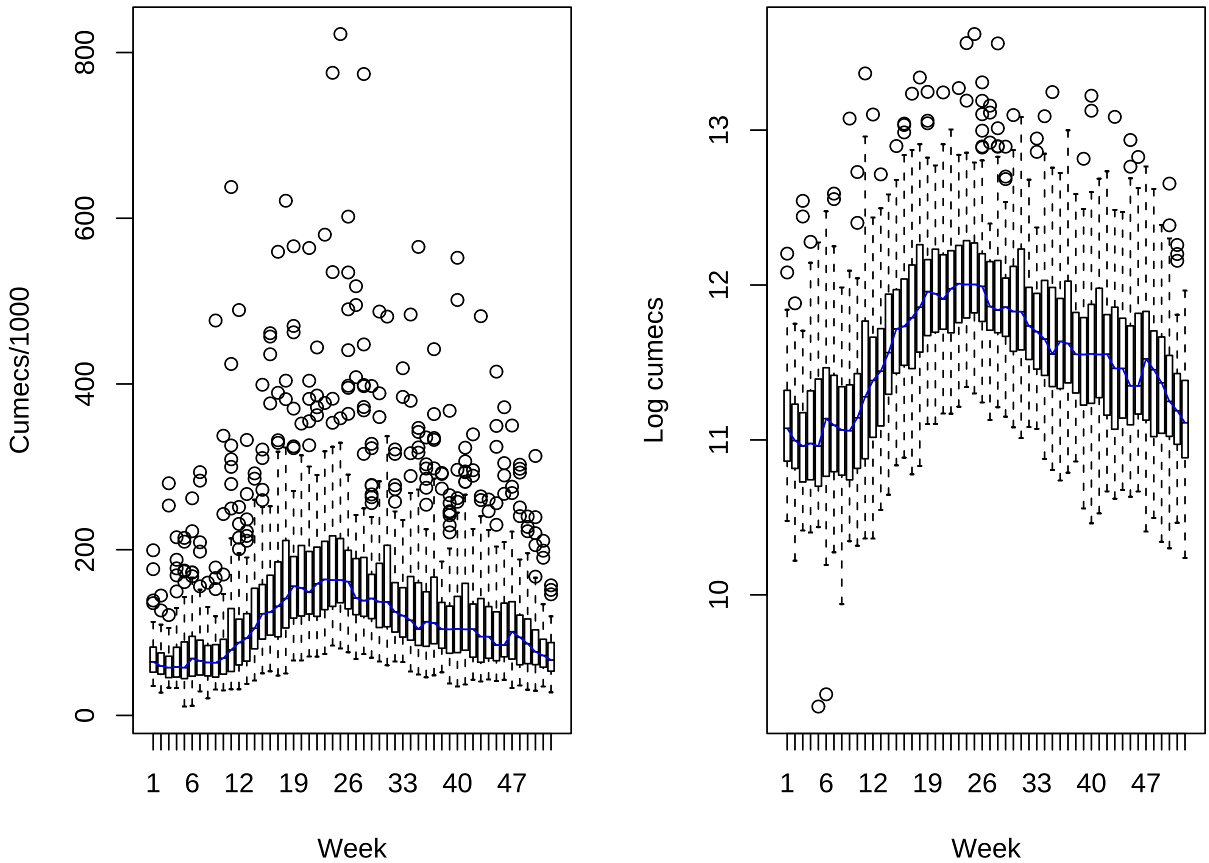
- are episodic in nature
- with an evolving seasonal pattern that
- appears to switch abruptly between levels at times which are earlier or later than those of the fixed mean seasonal pattern.



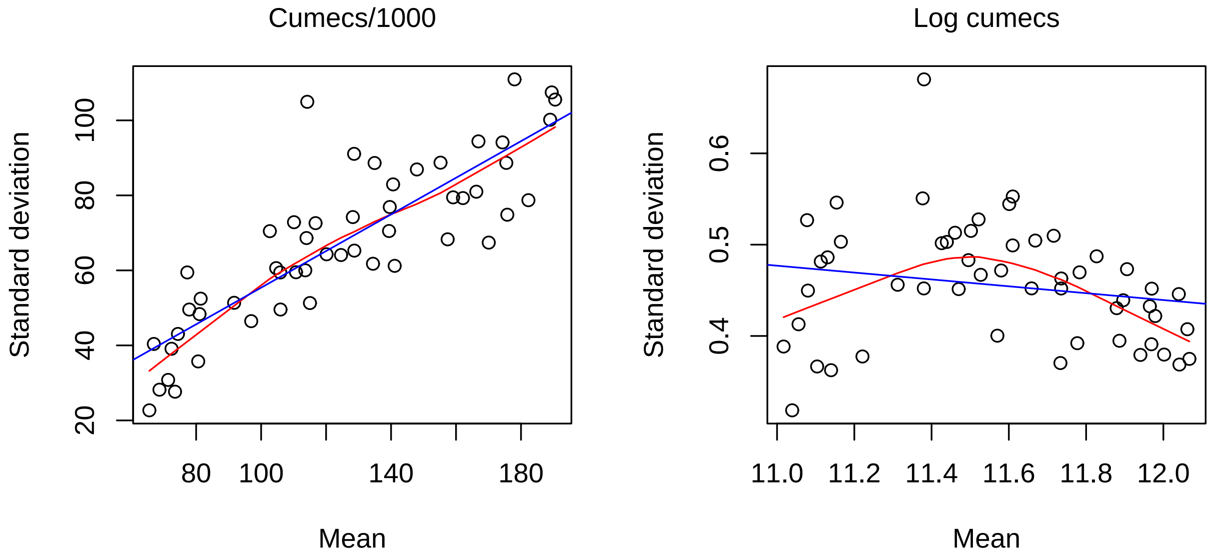
Benmore inflows over the 5 year period starting in the first week of 1981 with the mean seasonal pattern superimposed (blue).

The solid vertical grey lines mark calendar years and the dotted vertical lines mark nominal seasons.

Data transformation

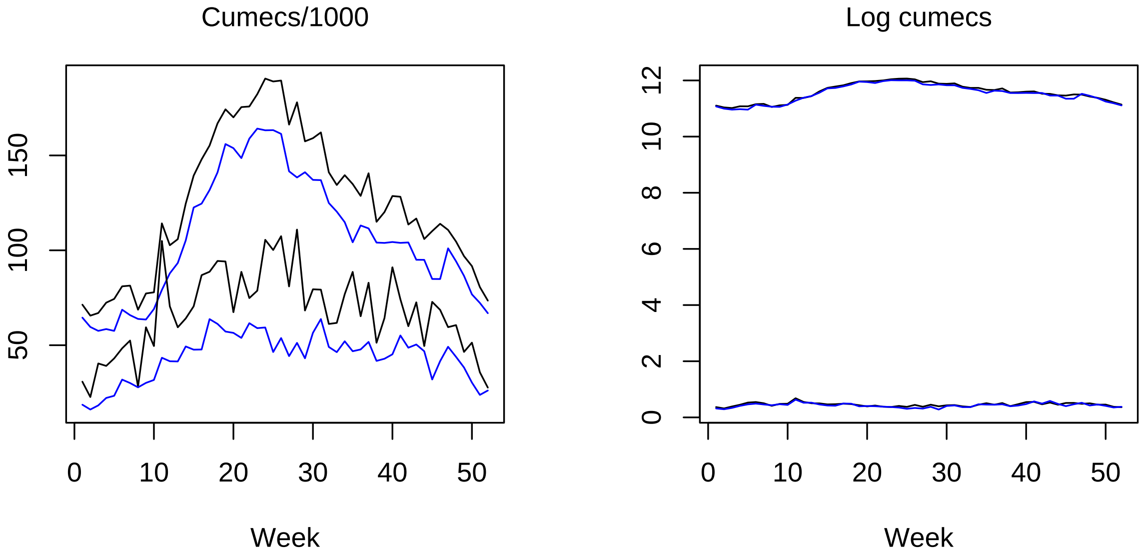


Boxplots of Benmore inflows (left panel) and logs of inflows (right panel) by week of the year with median seasonal patterns superimposed (blue).



Plots of standard deviations against means for each of the 52 series of Benmore inflows (left panel) and logs of inflows (right panel) by week of the year.

Robust *loess* smooths (red) and least–squares regression lines (blue) are superimposed.



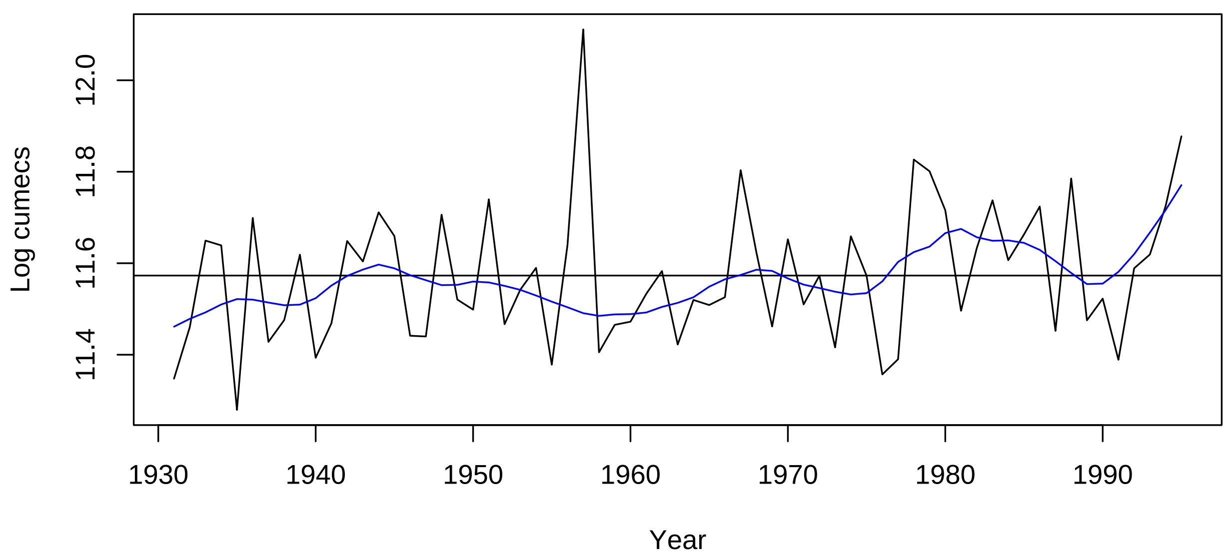
Plots of the means, standard deviations, medians and MADs for each of the 52 series of Benmore inflows (left panel) and logs of inflows (right panel) by week of the year.

The top two lines in each panel are the weekly means (black) and medians (blue); the bottom two lines are the weekly standard deviations (black) and MADs (blue).

Marginal distribution of Benmore inflows by week of the year appears to follow a three–parameter lognormal distribution and so an appropriate data transformation is the shifted log transformation.

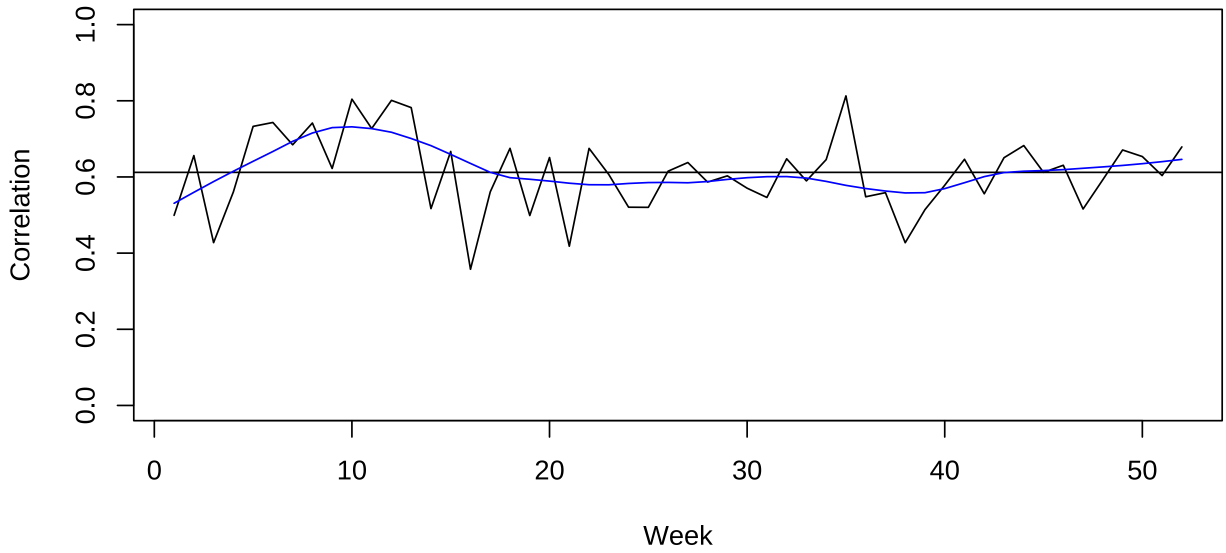
Dynamics

Annual timescale



Annual means of the logs of Benmore inflows. The horizontal line is the overall mean and a robust *loess* smooth (blue) has been superimposed (using a span of 11 years).

Evidence of long–term variability in level, possibly associated with ENSO or IPO (McKerchar et al 1998, 2003).



Correlations between the annual series for each week of the year and the series for the preceeding week for the logs of Benmore inflows. The horizontal line is the overall mean correlation and a robust smooth (blue) has been superimposed.

Supports PAR and PARMA framework and periodic correlation more generally.

Transition periods (autumn or spring) sometimes show a "bent stick" rather than linear relationship. Due to variable length seasons?

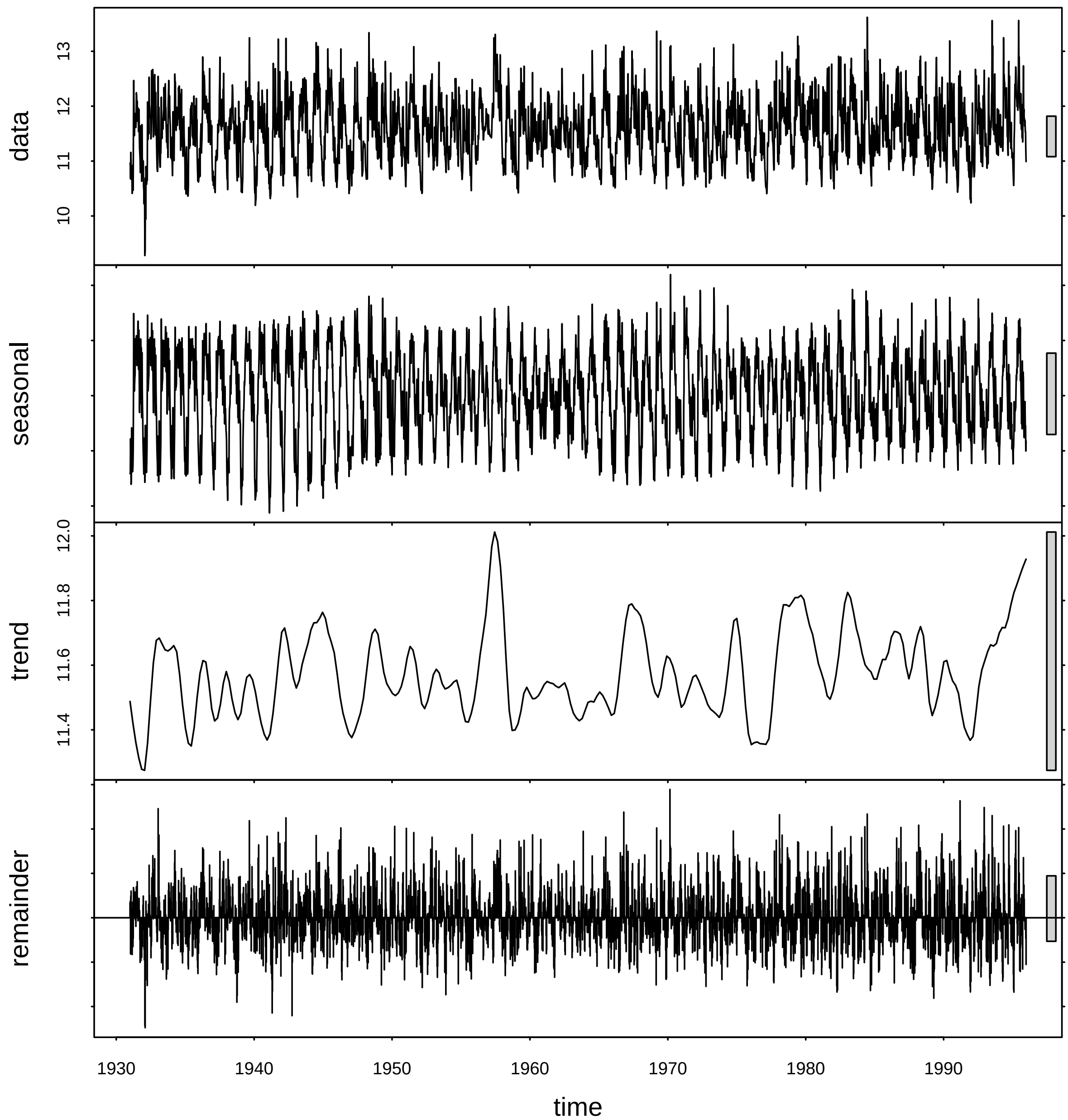
Weekly timescale

Now focus on the evolution of the log weekly inflows Y_t over weekly timescales and model Y_t as

$$Y_t = T_t + S_t + \epsilon_t$$

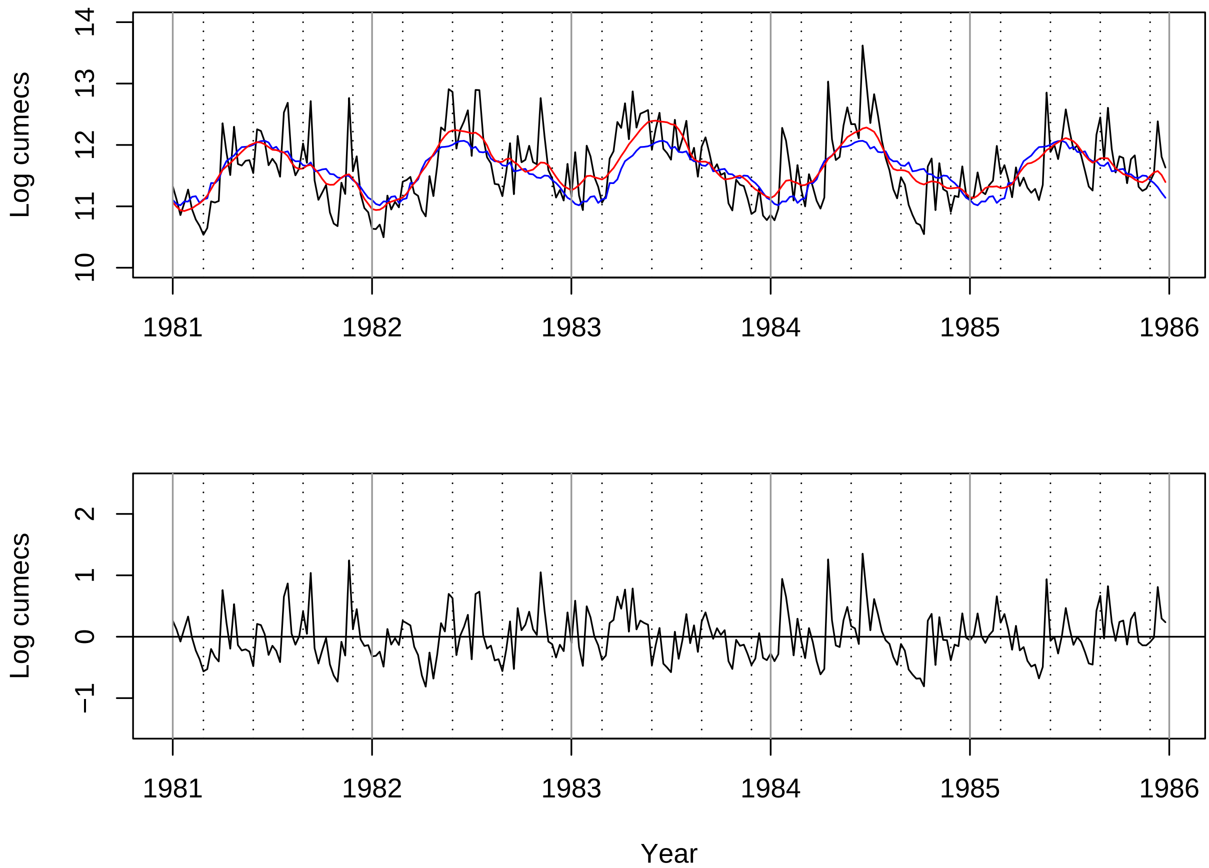
where T_t denotes an evolving trend or level, S_t denotes an evolving seasonal, and ϵ_t denotes the non–systematic irregular component.

One view of T_t and S_t is given by the STL decomposition of Y_t calculated using a moving trend window of 103 weeks (essentially 2 years) and a seasonal window of 7 years.



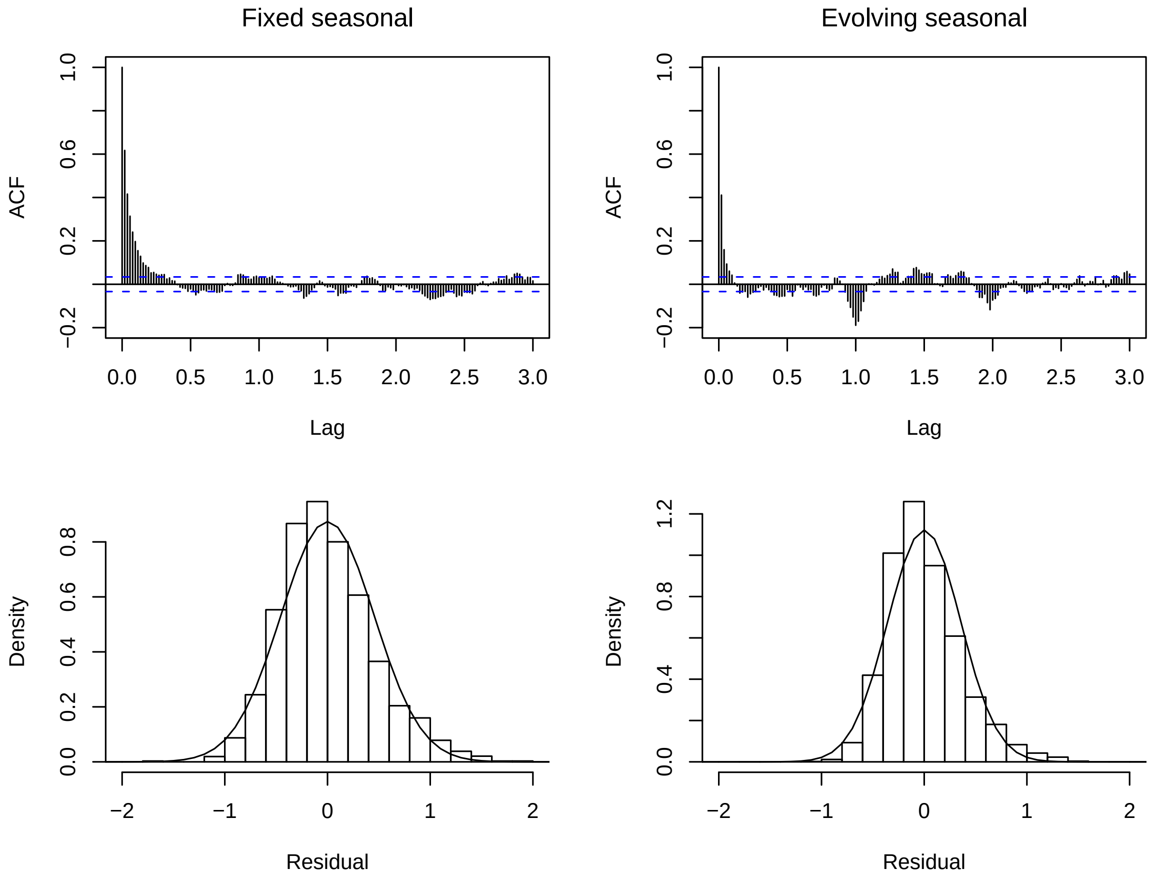
STL decomposition of the logs of Benmore inflows into trend, seasonal and irregular.

Plot too dense to see week to week movements and quality of estimated seasonal.



Logs of Benmore inflows (black) are plotted in the upper panel with the mean seasonal pattern (blue) and evolving seasonal systematic component $T_t + S_t$ (red) superimposed. The irregular (black) is plotted in the lower panel with a horizontal zero reference line superimposed.

Evolving seasonal systematic component better than fixed seasonal, but still not adaptive enough.



Autocorrelation functions and histograms of the residuals of the logs of Benmore inflows after fitting a fixed mean seasonal pattern (left panels) and evolving seasonal systematic component (right panels). A best fitting Gaussian distribution has been superimposed on the histograms.

The longer memory of residuals from fixed seasonal due mainly to evolving seasonal systematic component.

5. Evaluation

Key factors that any New Zealand model will need to take into account include:

- intra–annual weekly dynamics that are dominated by evolving seasonal patterns that are episodic in nature and which can switch abruptly between regimes at times that can be earlier or later than expected;

- inter–annual variability that may well be associated with climate forcing variables such as ENSO and IPO;

- highly skewed marginal distributions that vary according to season and which would appear to be well–modelled by the three–parameter lognormal distribution;

- joint distributions that exhibit periodic seasonal correlation.

SRA has recommended that:

- the linear parametric PARMA model and the nonparametric model of Srinivas and Srinivasan (2004) be fitted and their performance assessed; [Benchmark]

- a nonlinear HMM model with stochastically switching seasons be developed that extends the basic PARMA structure. Its sample path and risk forecasting performance should be evaluated against suitable benchmarks. [New research]

Comments and suggestions welcomed!

References and pdf version of review available from Peter Thomson (peter@statsresearch.co.nz).