Hidden Markov models for New Zealand hydro catchment inflows: a preliminary analysis

David Harte and Peter Thomson
Statistics Research Associates

Outline

1. Background
2. Strategy
3. A non-seasonal HMM for infows
4. Discovery data analysis
5. Seasonal HMMs for inflows

1. Background

The security of New Zealand's electricity supply is largely dependent on future annual patterns of hydro catchment inflows.

The NZ Electricity Commission oversees New Zealand's electricity sector. It is responsible for ensuring that demand can be met in a 1-in-60 dry year without the need for emergency measures.

The Commission 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.

The Commission wishes to develop a stochastic model for weekly inflows that

- captures the properties of historic inflows sufficiently accurately to be suitable for

- risk forecasting, particular of extreme sequences, and

- simulating realistic forward sample paths over

- seasonal to multi-year timescales.

Harte, Pickup and Thomson (2004) and Harte and Thomson (2006) recommended that a nonhomogeneous seasonal HMM (Hidden Markov Model) be developed for NZ weekly inflows that models the episodic seasonal regimes observed in the data.

This presentation reports on a preliminary HMM analysis undertaken for the Commission by SRA.

2. Strategy

Model outlier-corrected, weekly inflows X_t as

$$\phi_t(X_t) = \mu + T_t + \sigma Y_t$$

where

$$\begin{aligned} \phi_t(.) &= \text{suitable transformation} \\ \mu &= \text{long-term mean level} \\ T_t &= \text{smoothly evolving trend deviation} \\ Y_t &= \text{standardised, trend-adjusted, transformed data.} \end{aligned}$$

Objective: Use a homogeneous HMM to explore

- intra-annual seasonal dynamics of Y_t;
- potential structure of seasonal HMMs for Y_t.

Comments

- No attempt to model inter-annual trend T_t at this stage.

- Previous inflow analyses confirm episodic nature of seasonal regimes and desirability of switching models.

- Harte and Thomson (2006) show that $\phi_t(x) = \log(x - \theta_t)$ ($\theta_t = \theta_{t+52}$) eliminates the extreme skewness present in X_t, allowing HMM to focus on regime switching dynamics.

- The non-seasonal, homogeneous HMM fitted needs to be simple and sufficiently robust to reliably classify persistent states.

Although a non-seasonal HMM is inappropriate for out-of-sample prediction, its state classifications are useful for exploring the in-sample stochastic properties of seasonal inflow regimes.

3. A non-seasonal HMM for inflows

Consider modelling the transformed weekly inflows Y_t using the non-seasonal HMM

$$Y_t = \mu_{S_t} + \sigma_{S_t} Z_t$$

where

- S_t is an unobserved stationary Markov chain taking on values $1, \ldots, N$;
- Z_t is a Gaussian $AR(1)$ process, independent of S_t, with $E(Z_t) = 0$, $\text{Var}(Z_t) = 1$, $\text{cov}(Z_t, Z_{t-1}) = \rho$;
- $E(Y_t|S_t) = \mu_{S_t}$, $\text{Var}(Y_t|S_t) = \sigma^2_{S_t}$.

In general S_t is specified by $N(N-1)$ transition probabilities. For $N = 4$ this yields 12 parameters to estimate for S_t alone. Too expensive unless N is small. A major weakness!!

Following Buckle, Haugh and Thomson (2004) we restrict attention to $N = 4$ and specify S_t by two independent 2-state Markov chains C_t, V_t where

S_t	C_t	V_t	C_t regime	V_t regime	μ_{S_t}	σ_{S_t}
1	0	0	Low	Low	μ_1	σ_1
2	0	1	Low	High	μ_2	σ_2
3	1	0	High	Low	μ_3	σ_3
4	1	1	High	High	μ_4	σ_4

Now S_t has 4 parameters compared to 12 for the general chain.

BHT model is adopted to explore episodic seasonal regimes and dynamic structure of weekly inflows.

Note that S_t can be regarded as

- an approximation to a more general 4 state Markov chain; or
- a structural model of the underlying hydrological process.

Examples.

- General: States S_t have means that approximately follow a cyclic seasonal sequence.
- Structural: Seasonal C_t indexes primary flow regimes (high flow/charge and low flow/discharge) with non-seasonal V_t describing secondary flow characteristics (eg volatility).

4. Exploratory data analysis

Two representative hydro catchments considered.

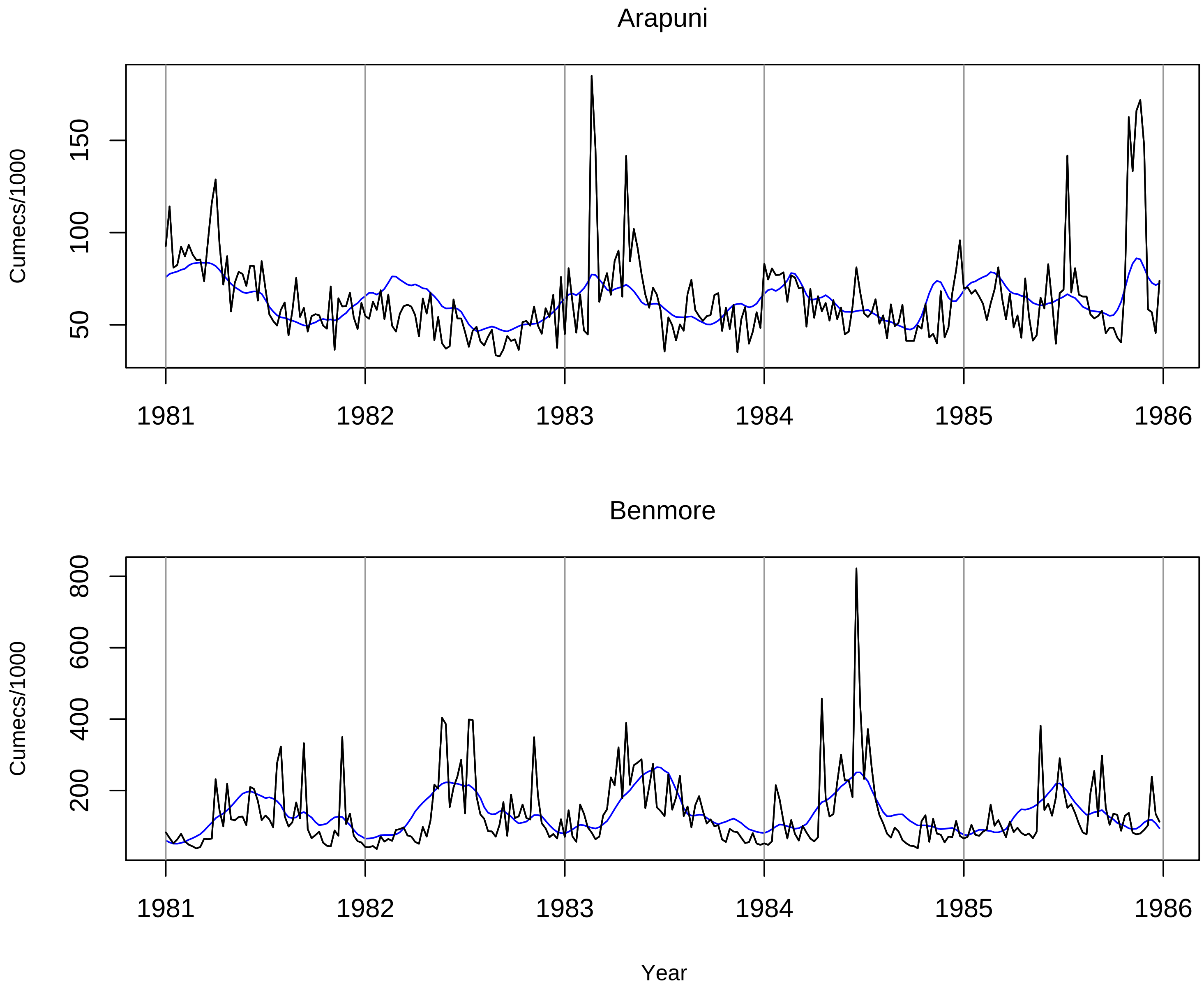
- Arapuni is dominated by North Island rainfall and lower topography.
- Benmore is dominated by South Island rainfall and snow as well as high mountains.
- 74 years of weekly inflows (1931–2004).
- Data prior adjusted for outliers (Arapuni 4; Benmore 3).

Inflows

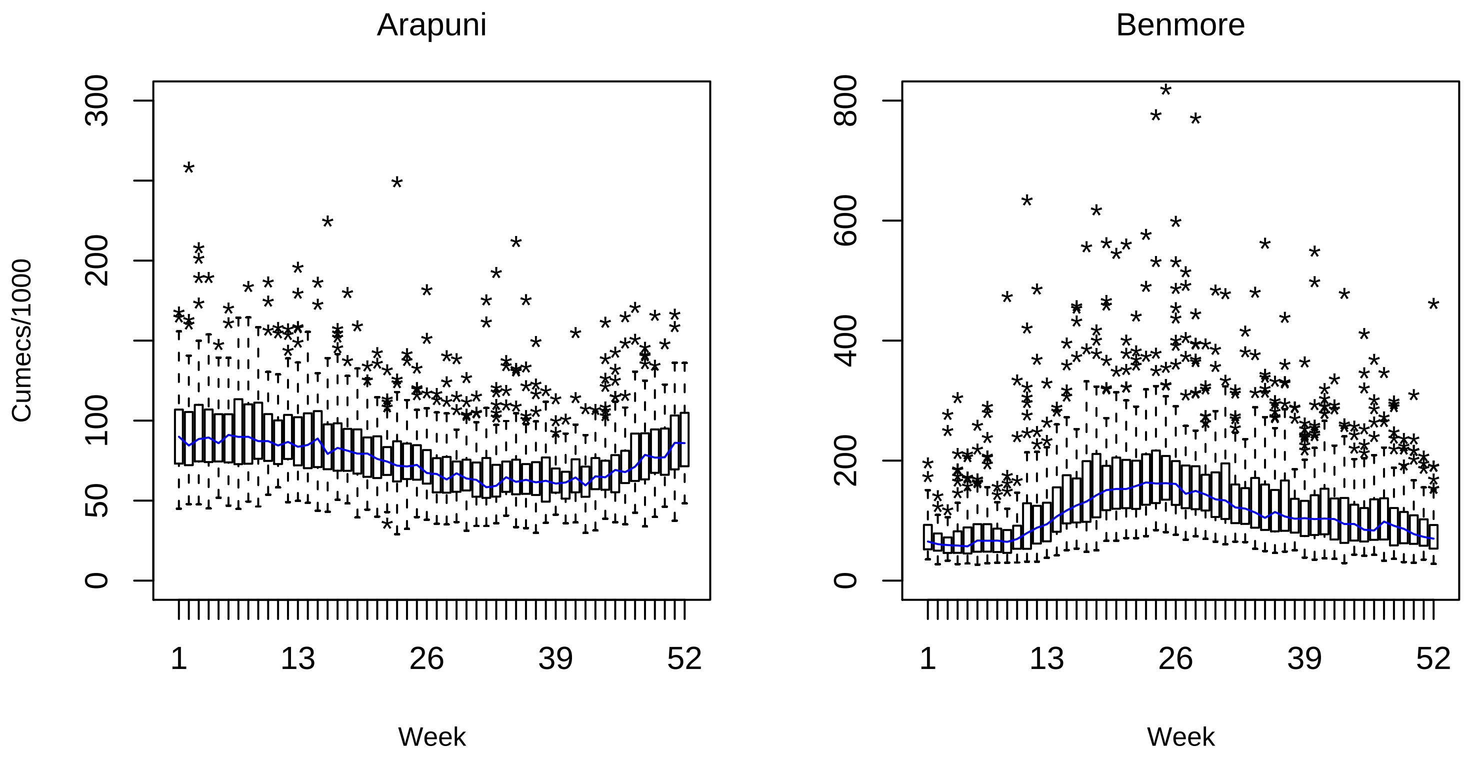
5 year sample

STL Trend

+ STL Seasonal



Boxplots of inflows by week of year



Transformation

Inflows X_t transformed by shifted-log transformation

$$W_t = \log(X_t - \theta_t)$$

where $\theta_t = \theta_{t+52}$ is minimum possible inflow.

Estimated using local maximum likelihood with moving estimation window of 13 weeks. Three local models considered.

- H&T 2006: $\theta_t = \theta$ and W_t independent Gaussian with constant mean and constant variance for each week;

- Modified: $\theta_t = \theta$ and W_t independent Gaussian with separate mean and variance for each week;

- Smooth modified using smoothness weights in log-likelihood.

A global (52 week) modified model was also fitted.

Estimates of θ_t

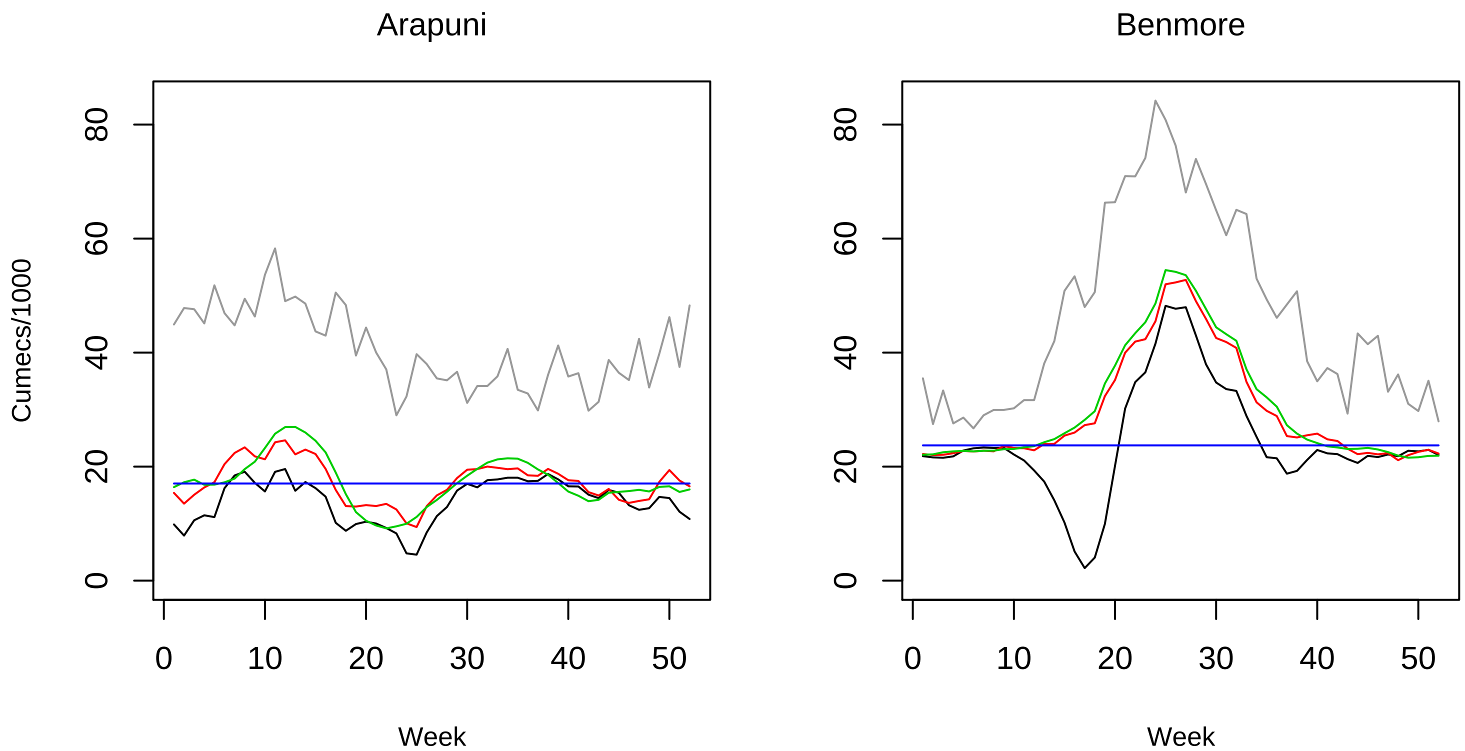
H&T 2006

Modified

Smooth modified

Global

Minimum



Global constant model $\theta_t = \theta$ chosen since it

- is a significant improvement over $\log X_t$ ($\theta_t = 0$);
- is a simple non-seasonal, time-homogeneous transformation;
- is similar in effect to modified estimates.

The W_t are now trend adjusted to give

$$W_t = \mu + T_t + \sigma Y_t$$

with T_t estimated by STL. This yields the standardised, trend-adjusted, transformed inflows

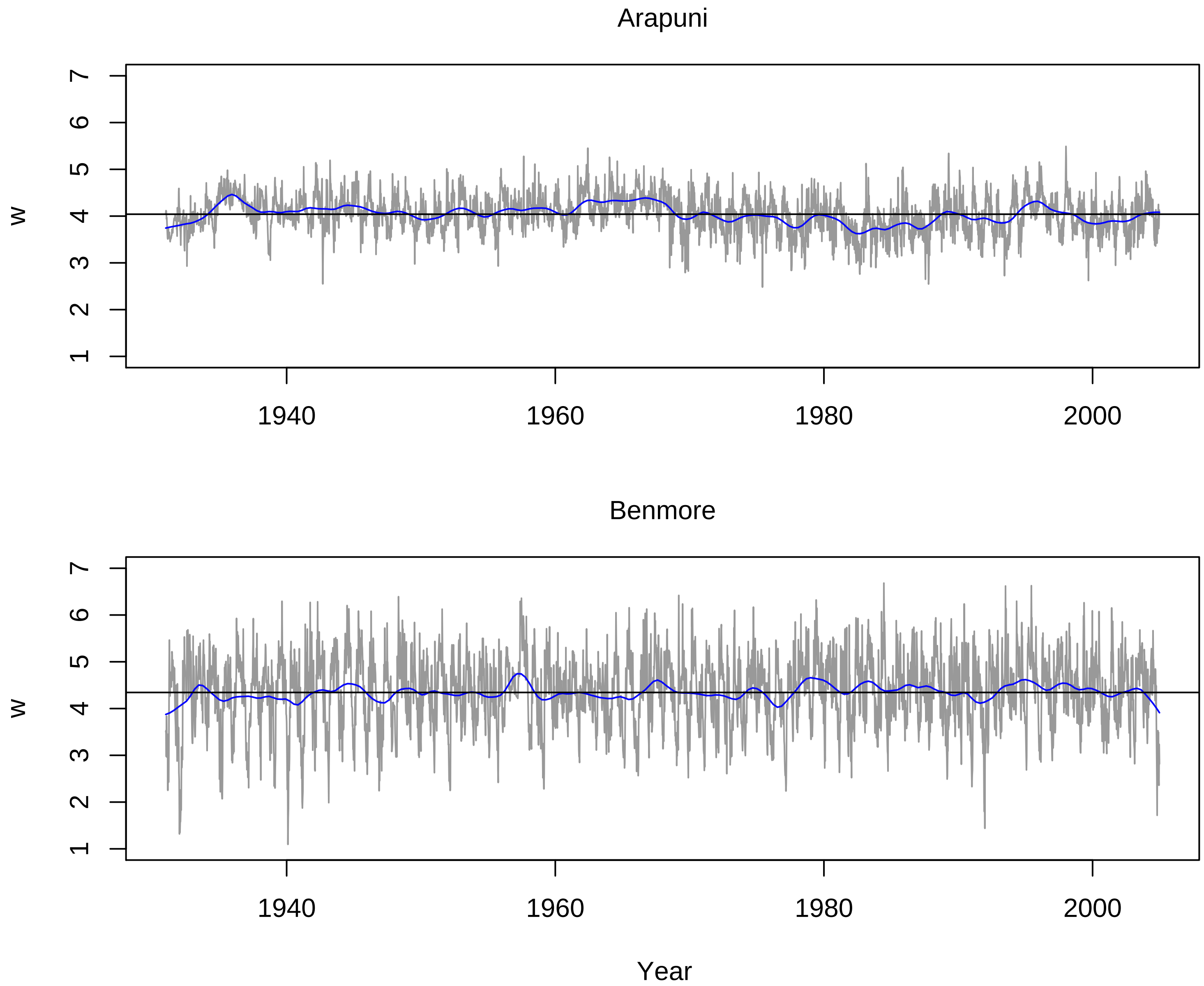
$$Y_t = (\log(X_t - \theta) - \mu - T_t)/\sigma.$$

The BHT model is fitted to the transformed weekly inflows Y_t.

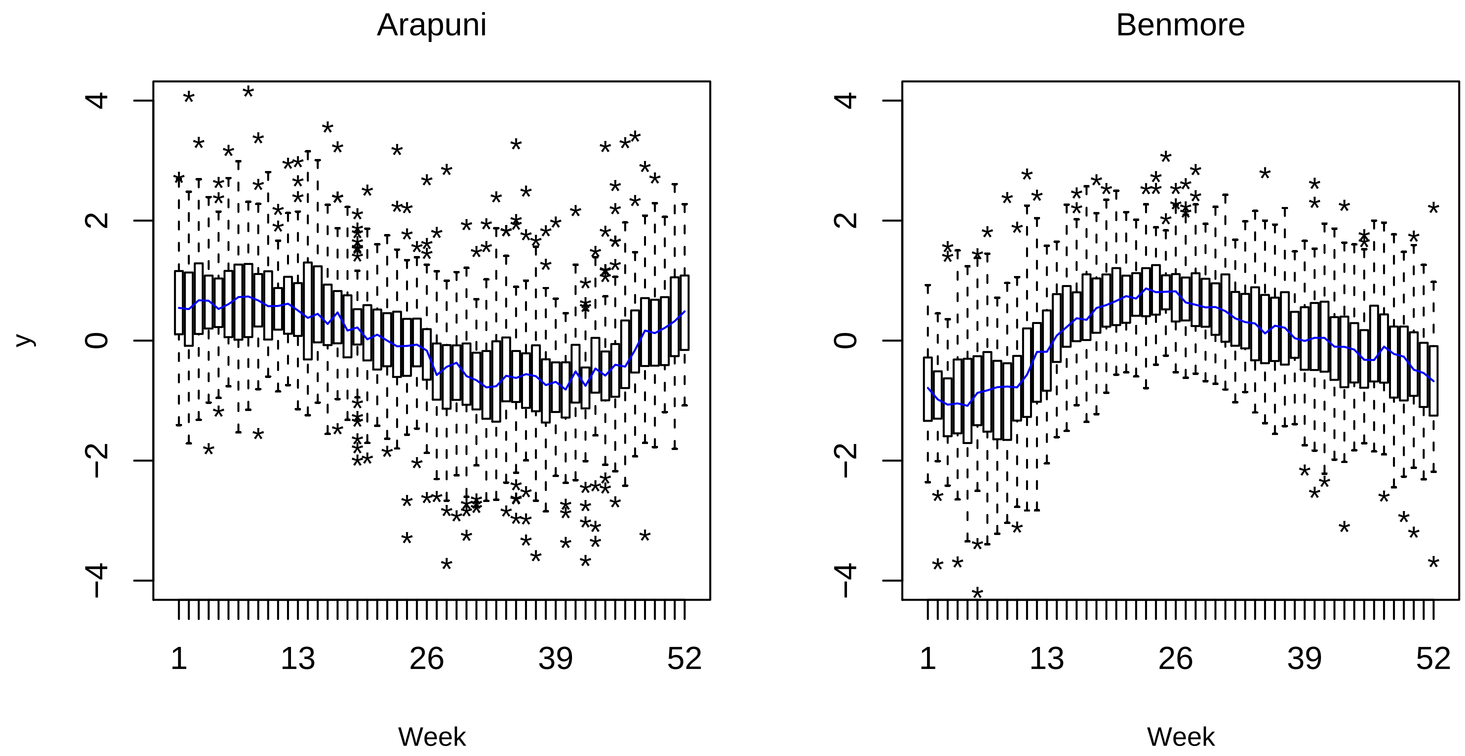
Transformed inflows

STL Trend

Mean



Boxplots of transformed inflows by week of year



Model fitting

BHT model fitted using EM and ML with model choice guided by AIC. Log-likelihood and EM depend on classification probabilities

$$\gamma_t(j) = P(S_t = j|\mathbf{Y}), \quad \gamma_t(j,k) = P(S_t = j, S_{t+1} = k|\mathbf{Y})$$

where $\mathbf{Y}$ denotes the data $Y_1, \ldots, Y_T$.

The $\gamma_t(j)$, $\gamma_t(j,k)$ are useful in their own right and also to extract S_t dependent quantities. For example, the best estimate of μ_{S_t} given the data is

$$E(\mu_{S_t}|\mathbf{Y}) = \sum_{j=1}^{4} \mu_j \gamma_t(j)$$

which we call the HMM trend.

The classification probabilities are used in this way to construct suitable diagnostics.

After re-labelling, the fitted parameters led to the mapping

S_t	C_t	V_t	C_t regime	V_t regime	μ_{S_t}	σ_{S_t}
1	0	0	Low	High	μ_1	σ_1
2	0	1	Low	Low	μ_2	σ_2
3	1	0	High	High	μ_3	σ_3
4	1	1	High	Low	μ_4	σ_4

where

$$\mu_1 < \mu_2 < \mu_4 < \mu_3$$

and

$$\sigma_1 = \sigma_3 > \sigma_2 = \sigma_4$$

Note mean flow hierarchy.

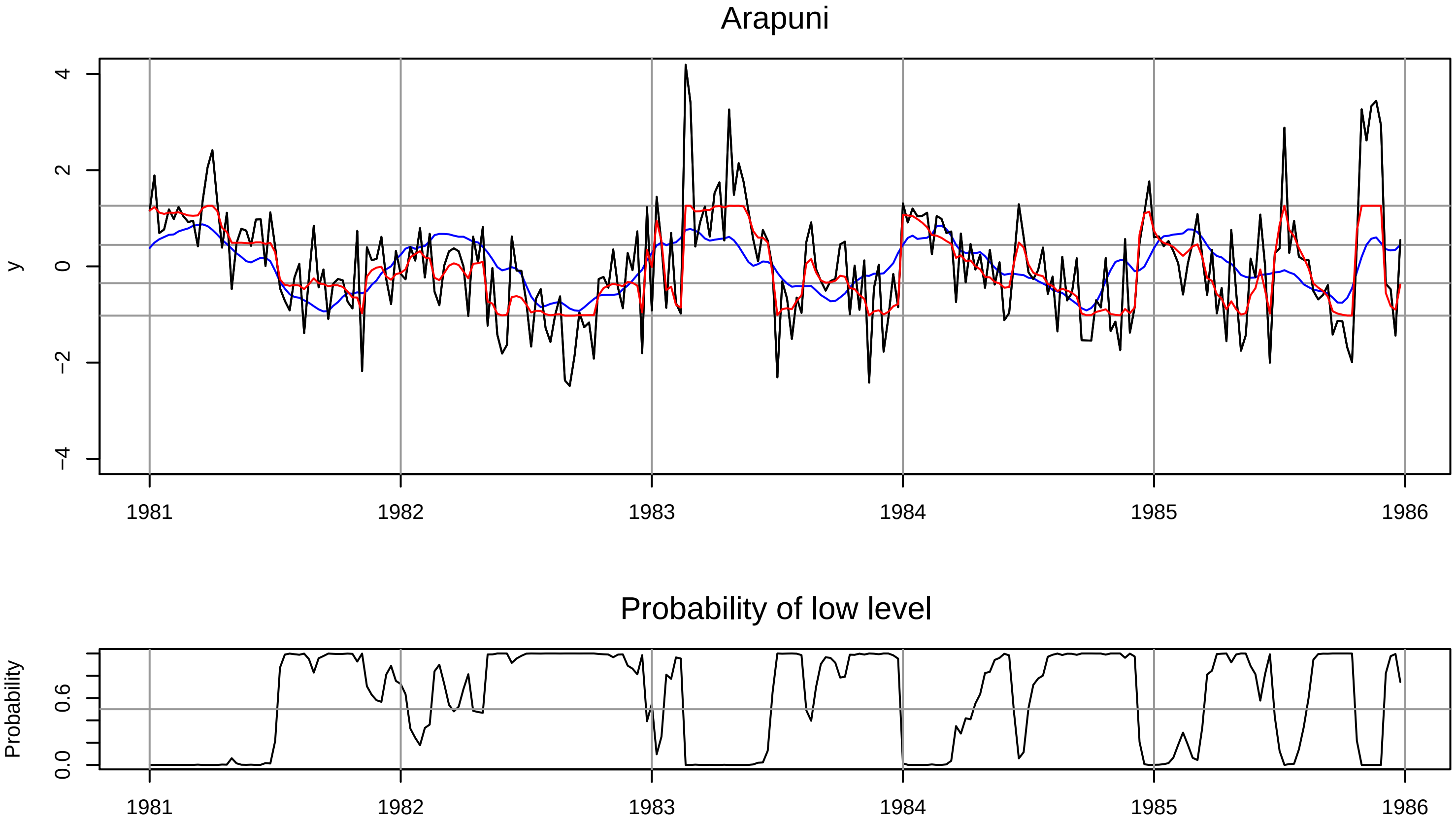
Arapuni transformed inflows (5 year sample)

STL Trend

+ STL Seasonal

HMM trend

Fitted means



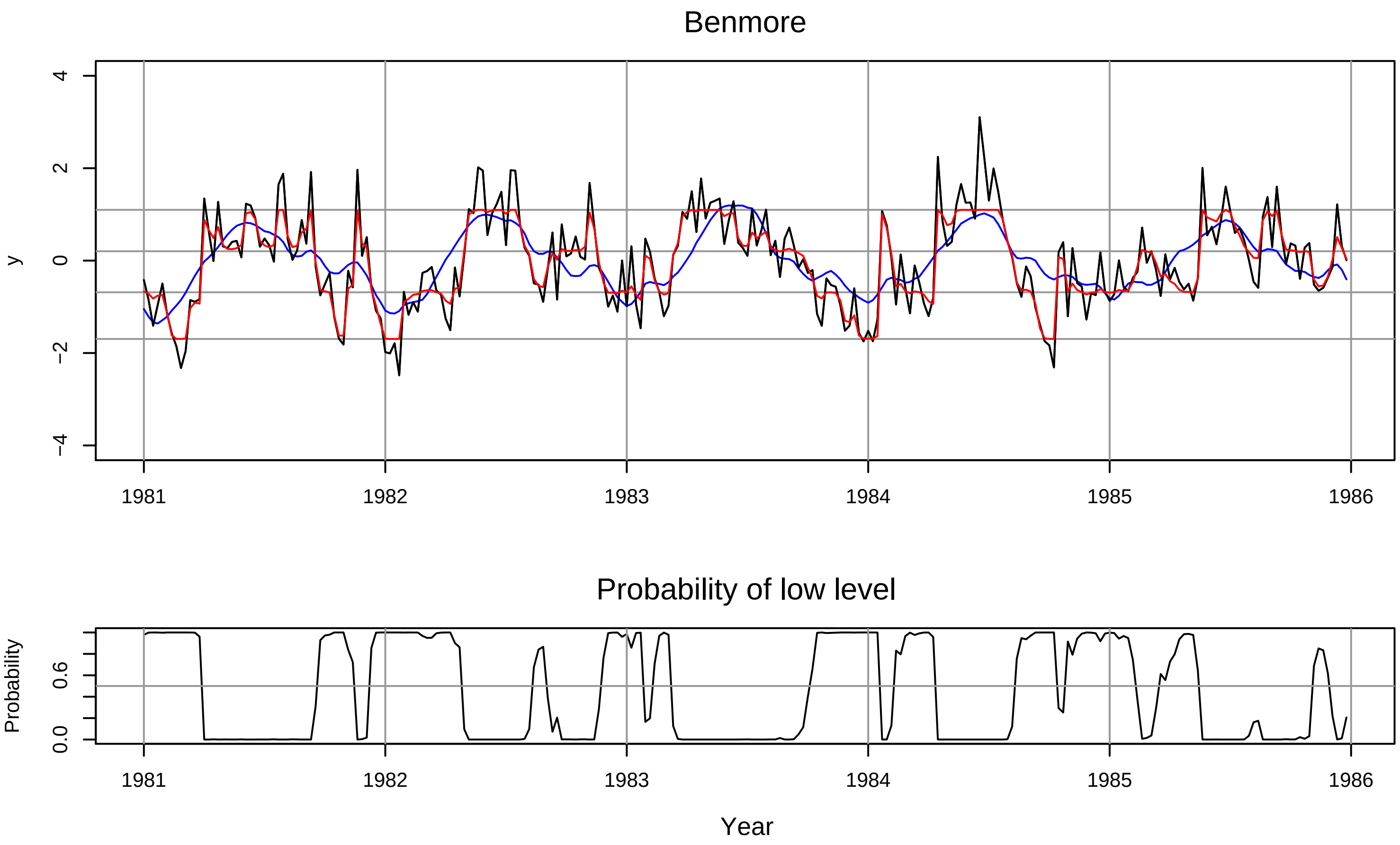
Benmore transformed inflows (5 year sample)

STL Trend

+ STL Seasonal

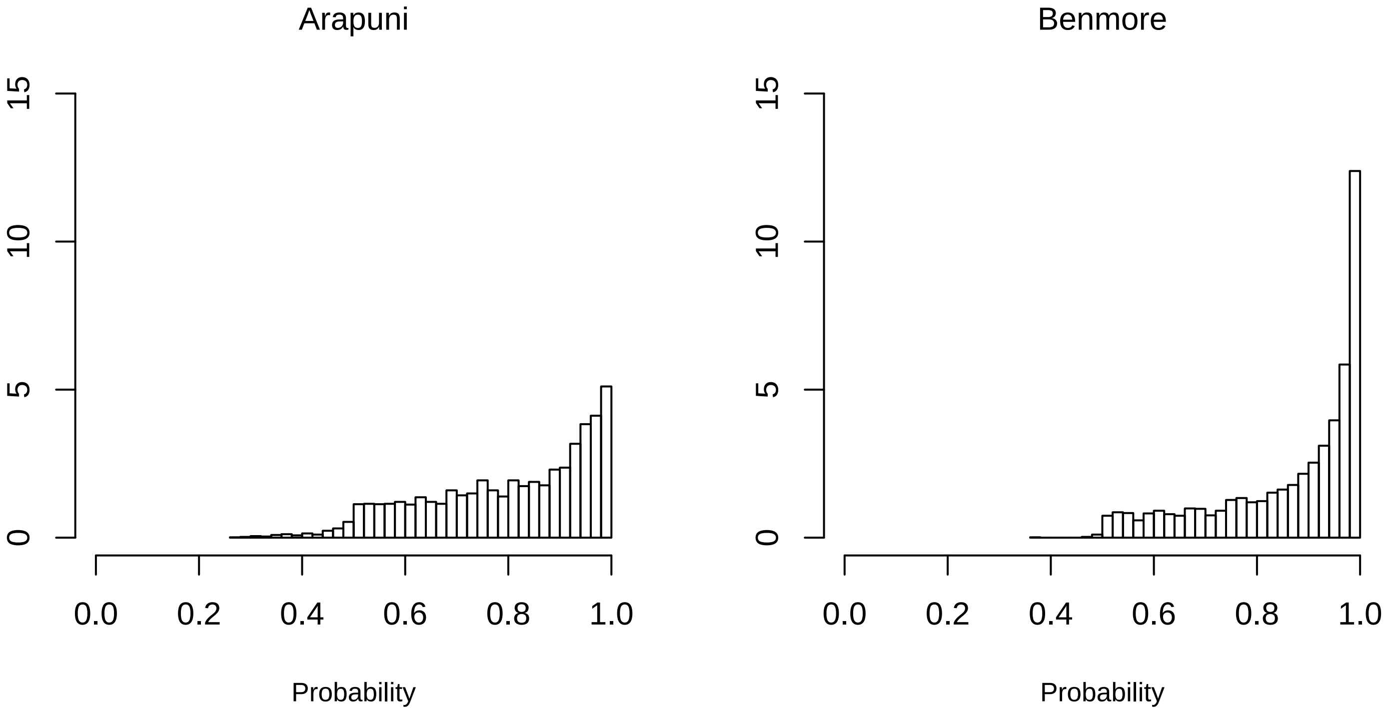
HMM trend

Fitted means



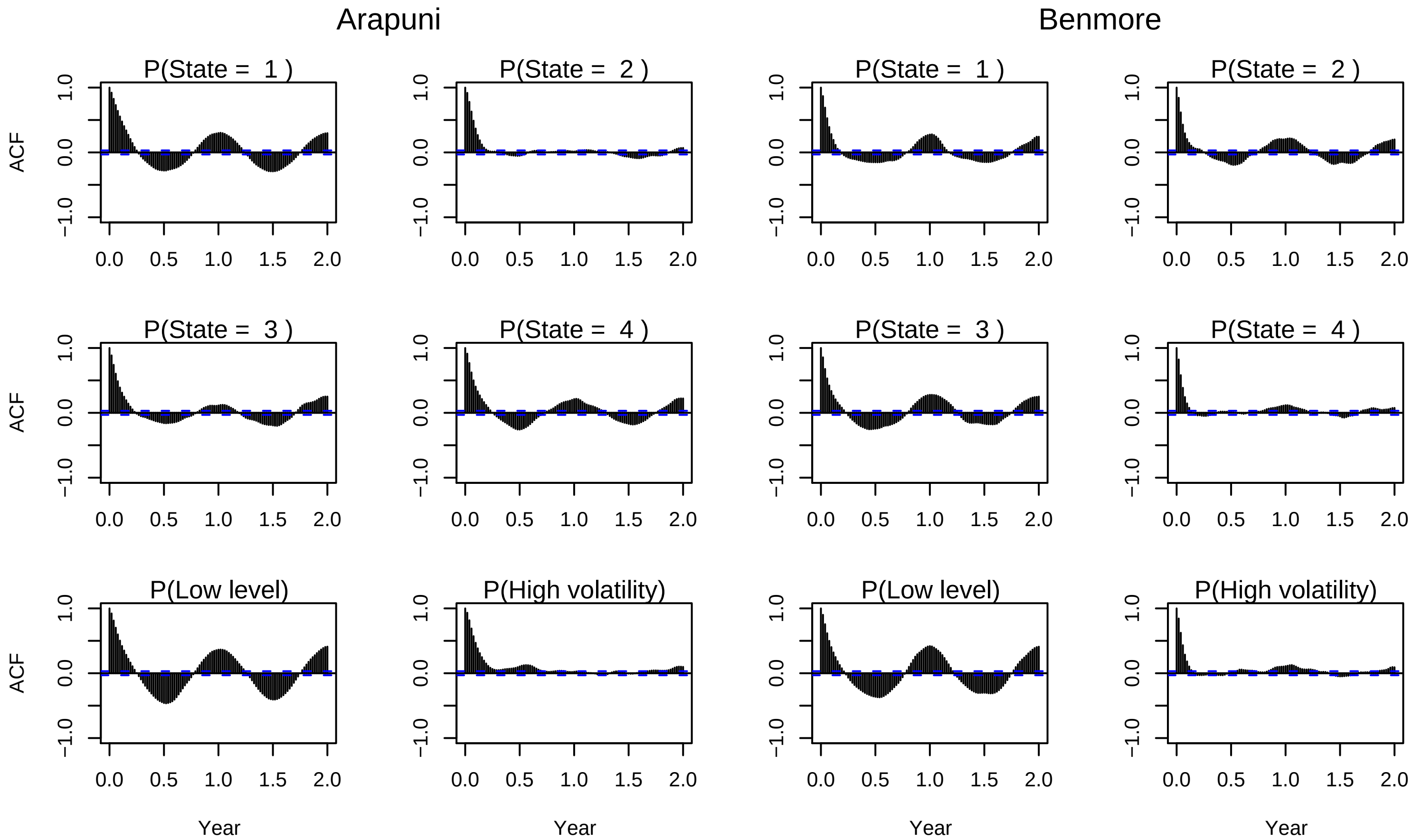
Transformed inflows

Histograms of maximum classification probabilities



Transformed inflows

ACFs of classification probabilities



For any state j, another view of its seasonality is provided by the proportion, over years, of visits to state j in week k of the year $(k = 1, \ldots, 52)$.

The best predictor of this proportion is

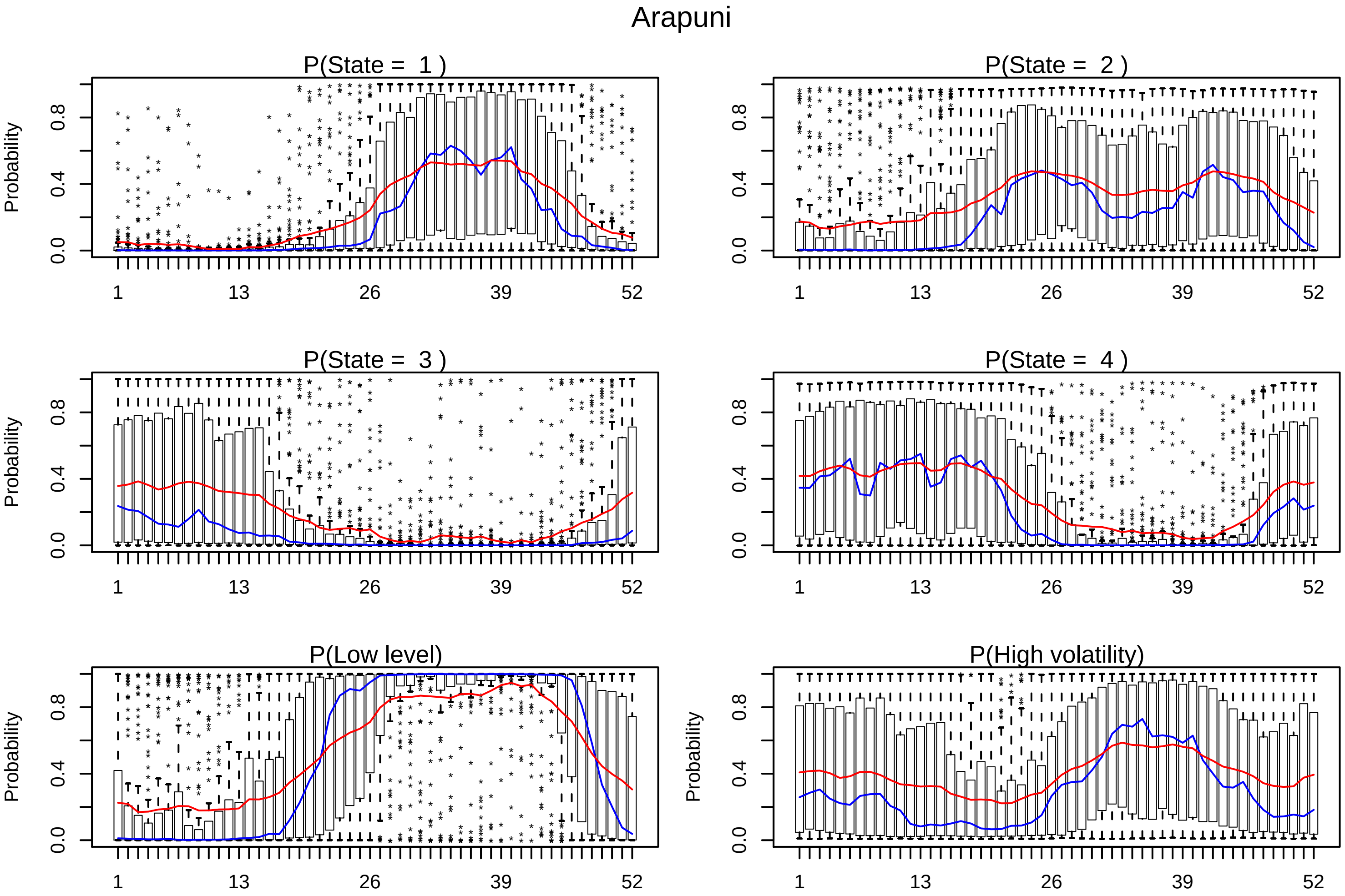
$$\frac{1}{74} \sum_t{}' \gamma_t(j) = \text{seasonal mean of } \gamma_t(j)$$

where $\sum_t'$ is over those t with week of the year k.

Arapuni transformed inflows

Boxplots of classification probabilities by week of the year

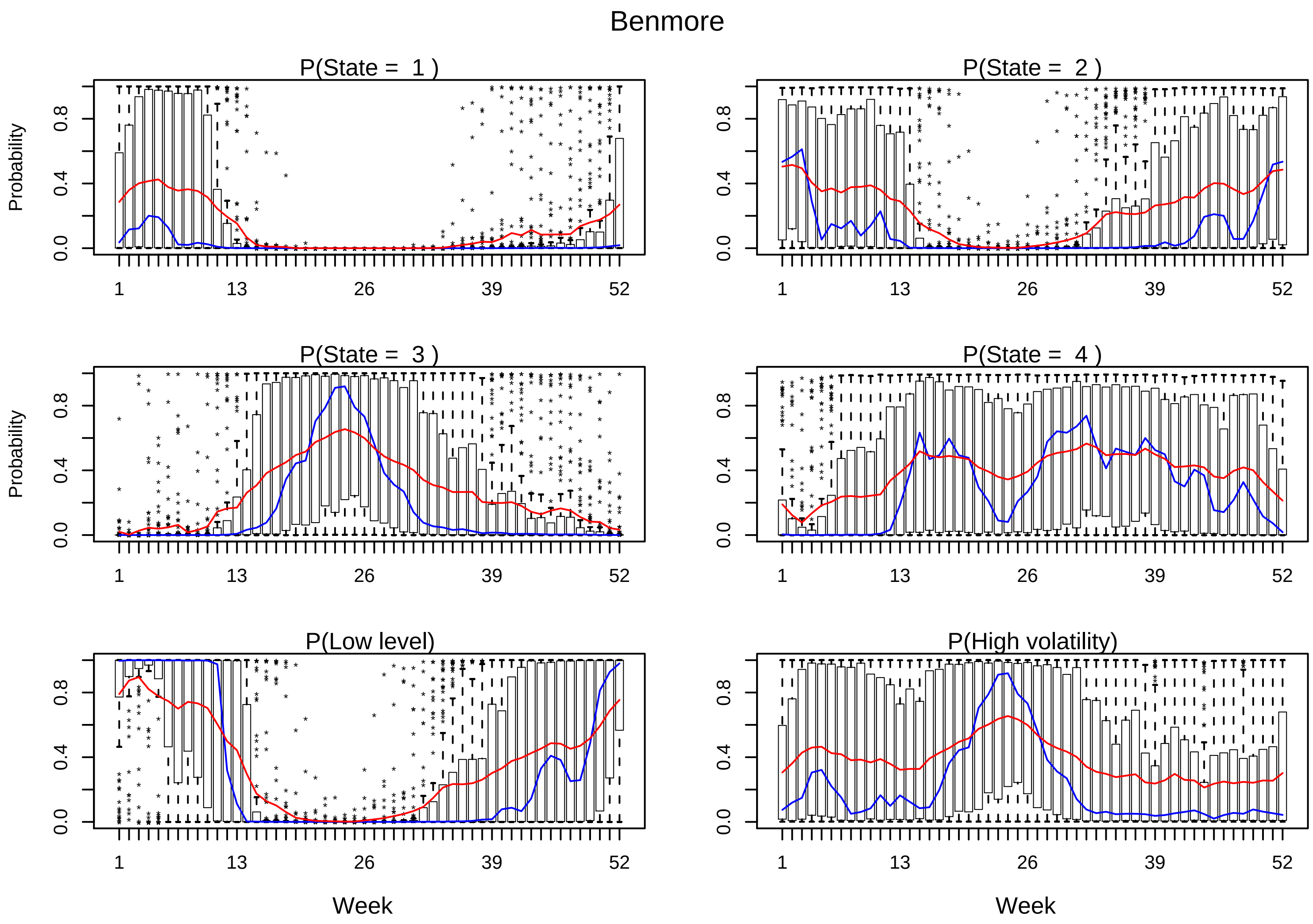
Seasonal mean



Benmore transformed inflows

Boxplots of classification probabilities by week of the year

Seasonal mean



Best predictors of the number of transitions from state j to state k and number of weeks in state j are given by

$$E(\sum_{t=1}^{T-1} I(S_t = j, S_{t+1} = k)|\mathbf{Y}) = \sum_{t=1}^{T-1} \gamma_t(j,k)$$
$$E(\sum_{t=1}^{T} I(S_t = j)|\mathbf{Y}) = \sum_{t=1}^{T} \gamma_t(j)$$

where $j, k = 1, \ldots, 4$ and $I(.)$ is the indicator function.

Using these estimated counts, simple moment estimators of the transition probabilities are given by

$$\tilde{P}_{jk} = \frac{\sum_{t=1}^{T-1} \gamma_t(j,k)}{\sum_{t=1}^{T} \gamma_t(j)} \quad (j, k = 1, \ldots, 4)$$

If the classification probabilities are reliable, $\tilde{P}_{jk}$ provides more robust estimates of the transition probabilities than ML.

	Arapuni				Benmore			
	$S_{t+1}=1$	$S_{t+1}=2$	$S_{t+1}=3$	$S_{t+1}=4$	$S_{t+1}=1$	$S_{t+1}=2$	$S_{t+1}=3$	$S_{t+1}=4$
$S_t=1$	0.84	0.10	0.06	0.01	0.79	0.13	0.05	0.03
$S_t=2$	0.08	0.84	0.01	0.07	0.10	0.77	0.02	0.11
$S_t=3$	0.04	0.01	0.79	0.15	0.00	0.01	0.80	0.19
$S_t=4$	0.01	0.10	0.07	0.82	0.00	0.09	0.11	0.80
$S_t=1$	0.81	0.12	0.06	0.01	0.73	0.16	0.09	0.02
$S_t=2$	0.08	0.85	0.01	0.07	0.10	0.79	0.01	0.10
$S_t=3$	0.08	0.01	0.79	0.12	0.05	0.01	0.77	0.17
$S_t=4$	0.01	0.08	0.08	0.83	0.01	0.05	0.11	0.83

Simple moment estimates $\tilde{P}_{jk}$ (top panel) and ML model-based estimates $\widehat{P}_{jk}$(bottom panel) of the transition probabilities.

Since $P(C_t = 0|\mathbf{Y})$ shows the strongest seasonality, consider onsets and durations of low and high flow regimes.

Classify inflow as low regime if $P(C_t = 0|\mathbf{Y}) > 0.5$ and high regime otherwise.

Apply a non-linear filter (censoring) that ignores short (implausible) durations.

Summary statistics for low and high flow regime classifications.

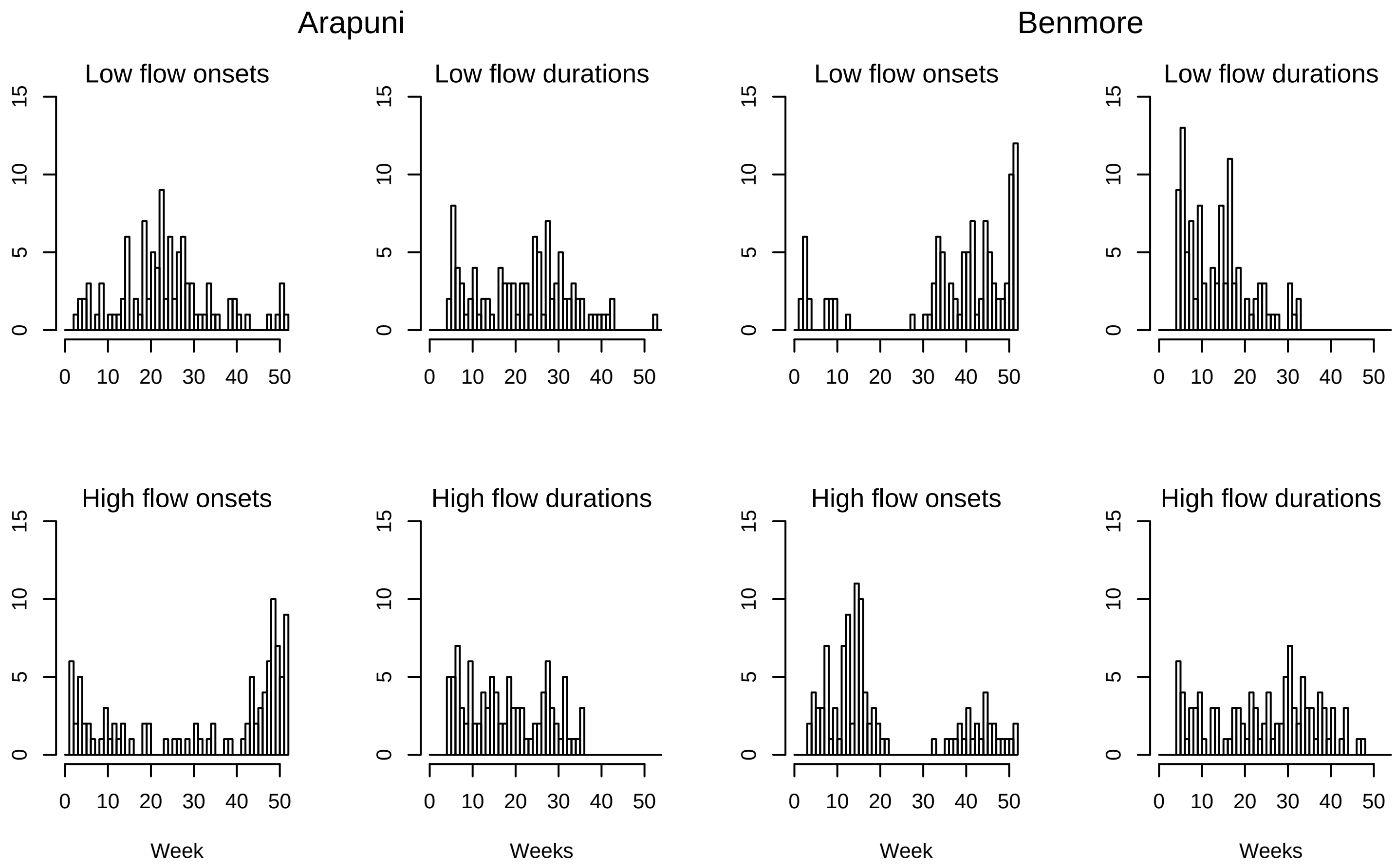
	Arapuni		Benmore	
	Low flows	High flows	Low flows	High flows
Number of regimes	139	139	141	142
Mean duration	14.90	12.55	9.50	17.48
Onset mode	22	48.5	50	14
Number of regimes	99	99	103	104
Mean duration	21.24	17.30	13.11	23.77
Onset mode	22	48	50	14
Number of regimes	75	75	74	75
Mean duration	28.16	22.72	17.12	34.07
Onset mode	22	48	50	14

Top panel applies to original classifications. Remaining panels to filtered regime classifications with

- durations of 3 or less months censored (middle panel);
- 6 or less months censored (bottom panel).

Transformed inflows

Histograms of onsets and durations of low and high flow regimes
($\leq$ 3 month durations censored)



Summary of findings

BHT model provides reasonably secure classification probabilities that give a better understanding of the seasonal dynamics.

In particular, the analysis shows

- strong empirical evidence for episodic seasonal regimes with varying onset and end times;
- that, in general, S_t tends to move mainly between adjacent states in the mean flow hierarchy, with the intermediate states used for (asymmetric) rising and falling flows;
- that, in terms of the BHT model, C_t needs to be seasonal, but V_t need not.

5. Seasonal HMMs for inflows

Consider a non-homogeneous generalisation of the BHT model given by

$$P(C_{t+1} = j | C_t = i) = p_{ij}(\tau) \quad (i, j = 0, 1)$$

where

$$p_{00}(\tau) = p_{00}^{(k)}, \quad p_{11}(\tau) = p_{11}^{(k)} \quad (\tau \in E_k;\ k = 1, \ldots, 4)$$

with τ denoting the week of the year for t.

The E_k are mutually exclusive time-of-year intervals with

$E_1 =$ only low flows $\quad E_2 =$ transition period

$E_3 =$ only high flows $\quad E_4 =$ transition period.

Now introduce absorbing states by requiring

$$\mathbf{P}^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \ \mathbf{P}^{(2)} = \begin{bmatrix} p_{00}^{(2)} & 1 - p_{00}^{(2)} \\ 0 & 1 \end{bmatrix}, \ \mathbf{P}^{(3)} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \ \mathbf{P}^{(4)} = \begin{bmatrix} 1 & 0 \\ p_{00}^{(4)} & 1 - p_{00}^{(4)} \end{bmatrix}$$

This ensures annual seasonality with just 2 free parameters.

The transition probabilities of V_t are defined by

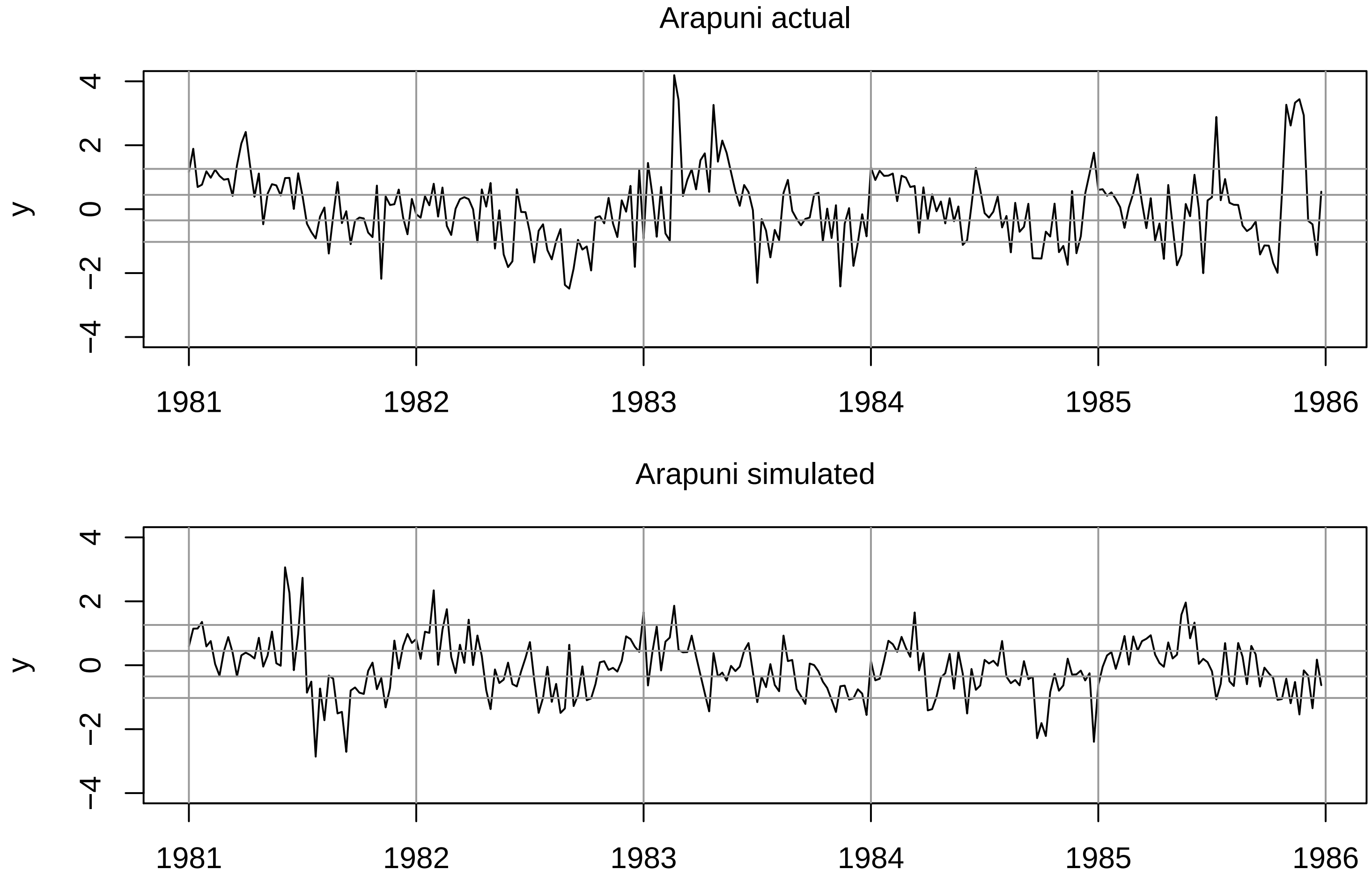
$$P(V_{t+1} = j | V_t = i) = \begin{cases} q_{ij}^{(0)} & (C_t = 0) \\ q_{ij}^{(1)} & (C_t = 1) \end{cases} \qquad (i, j = 0, 1)$$

This is a structural, seasonal HMM which is simple, physically interpretable, parsimonious and consistent with our findings.

A general seasonal HMM can be constructed similarly.

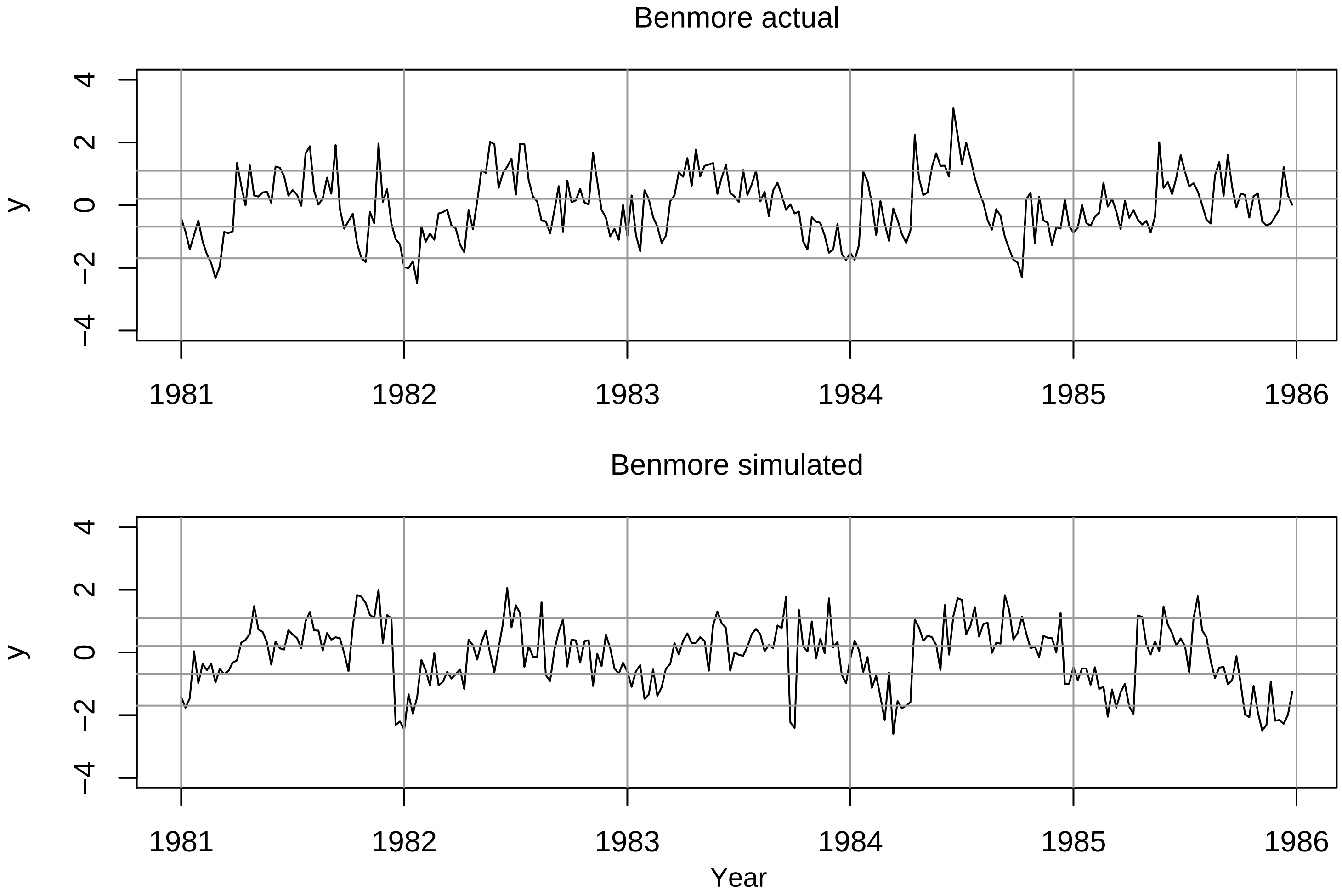
Arapuni transformed inflows (5 year sample)

Actual and simulated (illustrative only)



Benmore transformed inflows (5 year sample)

Actual and simulated (illustrative only)



Further work

These models need to be fitted to the data and their performance benchmarked against more conventional models (eg PARMA).

This will involve the development of suitable

- estimation procedures and algorithms;
- R programs and code;
- testing procedures.

Also need to devise ways of including the inter-annual trend.

For further details see:
http://www.electricitycommission.govt.nz/opdev/modelling/hydrology/index.html