Bayesian Minimal Description Lengths for Multiple Changepoint Detection

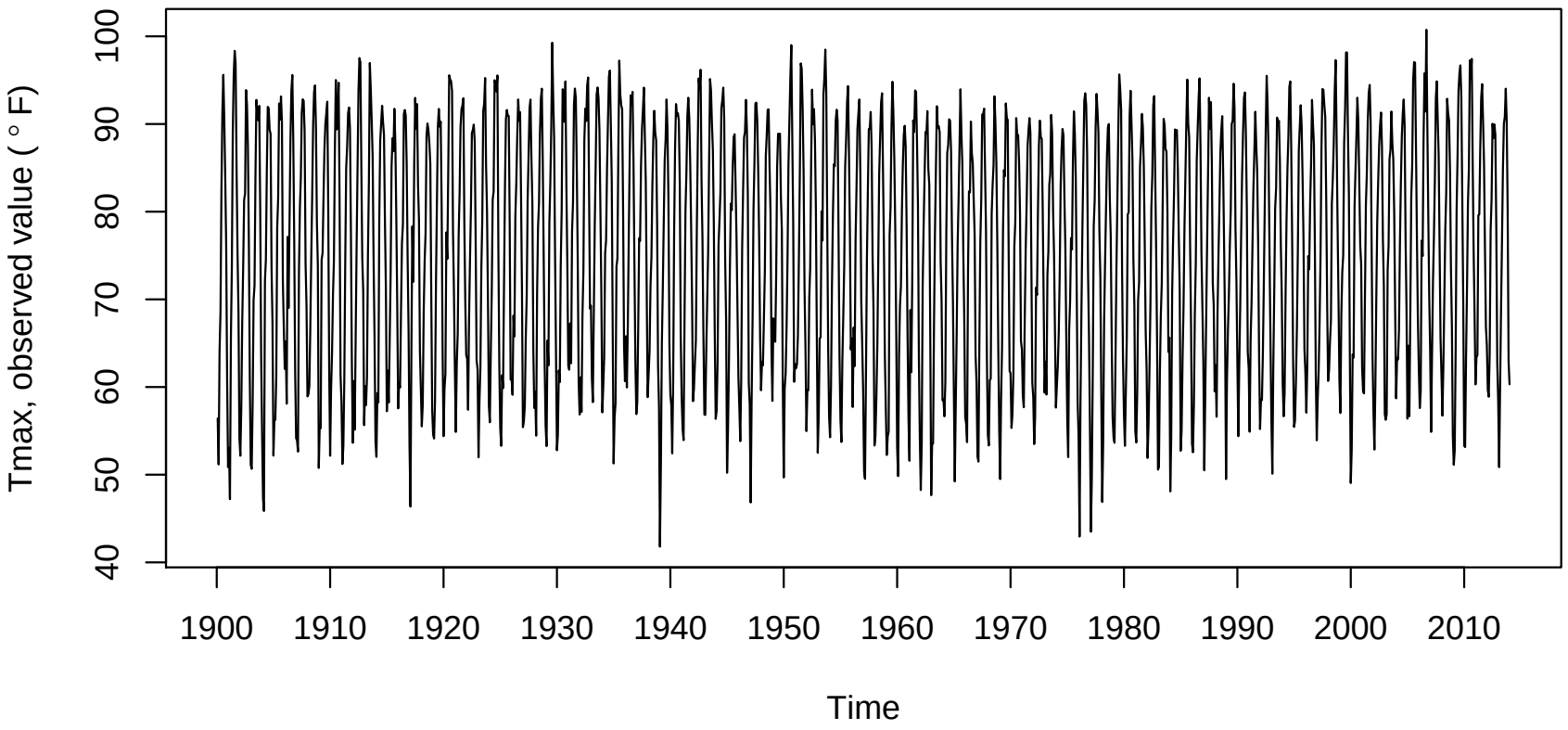
Yingbo Li

Dept of Mathematical Sciences, Clemson University

Co-authors: Robert Lund (Clemson University),
Anuradha Hewaarachchi (University of Kelaniya)

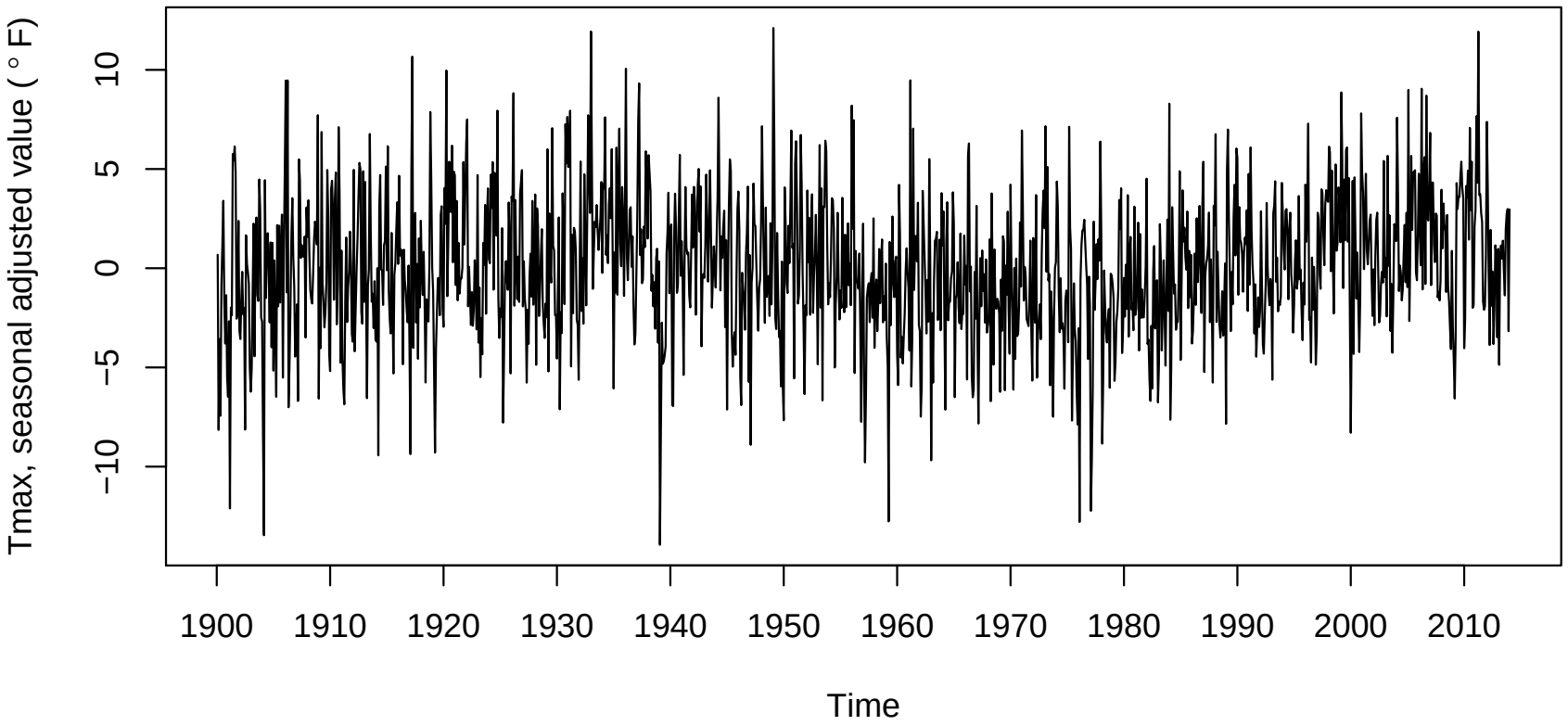
Monthly maximum temperature series in Tuscaloosa, AL

Observed data



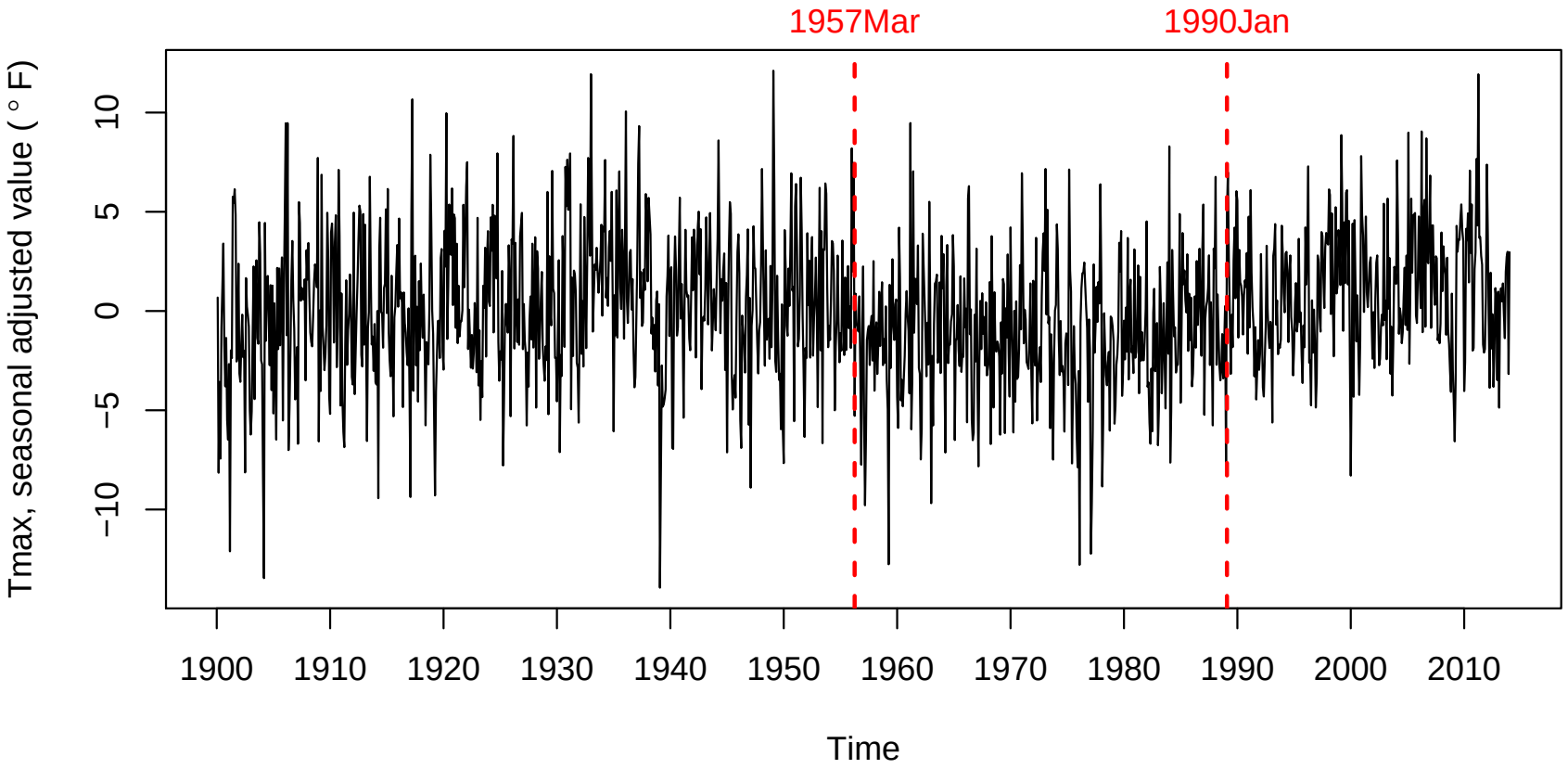
Monthly maximum temperature series in Tuscaloosa, AL

Observed data − sample seasonal mean



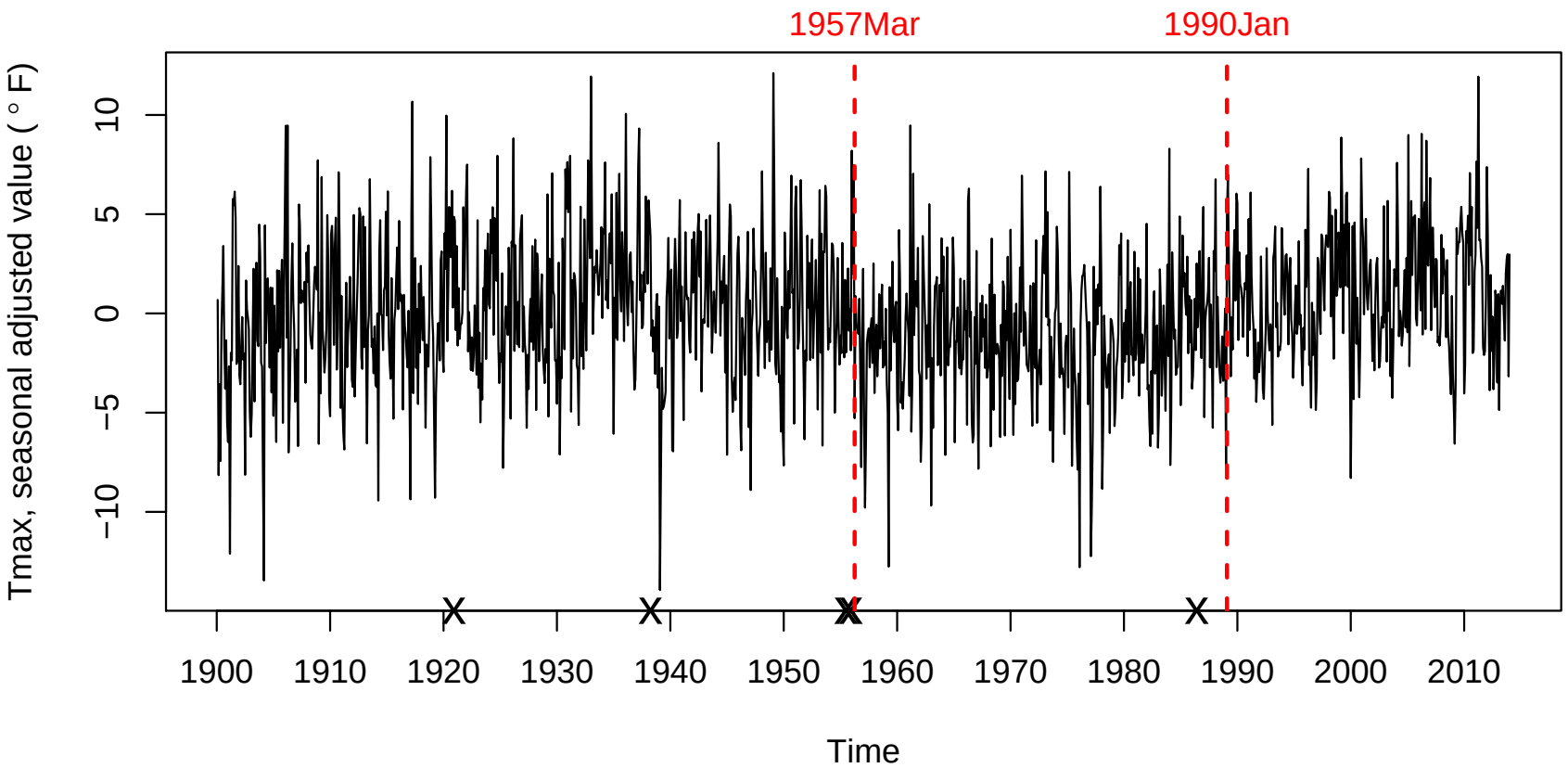
Monthly maximum temperature series in Tuscaloosa, AL

Observed data − sample seasonal mean



Monthly maximum temperature series in Tuscaloosa, AL

Observed data – sample seasonal mean

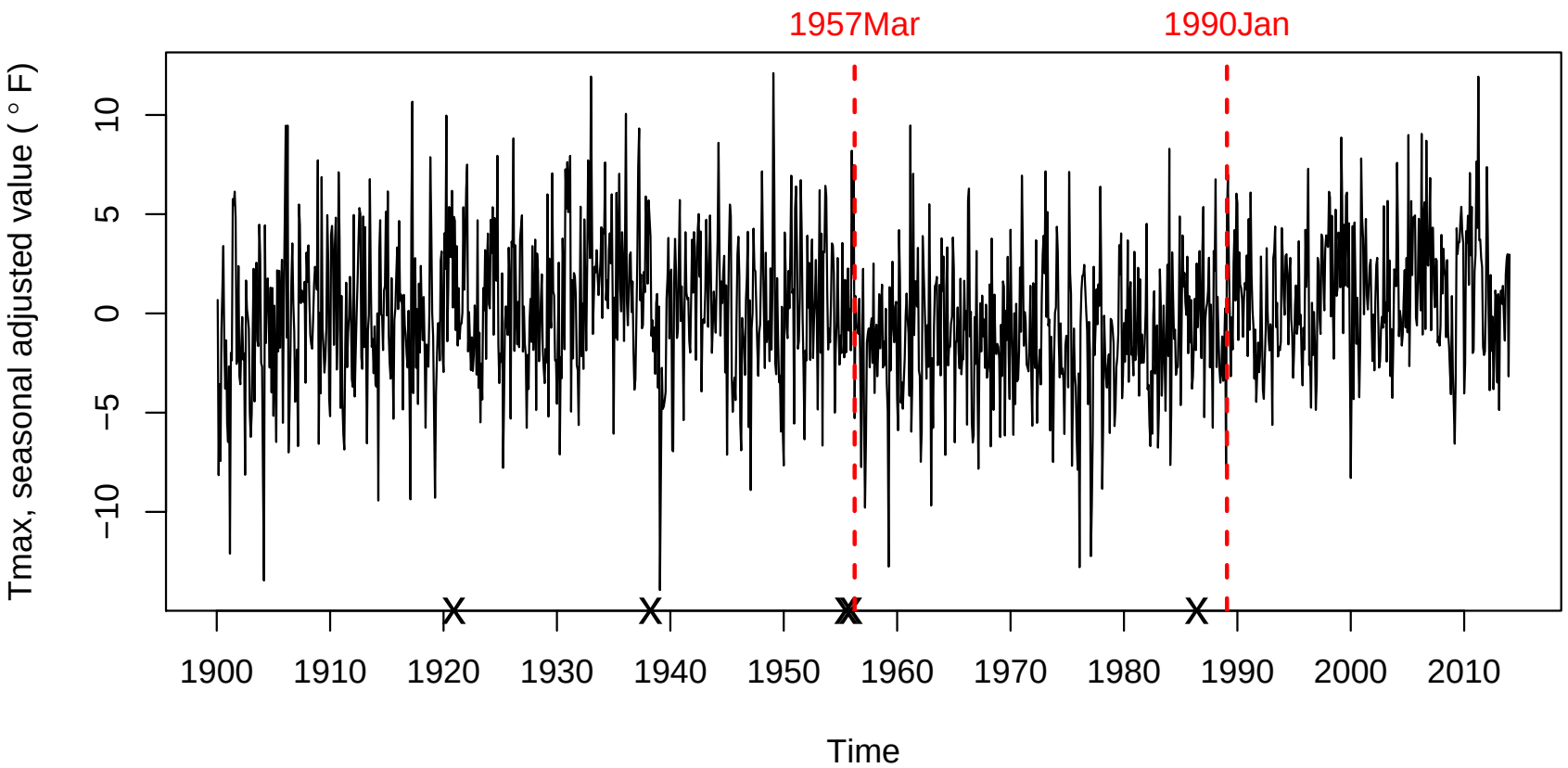


Metadata (station history logs)

- Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
- Instrumentation changes: 1956 Nov, 1987 May

Monthly maximum temperature series in Tuscaloosa, AL

Observed data − sample seasonal mean

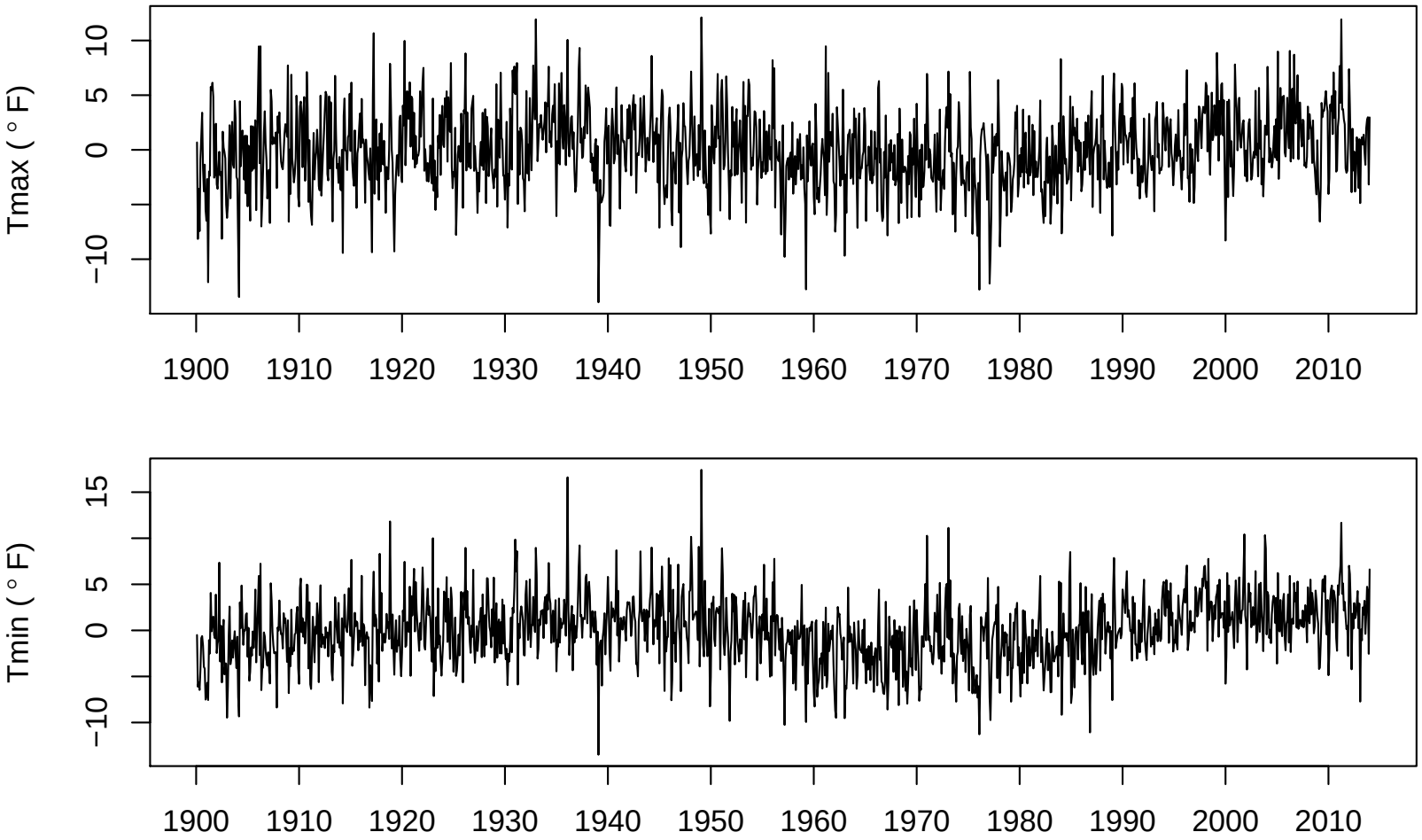


Metadata (station history logs): more likely to induce mean shifts

- Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
- Instrumentation changes: 1956 Nov, 1987 May

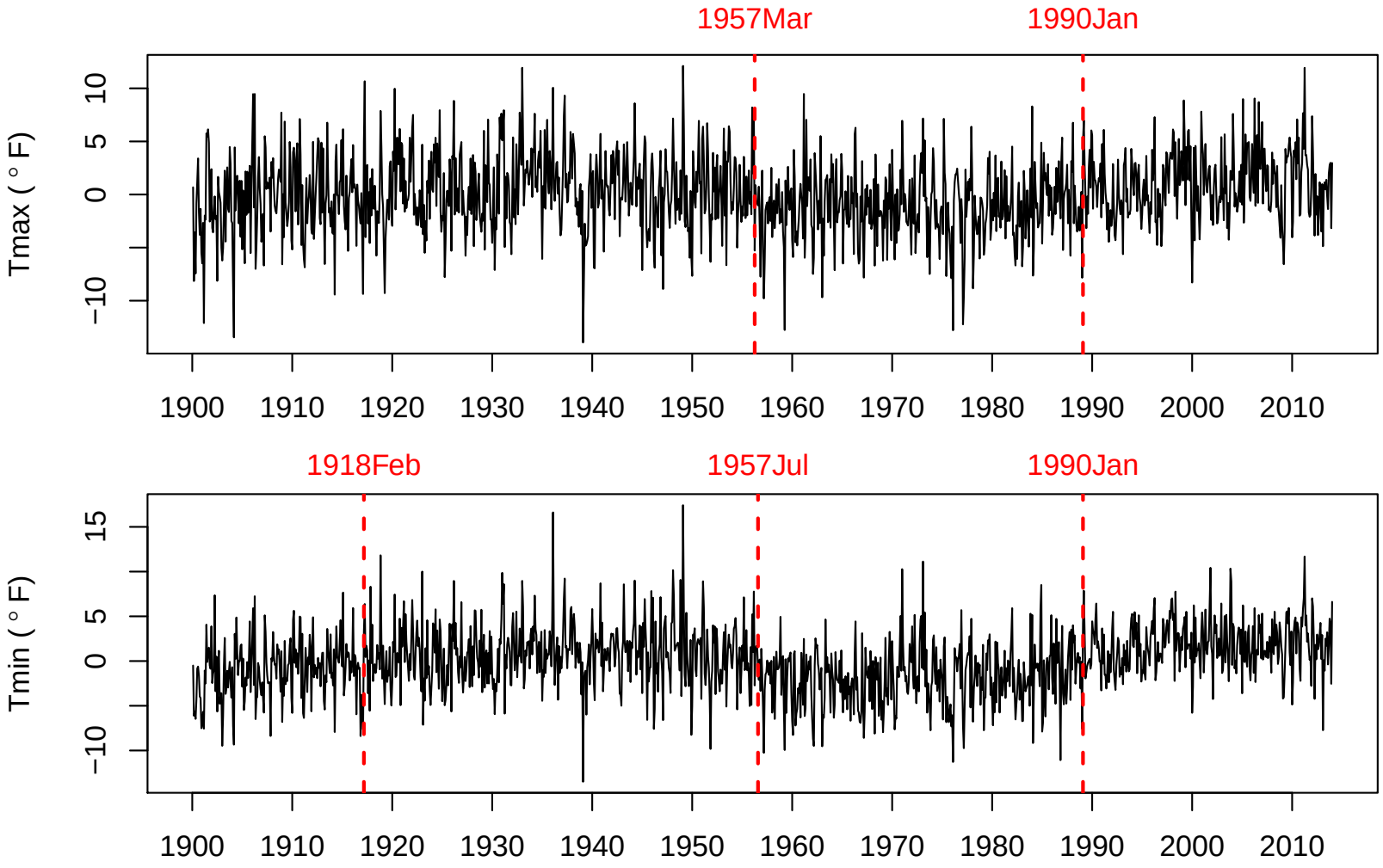
Monthly maximum and minimum temperature series

Observed data − sample seasonal mean



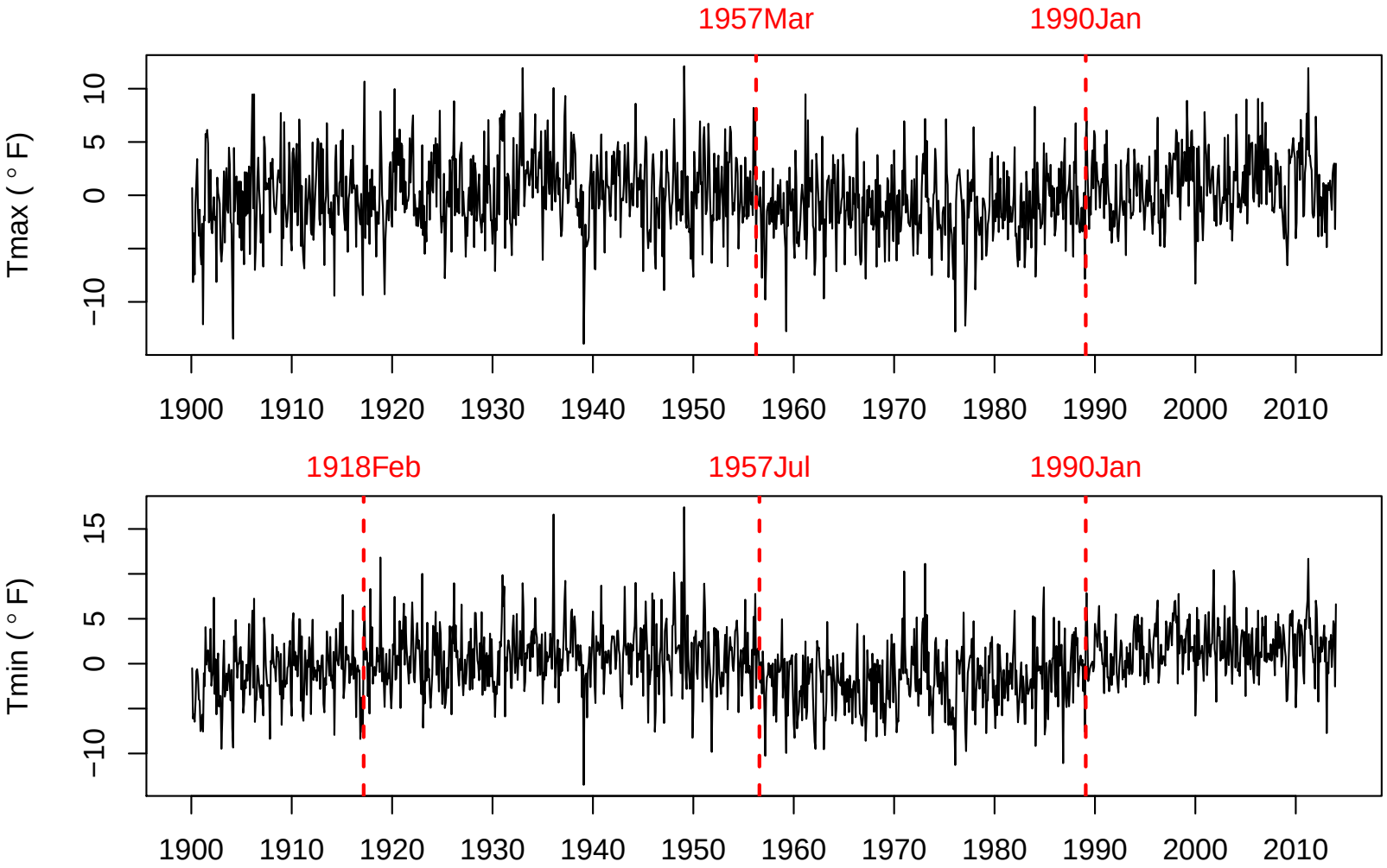
Monthly maximum and minimum temperature series

Observed data − sample seasonal mean



Monthly maximum and minimum temperature series

Observed data − sample seasonal mean



Tmax and Tmin are likely to shift at the same time.

MDL for Changepiont Detection in Piecewise AR Series

- Automatic MDL by Davis et al. [2006]: a penalized likelihood, with penalty being the code length (CL) of parameters

$$\underbrace{\log(m)}_{m} + \underbrace{(m+1)\log(N)}_{\text{regime lengths}} + \underbrace{\sum_{r=1}^{m+1} \log p_r}_{\text{AR orders}} + \underbrace{\sum_{r=1}^{m+1} \frac{p_r+2}{2} \log N_r}_{\text{AR coefficients}}$$

- Automatic MDL rules:
  - CL of an unbounded positive integer I: $\log(I)$
  - CL of a positive integer bounded above by U: $\log(U)$
  - CL of the MLE of a real-valued parameter estimated by N observations: $\frac{1}{2}\log N$

Model Selection Using MDL Principle

- Description length [Risanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.

Model Selection Using MDL Principle

- Description length [Risanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

$$\mathcal{L}(\mathbf{X}, \theta) = \underbrace{\mathcal{L}(\mathbf{X} \mid \theta)}_{\text{transmit } \mathbf{X}} + \underbrace{\mathcal{L}(\theta)}_{\text{transmit } \theta}$$

Model Selection Using MDL Principle

- Description length [Risanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

$$\mathcal{L}(\mathbf{X}, \theta) = \underbrace{-\log f(\mathbf{X} \mid \theta)}_{\text{transmit } \mathbf{X}} - \underbrace{\log \pi(\theta)}_{\text{transmit } \theta}$$

Model Selection Using MDL Principle

- Description length [Risanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

$$\mathcal{L}(\mathbf{X}, \theta) = \underbrace{-\log f(\mathbf{X} \mid \theta)}_{\text{transmit } \mathbf{X}} - \underbrace{\log \pi(\theta)}_{\text{transmit } \theta}$$

- Mixture MDL

$$\mathcal{L}(\mathbf{X}) = -\log \underbrace{\int f(\mathbf{X} \mid \theta)\pi(\theta)d\theta}_{\text{marginal likelihood}}$$

Model Selection Using MDL Principle

- Description length [Risanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

$$\mathcal{L}(\mathbf{X}, \theta) = \underbrace{-\log f(\mathbf{X} \mid \theta)}_{\text{transmit } \mathbf{X}} - \underbrace{\log \pi(\theta)}_{\text{transmit } \theta}$$

- Mixture MDL

$$\mathcal{L}(\mathbf{X}) = -\log \underbrace{\int f(\mathbf{X} \mid \theta)\pi(\theta)d\theta}_{\text{marginal likelihood}}$$

  - If the prior $\pi(\theta \mid \tau)$, combine with two-part MDL:

$$\mathcal{L}(\mathbf{X}, \hat{\tau}) = -\log \int f(\mathbf{X} \mid \theta)\pi(\theta \mid \hat{\tau})d\theta - \log \pi(\hat{\tau})$$

    Closely related with empirical Bayes marginal likelihood.

Bayesian MDL

- Observed time series: $\mathbf{X}_{1:N} = (X_1, X_2, \ldots, X_N)'$
- Suppose m changepoints $1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N$ partition the timeline $m+1$ distinct regimes (segments).

$$\text{Time } t \text{ is in regime } r \iff \tau_{r-1} \leq t < \tau_r$$

Bayesian MDL

- Observed time series: $\mathbf{X}_{1:N} = (X_1, X_2, \ldots, X_N)'$
- Suppose m changepoints $1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N$ partition the timeline $m+1$ distinct regimes (segments).

$$\text{Time } t \text{ is in regime } r \iff \tau_{r-1} \leq t < \tau_r$$

- Any time in $t = \{p+1, p+2, \ldots, N\}$ can be a changepoint

Definition

Denote a *multiple changepoint configuration* as a indicator vector

$$\boldsymbol{\eta} = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)', \quad \text{such that}$$

$$\eta_t = \begin{cases} 1, & \text{if time } t \text{ is a changepoint} \\ 0, & \text{if time } t \text{ is not a changepoint} \end{cases}$$

Bayesian MDL

- Observed time series: $\mathbf{X}_{1:N} = (X_1, X_2, \ldots, X_N)'$
- Suppose m changepoints $1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N$ partition the timeline $m+1$ distinct regimes (segments).

$$\text{Time } t \text{ is in regime } r \iff \tau_{r-1} \leq t < \tau_r$$

- Any time in $t = \{p+1, p+2, \ldots, N\}$ can be a changepoint

Definition

Denote a *multiple changepoint configuration* as a indicator vector

$$\boldsymbol{\eta} = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)', \quad \text{such that}$$

$$\eta_t = \begin{cases} 1, & \text{if time } t \text{ is a changepoint} \\ 0, & \text{if time } t \text{ is not a changepoint} \end{cases}$$

- Number of changepoints in $\boldsymbol{\eta}$: $m = \sum_{t=p+1}^{N} \eta_t$
- Total number of models: 2^{N-p}

Prior distribution $\pi(\boldsymbol{\eta})$: Beta-Binomial

- If time t is not in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(1)}\right), \quad \rho^{(1)} \sim \text{Beta}\left(a, b^{(1)}\right)$$

- If time t is in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(2)}\right), \quad \rho^{(2)} \sim \text{Beta}\left(a, b^{(2)}\right)$$

Prior distribution $\pi(\boldsymbol{\eta})$: Beta-Binomial

- If time t is not in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(1)}\right), \quad \rho^{(1)} \sim \text{Beta}\left(a, b^{(1)}\right)$$

- If time t is in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(2)}\right), \quad \rho^{(2)} \sim \text{Beta}\left(a, b^{(2)}\right)$$

- $\pi(\boldsymbol{\eta})$ has a closed form:

$$\pi(\boldsymbol{\eta}) = \prod_{k=1}^{2} \int_0^1 \left[\prod_{t^{(k)}} \pi\left(\eta_{t^{(k)}} \mid \rho^{(k)}\right) \right] \pi\left(\rho^{(k)}\right) d\rho^{(k)}$$

Prior distribution $\pi(\boldsymbol{\eta})$: Beta-Binomial

- If time t is not in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(1)}\right), \quad \rho^{(1)} \sim \text{Beta}\left(a, b^{(1)}\right)$$

- If time t is in metadata

$$\eta_t \overset{iid}{\sim} \text{Bernoulli}\left(\rho^{(2)}\right), \quad \rho^{(2)} \sim \text{Beta}\left(a, b^{(2)}\right)$$

- $\pi(\boldsymbol{\eta})$ has a closed form:

$$\pi(\boldsymbol{\eta}) \propto \prod_{k=1}^{2} \Gamma\left(a + m^{(k)}\right) \Gamma\left(b^{(k)} + N^{(k)} - m^{(k)}\right)$$

Number of:

- undocumented times $N^{(1)}$, undocumented changepoints $m^{(1)}$
- documented times $N^{(2)}$, documented changepoints $m^{(2)}$

Parameter elicitation

- Metadata times are more likely to be changepoints

$$E\left(\rho^{(1)}\right) = \frac{a}{a+b^{(1)}} < \frac{a}{a+b^{(2)}} = E\left(\rho^{(2)}\right)$$

Parameter elicitation

- Metadata times are more likely to be changepoints

$$E\left(\rho^{(1)}\right) = \frac{a}{a+b^{(1)}} < \frac{a}{a+b^{(2)}} = E\left(\rho^{(2)}\right)$$

- US temperature: 6 changepoints per century [Mitchell, 1953],

$$0.005 \text{ changepoint / month}$$

Default parameters:

$$a = 1,\ b^{(1)} = 239 \Longrightarrow E\left(\rho^{(1)}\right) = 0.004$$
$$b^{(2)} = 47 \Longrightarrow E\left(\rho^{(2)}\right) = 0.021$$

Likelihood: multivariate normal

Under a specific model $\boldsymbol{\eta}$,

$$X_t = s_{v(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$

$v(t)$ time t is in season $v(t)$

$r(t)$ time t is in regime $r(t)$

s_v seasonal mean, $v = 1, 2, \ldots, 12$.

μ_r regime mean, $r = 1, 2, \ldots, m+1$

$\{\epsilon_t\}$ Gaussian AR(p) errors, with white noise variance σ^2 and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$

Likelihood: multivariate normal

Under a specific model $\boldsymbol{\eta}$,

$$X_t = s_{v(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$

$v(t)$ time t is in season $v(t)$

$r(t)$ time t is in regime $r(t)$

s_v seasonal mean, $v = 1, 2, \ldots, 12$.

μ_r regime mean, $r = 1, 2, \ldots, m+1$

$\{\epsilon_t\}$ Gaussian AR(p) errors, with white noise variance σ^2 and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
- Likelihood function: $f\left(\mathbf{X}_{(p+1):N} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\eta}, \mathbf{X}_{1:p}\right)$

Likelihood: multivariate normal

Under a specific model $\boldsymbol{\eta}$,

$$X_t = s_{v(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$

$v(t)$ time t is in season $v(t)$

$r(t)$ time t is in regime $r(t)$

s_v seasonal mean, $v = 1, 2, \ldots, 12$.

μ_r regime mean, $r = 1, 2, \ldots, m+1$

$\{\epsilon_t\}$ Gaussian AR(p) errors, with white noise variance σ^2 and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
- Likelihood function: $f\left(\mathbf{X}_{(p+1):N} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\eta}, \mathbf{X}_{1:p}\right)$

Likelihood: multivariate normal

Under a specific model $\boldsymbol{\eta}$,

$$X_t = s_{v(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$

$v(t)$ time t is in season $v(t)$

$r(t)$ time t is in regime $r(t)$

s_v seasonal mean, $v = 1, 2, \ldots, 12$.

μ_r regime mean, $r = 1, 2, \ldots, m+1$

$\{\epsilon_t\}$ Gaussian AR(p) errors, with white noise variance σ^2 and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
- Likelihood function: $f\left(\mathbf{X}_{(p+1):N} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\eta}, \mathbf{X}_{1:p}\right)$
- Using mixture MDL on $\boldsymbol{\mu}$, and two-part MDL on the rest.

BMDL: has a closed form

$$\mathrm{BMDL}(\boldsymbol{\eta}) = \underbrace{\mathcal{L}(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\boldsymbol{\phi}}, \boldsymbol{\eta})}_{\text{mixture MDL}} + \underbrace{\mathcal{L}(\hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\boldsymbol{\phi}} \mid \boldsymbol{\eta})}_{\text{two-part MDL}} + \underbrace{\mathcal{L}(\boldsymbol{\eta})}_{\text{two-part MDL}}$$

BMDL: has a closed form

$$\text{BMDL}(\boldsymbol{\eta}) = \underbrace{\mathcal{L}(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\boldsymbol{\phi}}, \boldsymbol{\eta})}_{\text{mixture MDL}} + \underbrace{\mathcal{L}(\hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\boldsymbol{\phi}} \mid \boldsymbol{\eta})}_{\text{two-part MDL}} + \underbrace{\mathcal{L}(\boldsymbol{\eta})}_{\text{two-part MDL}}$$

- Prior distribution on $\boldsymbol{\mu} = (\mu_2, \mu_3, \ldots, \mu_{m+1})'$

$$\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta} \sim \text{N}(\mathbf{0}, \nu\sigma^2 \mathbf{I}_m).$$

  - ν is pre-specified; default $\nu = 5$.
  - Normal-normal conjugacy: marginal likelihood has a closed form

$$f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\eta}) = \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta}) d\boldsymbol{\mu}$$

BMDL: has a closed form

$$\text{BMDL}(\boldsymbol{\eta}) = \underbrace{\mathcal{L}(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta})}_{\text{mixture MDL}} + \underbrace{\mathcal{L}(\hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi} \mid \boldsymbol{\eta})}_{\text{two-part MDL}} + \underbrace{\mathcal{L}(\boldsymbol{\eta})}_{\text{two-part MDL}}$$

- Prior distribution on $\boldsymbol{\mu} = (\mu_2, \mu_3, \ldots, \mu_{m+1})'$

$$\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta} \sim \text{N}(\mathbf{0}, \nu\sigma^2 \mathbf{I}_m).$$

  - ν is pre-specified; default $\nu = 5$.
  - Normal-normal conjugacy: marginal likelihood has a closed form

$$f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}) = \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta}) d\boldsymbol{\mu}$$

- $\hat{\mathbf{s}}, \hat{\sigma}^2 = \arg\max f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta})$ have closed forms.
- $\hat{\phi}$: Yule-Walker estimator.

BMDL: has a closed form

$$
\begin{aligned}
\text{BMDL}(\boldsymbol{\eta}) = \mathcal{L}(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}) & \quad + \mathcal{L}(\hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi} \mid \boldsymbol{\eta}) & \quad + \mathcal{L}(\boldsymbol{\eta}) \\
- \log f(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}) & \quad + \tfrac{p+13}{2} \log(N-p) & \quad - \log \pi(\boldsymbol{\eta})
\end{aligned}
$$

(mixture MDL; two-part MDL; two-part MDL)

- Prior distribution on $\boldsymbol{\mu} = (\mu_2, \mu_3, \ldots, \mu_{m+1})'$

$$\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, \nu\sigma^2 \mathbf{I}_m).$$

  - ν is pre-specified; default $\nu = 5$.
  - Normal-normal conjugacy: marginal likelihood has a closed form

$$f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}) = \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta}) d\boldsymbol{\mu}$$

- $\hat{\mathbf{s}}, \hat{\sigma}^2 = \arg\max f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta})$ have closed forms.
- $\hat{\phi}$: Yule-Walker estimator.

BMDL: has a closed form

$$
\begin{aligned}
\text{BMDL}(\boldsymbol{\eta}) = \mathcal{L}(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}) & \quad + \mathcal{L}(\hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi} \mid \boldsymbol{\eta}) & \quad + \mathcal{L}(\boldsymbol{\eta}) \\
- \log f(\mathbf{X} \mid \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}) & \quad + \text{constant} & \quad - \log \pi(\boldsymbol{\eta})
\end{aligned}
$$

mixture MDL | two-part MDL | two-part MDL

- Prior distribution on $\boldsymbol{\mu} = (\mu_2, \mu_3, \ldots, \mu_{m+1})'$

$$\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta} \sim \text{N}(\mathbf{0}, \nu\sigma^2 \mathbf{I}_m).$$

  - ν is pre-specified; default $\nu = 5$.
  - Normal-normal conjugacy: marginal likelihood has a closed form

$$f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}) = \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \sigma^2, \boldsymbol{\eta}) d\boldsymbol{\mu}$$

- $\hat{\mathbf{s}}, \hat{\sigma}^2 = \arg\max f(\mathbf{X} \mid \mathbf{s}, \sigma^2, \phi, \boldsymbol{\eta})$ have closed forms.
- $\hat{\phi}$: Yule-Walker estimator.

Bayesian MDL formula

BMDL (for model $\boldsymbol{\eta}$)

$$BMDL(\boldsymbol{\eta}) = \frac{N-p}{2}\log\left(\hat{\sigma}^2\right) + \frac{m}{2}\log(\nu) + \frac{1}{2}\log\left(\left|\widehat{\mathbf{D}}'\widehat{\mathbf{D}} + \frac{\mathbf{I}_m}{\nu}\right|\right)$$
$$- \sum_{k=1}^{2}\log\left[\Gamma\left(a + m^{(k)}\right)\Gamma\left(b^{(k)} + N^{(k)} - m^{(k)}\right)\right].$$

Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\boldsymbol{\eta}$

$$p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X}) \propto \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \hat{\sigma}^2, \boldsymbol{\eta}) d\boldsymbol{\mu} \cdot \pi(\boldsymbol{\eta})$$

- BMDL is closely related to EB.

$$\mathrm{BMDL}(\boldsymbol{\eta}) = -\log \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \hat{\sigma}^2, \boldsymbol{\eta}) d\boldsymbol{\mu} - \log \pi(\boldsymbol{\eta})$$

Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\boldsymbol{\eta}$

$$p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X}) \propto \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \hat{\sigma}^2, \boldsymbol{\eta}) d\boldsymbol{\mu} \cdot \pi(\boldsymbol{\eta})$$

- BMDL is closely related to EB.

$$\mathrm{BMDL}(\boldsymbol{\eta}) = -\log p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X})$$

Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\boldsymbol{\eta}$

$$p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X}) \propto \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \hat{\sigma}^2, \boldsymbol{\eta}) d\boldsymbol{\mu} \cdot \pi(\boldsymbol{\eta})$$

- BMDL is closely related to EB.

$$\mathrm{BMDL}(\boldsymbol{\eta}) = -\log p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X})$$

- Borrow stochastic model search algorithms from Bayesian model selection literature: Metropolis-Hastings (MCMC).

$$\boldsymbol{\eta}^{[0]} \to \boldsymbol{\eta}^{[1]} \to \boldsymbol{\eta}^{[2]} \to \cdots \to \boldsymbol{\eta}^{[t]} \to \cdots$$

Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\boldsymbol{\eta}$

$$p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X}) \propto \int f\left(\mathbf{X} \mid \boldsymbol{\mu}, \hat{\mathbf{s}}, \hat{\sigma}^2, \hat{\phi}, \boldsymbol{\eta}\right) \pi(\boldsymbol{\mu} \mid \hat{\sigma}^2, \boldsymbol{\eta}) d\boldsymbol{\mu} \cdot \pi(\boldsymbol{\eta})$$

- BMDL is closely related to EB.

$$\mathrm{BMDL}(\boldsymbol{\eta}) = -\log p_{\mathrm{EB}}(\boldsymbol{\eta} \mid \mathbf{X})$$

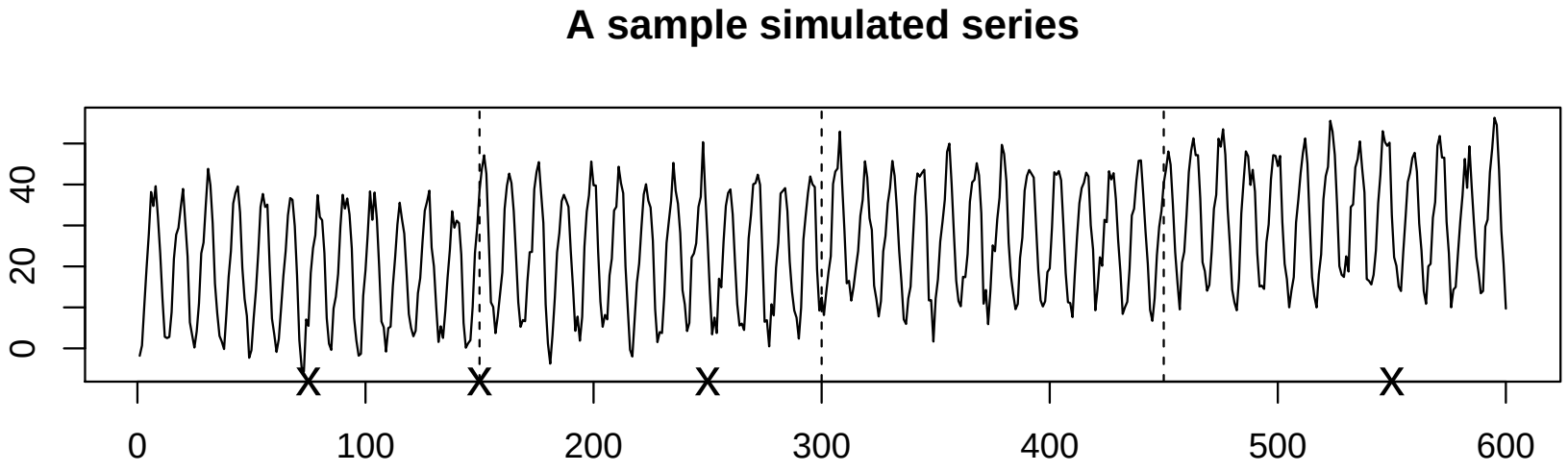
- Borrow stochastic model search algorithms from Bayesian model selection literature: Metropolis-Hastings (MCMC).

$$\boldsymbol{\eta}^{[0]} \to \boldsymbol{\eta}^{[1]} \to \boldsymbol{\eta}^{[2]} \to \cdots \to \boldsymbol{\eta}^{[t]} \to \cdots$$

- Metropolis-Hastings algorithm [George and McCulloch, 1997]
- R package `BayesMDL`.

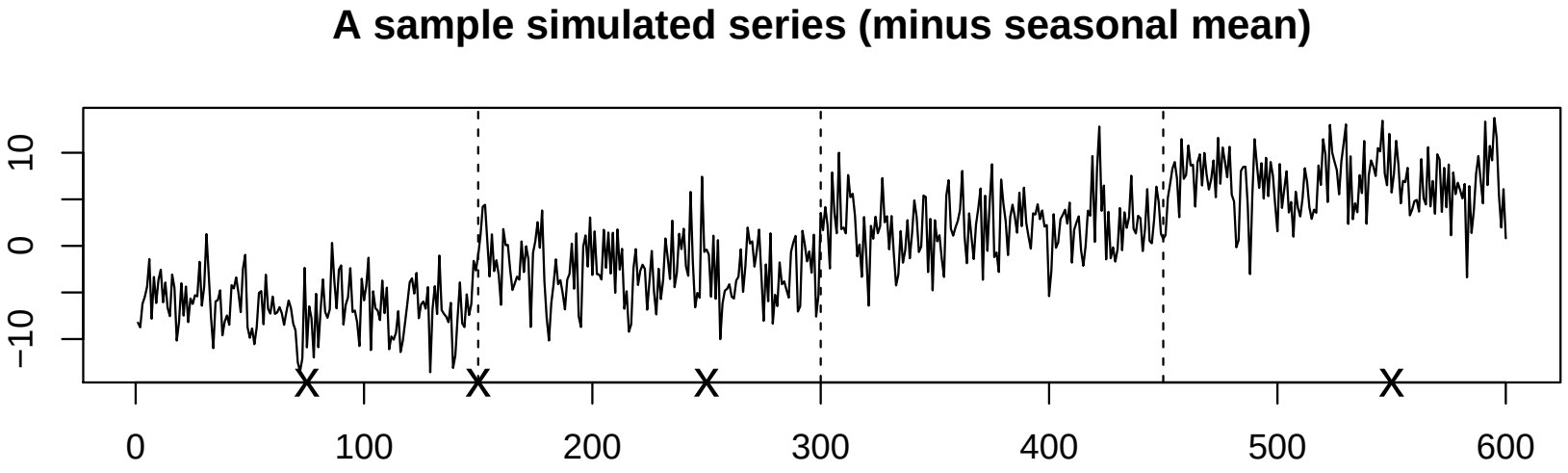
Scenario 1: monotonic shifts $\boldsymbol{\mu} = (0, \Delta, 2\Delta, 3\Delta)'$

$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$



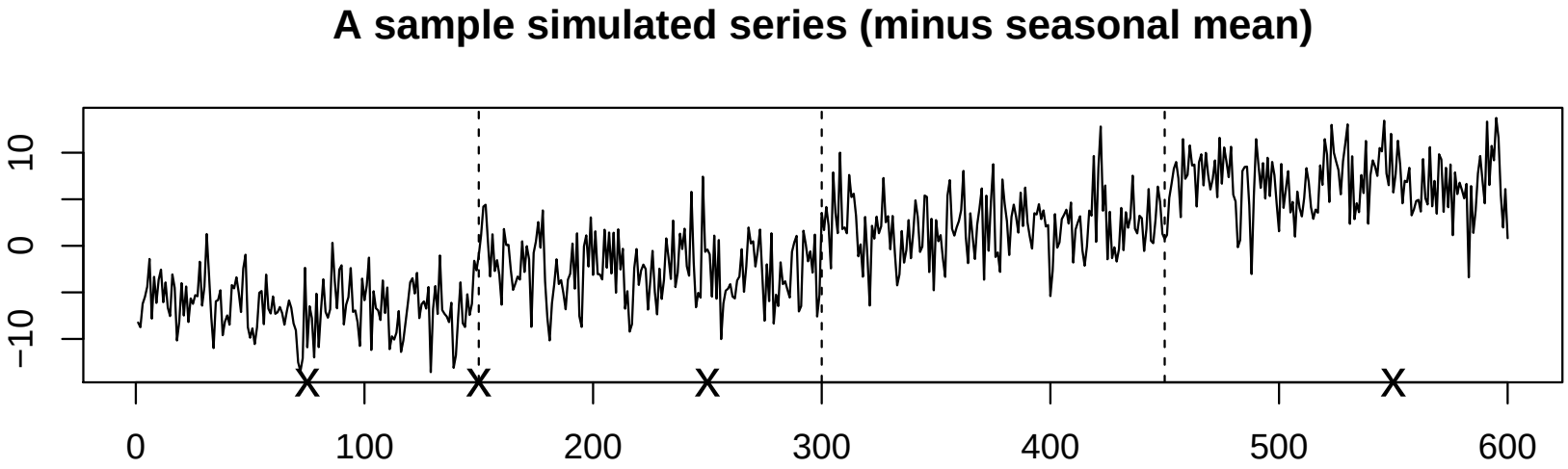
Scenario 1: monotonic shifts $\boldsymbol{\mu} = (0, \Delta, 2\Delta, 3\Delta)'$

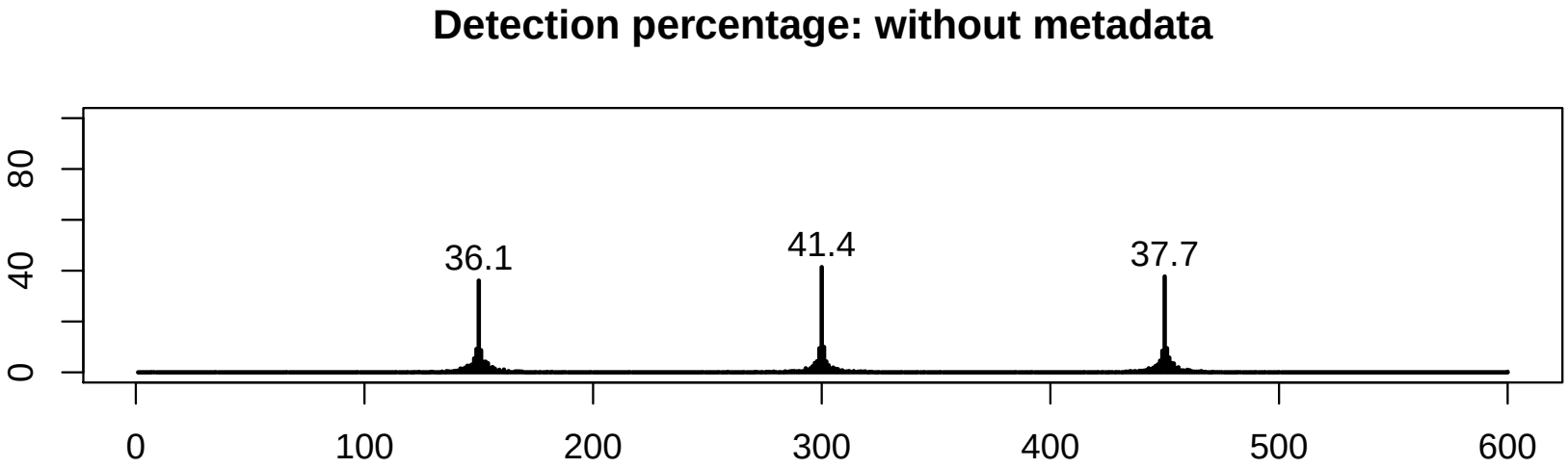
$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$



Scenario 1: monotonic shifts $\boldsymbol{\mu} = (0, \Delta, 2\Delta, 3\Delta)'$

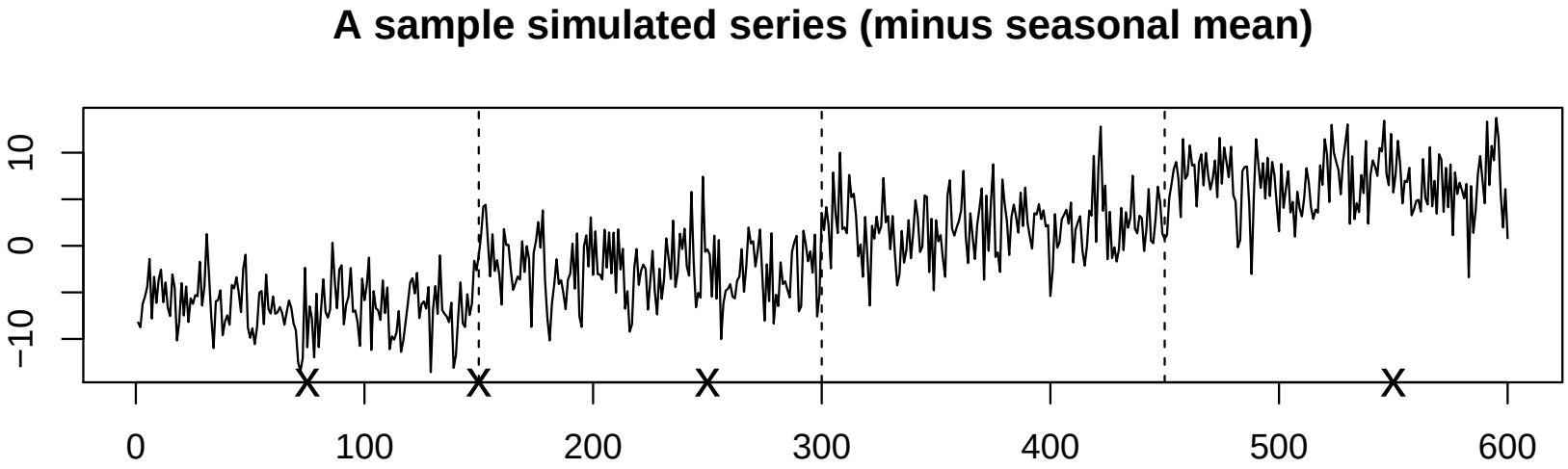
$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

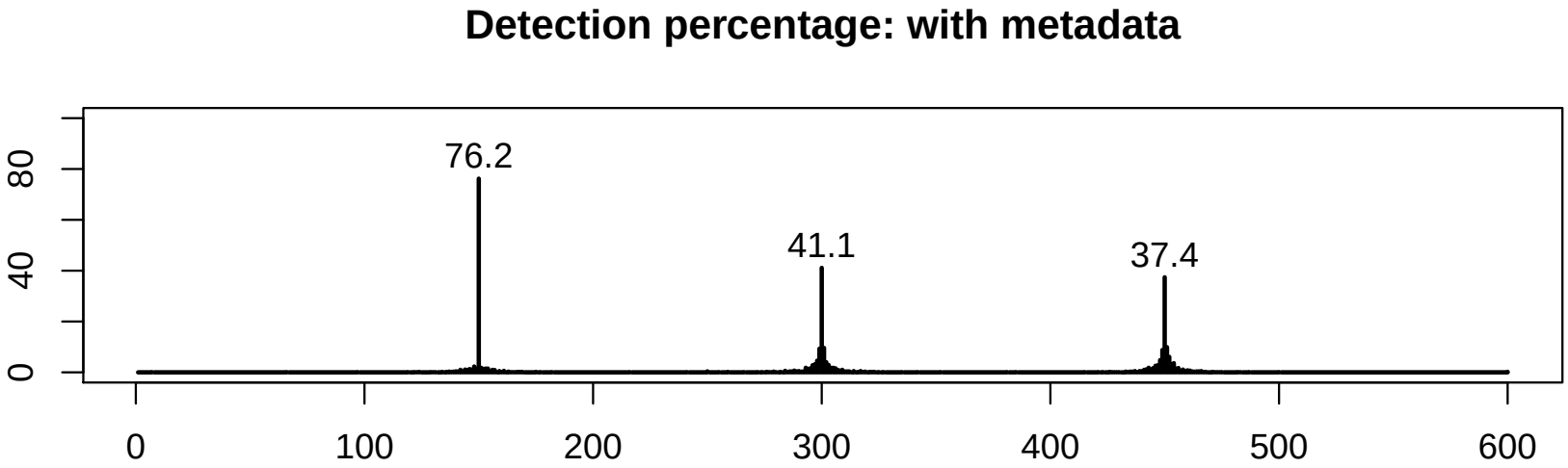




Scenario 1: monotonic shifts $\boldsymbol{\mu} = (0, \Delta, 2\Delta, 3\Delta)'$

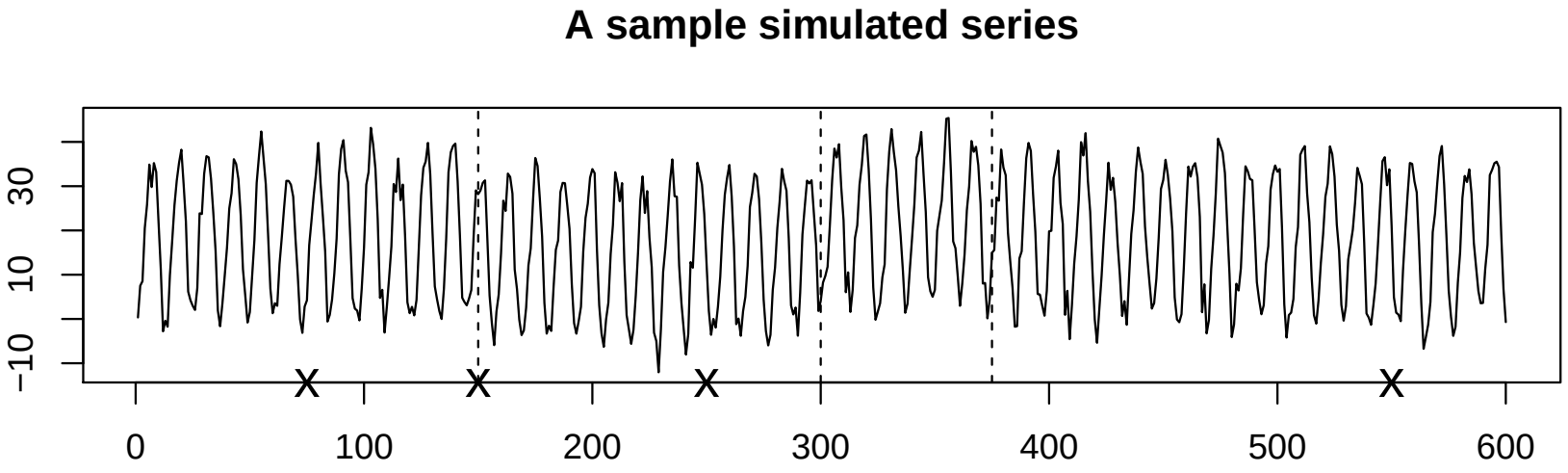
$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$





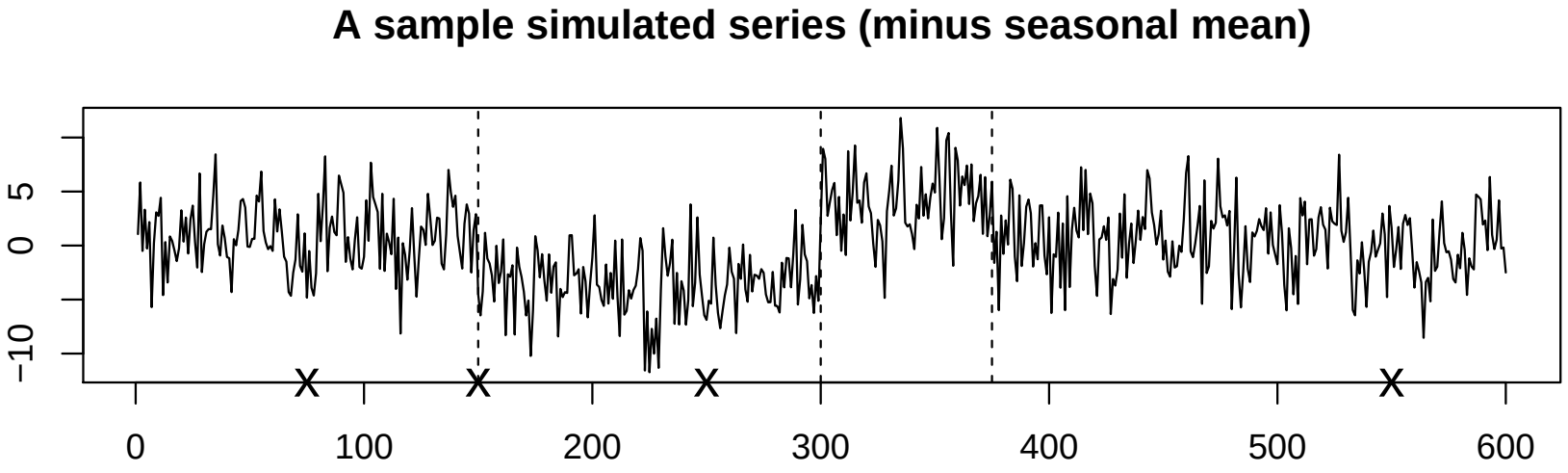
Scenario 2: non-monotonic shifts $\boldsymbol{\mu} = (0, -\Delta, \Delta, 0)'$

$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$



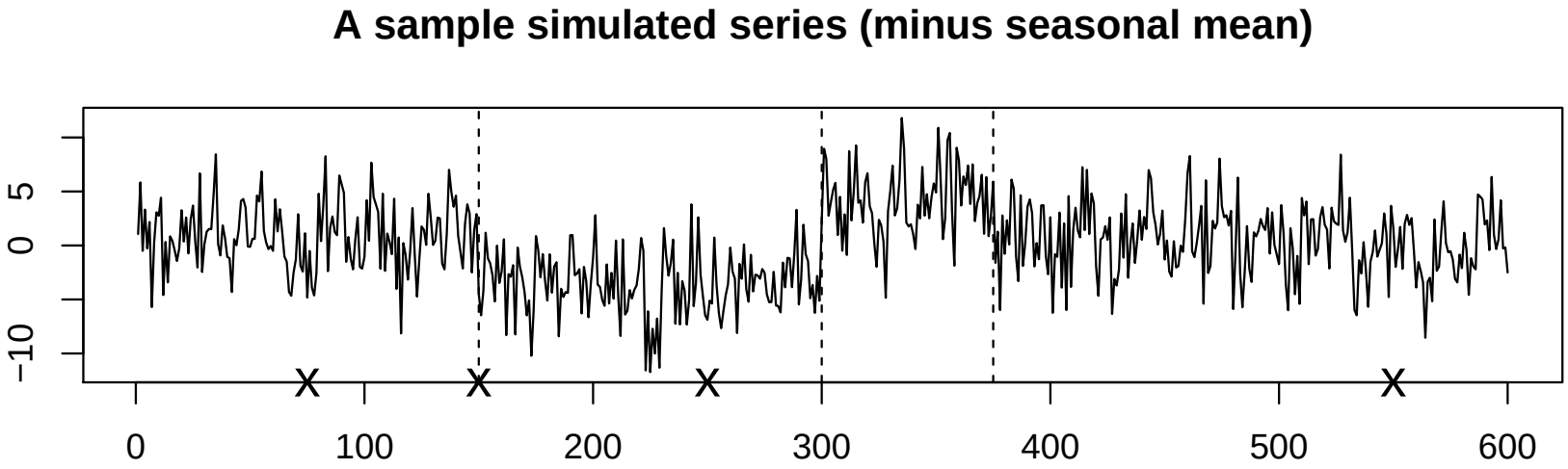
Scenario 2: non-monotonic shifts $\boldsymbol{\mu} = (0, -\Delta, \Delta, 0)'$

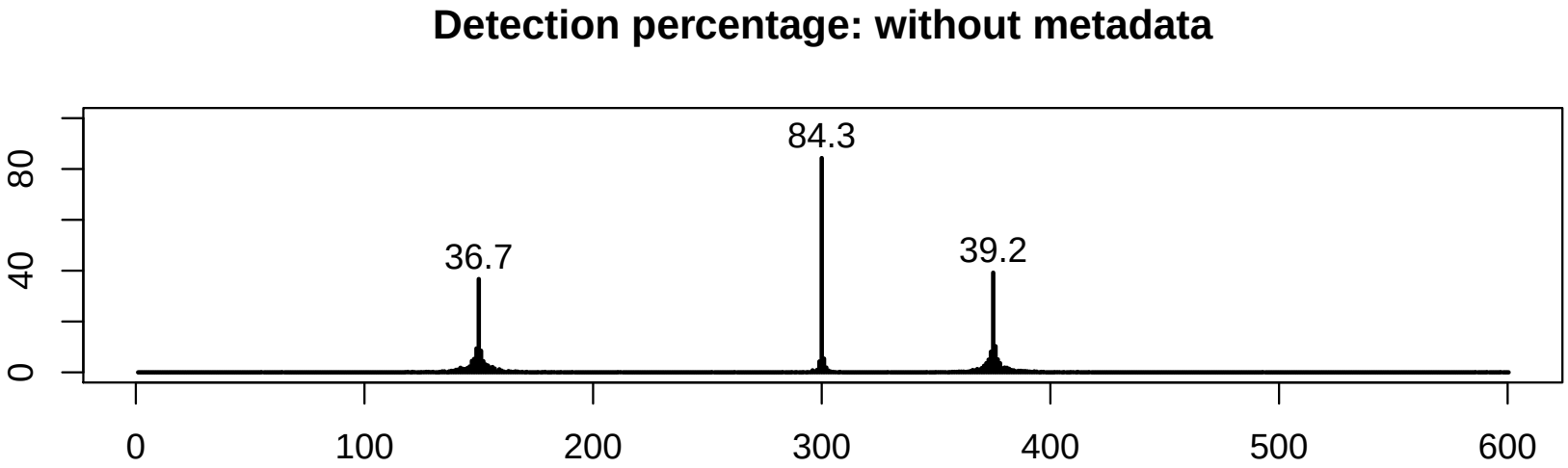
$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$



Scenario 2: non-monotonic shifts $\boldsymbol{\mu} = (0, -\Delta, \Delta, 0)'$

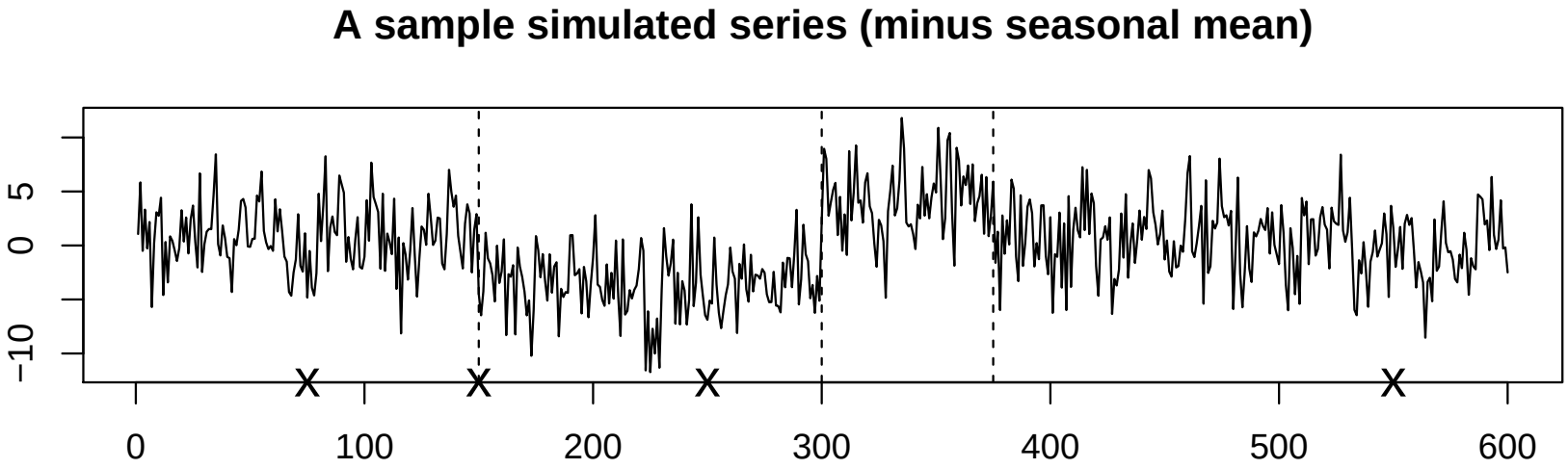
$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

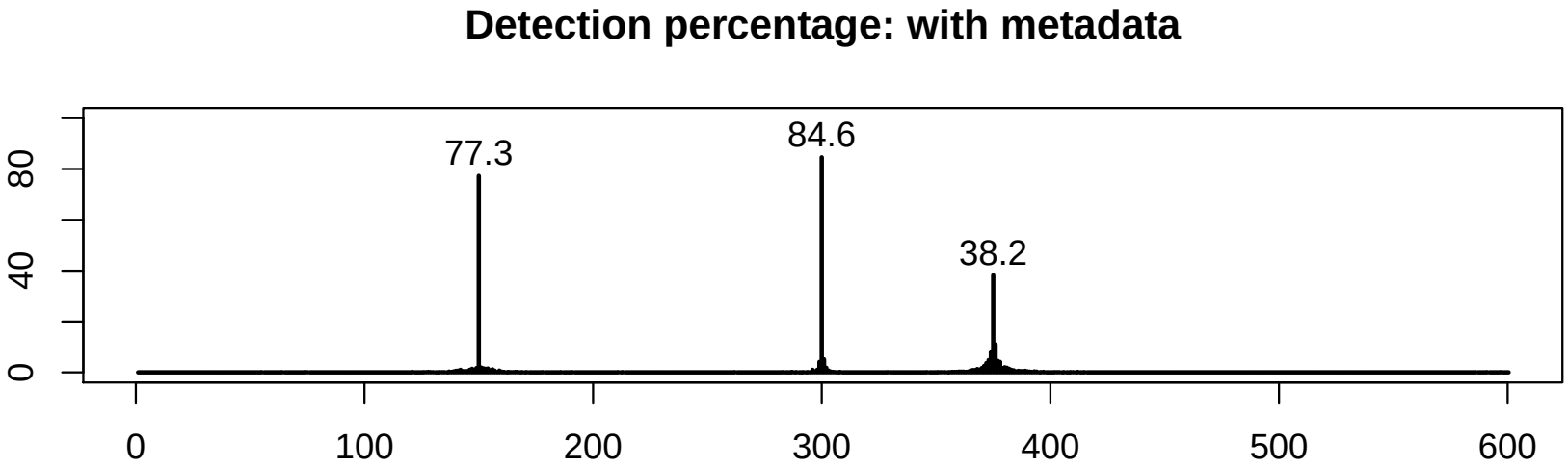




Scenario 2: non-monotonic shifts $\boldsymbol{\mu} = (0, -\Delta, \Delta, 0)'$

$p = 3, \boldsymbol{\phi} = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$





Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$
  Scale time to $[0, 1]$ by mapping time t to t/N.

$$\left.\begin{array}{ll} \boldsymbol{\lambda}: & 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\ \boldsymbol{\eta}: & 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \end{array}\right\} \longleftrightarrow \eta_r = \lfloor \lambda_r N \rfloor$$

Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$
  Scale time to $[0, 1]$ by mapping time t to t/N.

$$\left.\begin{array}{ll} \boldsymbol{\lambda}: & 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\ \boldsymbol{\eta}: & 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \end{array}\right\} \longleftrightarrow \eta_r = \lfloor \lambda_r N \rfloor$$

- True model $\boldsymbol{\lambda}^0$ has m^0 changepoints.
  The true number of changepoints m^0 is unknown.

Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$
  Scale time to $[0, 1]$ by mapping time t to t/N.

$$\left.\begin{array}{ll} \boldsymbol{\lambda}: & 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\ \boldsymbol{\eta}: & 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \end{array}\right\} \longleftrightarrow \eta_r = \lfloor \lambda_r N \rfloor$$

- True model $\boldsymbol{\lambda}^0$ has m^0 changepoints.
  The true number of changepoints m^0 is unknown.
- Consider all relative changepoint configurations in

$$\boldsymbol{\Lambda} = \{\boldsymbol{\lambda} : 0 \leq m \leq M, \min_{r=1,2,\ldots,m+1} \lambda_r - \lambda_{r-1} \geq d\}$$

  - M: a large integer, fixed, $M > m^0$
  - d: a very small positive constant

Asymptotic selection consistency

- The estimated relative changepoint model:

$$\hat{\boldsymbol{\lambda}}_N = \arg\min_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \ \mathrm{BMDL}(\boldsymbol{\lambda}),$$

- The estimated number of changepoints is $\hat{m}_N = |\hat{\boldsymbol{\lambda}}_N|$.

Theorem (Consistency for changepoint model estimation)

As $N \to \infty$, we have

$$\hat{m}_N \xrightarrow{P} m^0, \quad \textit{and} \quad \hat{\boldsymbol{\lambda}}_N \xrightarrow{P} \boldsymbol{\lambda}^0.$$

Furthermore, for each $r = 1, \ldots, m^0$,

$$\left|\hat{\lambda}_r - \lambda_r^0\right| = O_P\left(\frac{1}{N}\right).$$

Parameter estimation

Under the estimated chagnepoint model $\hat{\boldsymbol{\lambda}}$,

- Yule-Walker estimator for ϕ.
- Optimizers of BMDL for σ^2 and $\mathbf{s}$.
- Conditional posterior mean for $\boldsymbol{\mu}$.

Theorem (Consistency for parameter estimation)

As $N \to \infty$, all parameter estimators converge to the true values,

$$\hat{\boldsymbol{\mu}}_N \xrightarrow{P} \boldsymbol{\mu}^0, \quad \hat{\mathbf{s}}_N \xrightarrow{P} \mathbf{s}^0, \quad \hat{\sigma}_N^2 \xrightarrow{P} (\sigma^2)^0, \quad \hat{\phi}_N \xrightarrow{P} \phi^0.$$

Outline

Joint detection of Tmax and Tmin

- Any time in $t = \{p+1, p+2, \ldots, N\}$ can be a changepoint, for either Tmax or Tmin, or both.
- If both, can shift in the same or opposite directions.

Joint detection of Tmax and Tmin

- Any time in $t = \{p+1, p+2, \ldots, N\}$ can be a changepoint, for either Tmax or Tmin, or both.
- If both, can shift in the same or opposite directions.

Definition (bivariate changepoint model $\boldsymbol{\eta}$)

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_{p+1}, \boldsymbol{\eta}_{p+2}, \ldots, \boldsymbol{\eta}_N)' \in \mathbb{R}^{(N-p)\times 2},$$

$$\boldsymbol{\eta}_t = \begin{cases} (1,1)', & \text{if time } t \text{ is a concurrent changepoint} \\ (1,0)', & \text{if time } t \text{ is only a changepoint for Tmax} \\ (0,1)', & \text{if time } t \text{ is only a changepoint for Tmin} \\ (0,0)', & \text{if time } t \text{ is not a changepoint} \end{cases}$$

- Number of changepoints in $\boldsymbol{\eta}$: m_1 (Tmax), m_2 (Tmin)

Joint detection of Tmax and Tmin

- Any time in $t = \{p+1, p+2, \ldots, N\}$ can be a changepoint, for either Tmax or Tmin, or both.
- If both, can shift in the same or opposite directions.

Definition (bivariate changepoint model $\boldsymbol{\eta}$)

$$\boldsymbol{\eta} = (\boldsymbol{\eta}_{p+1}, \boldsymbol{\eta}_{p+2}, \ldots, \boldsymbol{\eta}_N)' \in \mathbb{R}^{(N-p)\times 2},$$

$$\boldsymbol{\eta}_t = \begin{cases} (1,1)', & \text{if time } t \text{ is a concurrent changepoint} \\ (1,0)', & \text{if time } t \text{ is only a changepoint for Tmax} \\ (0,1)', & \text{if time } t \text{ is only a changepoint for Tmin} \\ (0,0)', & \text{if time } t \text{ is not a changepoint} \end{cases}$$

- Number of changepoints in $\boldsymbol{\eta}$: m_1 (Tmax), m_2 (Tmin)
- Dirichlet-Multinomial prior: closed form $\pi(\boldsymbol{\eta})$.
- Hyper parameter choices: encourage concurrent shifts.

Bivariate model

Under a specific model $\boldsymbol{\eta}$,

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} s_{v,1} \\ s_{v,2} \end{pmatrix} + \begin{pmatrix} \mu_{r_1,1} \\ \mu_{r_2,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}, \quad t = 1, 2, \ldots, N.$$

- v time t is in season v
- r_i time t is in regime r_i, where $i = 1(\text{Tmax}), 2(\text{Tmin})$
- $s_{v,i}$ seasonal mean; $\mathbf{s}_1 \in \mathbb{R}^{12}, \mathbf{s}_2 \in \mathbb{R}^{12}$
- $\mu_{r_i,i}$ regime mean; $\boldsymbol{\mu}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\mu}_2 \in \mathbb{R}^{m_2}$
- $\{\boldsymbol{\epsilon}_t\}$ Gaussian VAR(p) errors, with white noise covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ and autoregression coefficients $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p \in \mathbb{R}^{2\times 2}$.

- Likelihood: normal

Bivariate model

Under a specific model $\boldsymbol{\eta}$,

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} s_{v,1} \\ s_{v,2} \end{pmatrix} + \begin{pmatrix} \mu_{r_1,1} \\ \mu_{r_2,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}, \quad t = 1, 2, \ldots, N.$$

v time t is in season v

r_i time t is in regime r_i, where $i = 1(\text{Tmax}), 2(\text{Tmin})$

$s_{v,i}$ seasonal mean; $\mathbf{s}_1 \in \mathbb{R}^{12}, \mathbf{s}_2 \in \mathbb{R}^{12}$

$\mu_{r_i,i}$ regime mean; $\boldsymbol{\mu}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\mu}_2 \in \mathbb{R}^{m_2}$

$\{\epsilon_t\}$ Gaussian VAR(p) errors, with white noise covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ and autoregression coefficients $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p \in \mathbb{R}^{2\times 2}$.

- Likelihood: normal
- Prior $\pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$: normal
- $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$: closed forms
- $\widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\Phi}}_1, \ldots, \widehat{\boldsymbol{\Phi}}_p$: Yule-Walker estimators

Bivariate model

Under a specific model $\boldsymbol{\eta}$,

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} s_{v,1} \\ s_{v,2} \end{pmatrix} + \begin{pmatrix} \mu_{r_1,1} \\ \mu_{r_2,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}, \quad t = 1, 2, \ldots, N.$$

v time t is in season v

r_i time t is in regime r_i, where $i = 1(\text{Tmax}), 2(\text{Tmin})$

$s_{v,i}$ seasonal mean; $\mathbf{s}_1 \in \mathbb{R}^{12}, \mathbf{s}_2 \in \mathbb{R}^{12}$

$\mu_{r_i,i}$ regime mean; $\boldsymbol{\mu}_1 \in \mathbb{R}^{m_1}, \boldsymbol{\mu}_2 \in \mathbb{R}^{m_2}$

$\{\epsilon_t\}$ Gaussian VAR(p) errors, with white noise covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times 2}$ and autoregression coefficients $\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p \in \mathbb{R}^{2\times 2}$.

- Likelihood: normal
- Prior $\pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$: normal
- $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$: closed forms
- $\widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\Phi}}_1, \ldots, \widehat{\boldsymbol{\Phi}}_p$: Yule-Walker estimators

BMDL has a closed form.

Combining simulated data Scenario 1 and 2

Two concurrent shifts: 150 (↑↓), 300 (↑↑).

- Tmax (Series 1): shifts at 150, 300, 450. $\boldsymbol{\mu}_1 = (0, \Delta, 2\Delta, 3\Delta)'$
- Tmin (Series 2): shifts at 150, 300, 375. $\boldsymbol{\mu}_2 = (0, -\Delta, \Delta, 0)'$

Combining simulated data Scenario 1 and 2

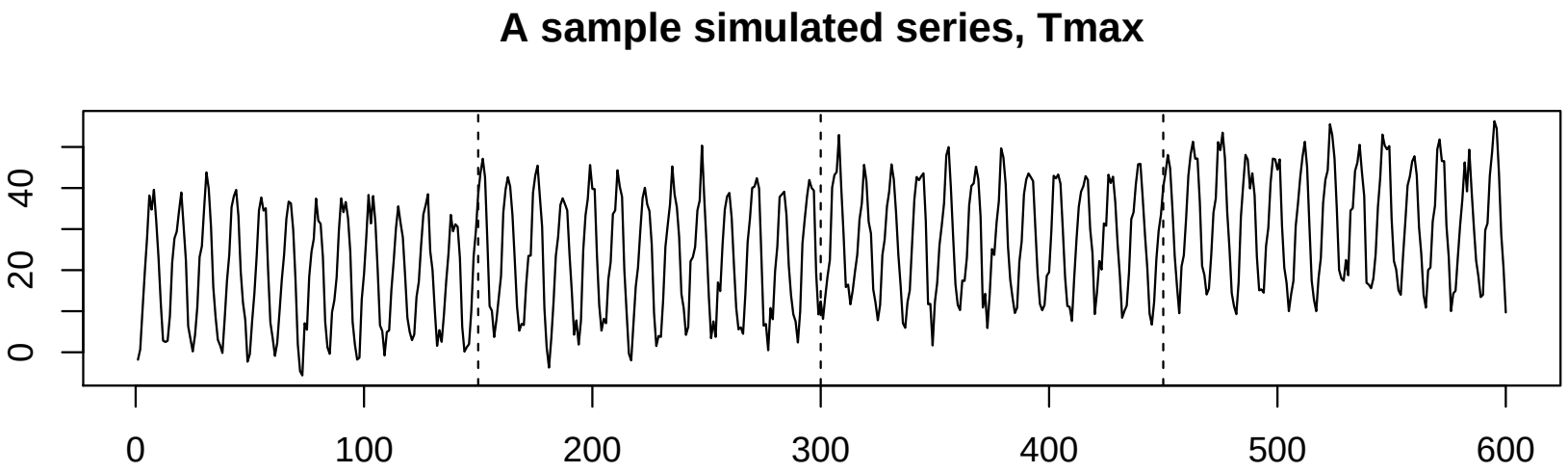
Two concurrent shifts: 150 (↑↓), 300 (↑↑).

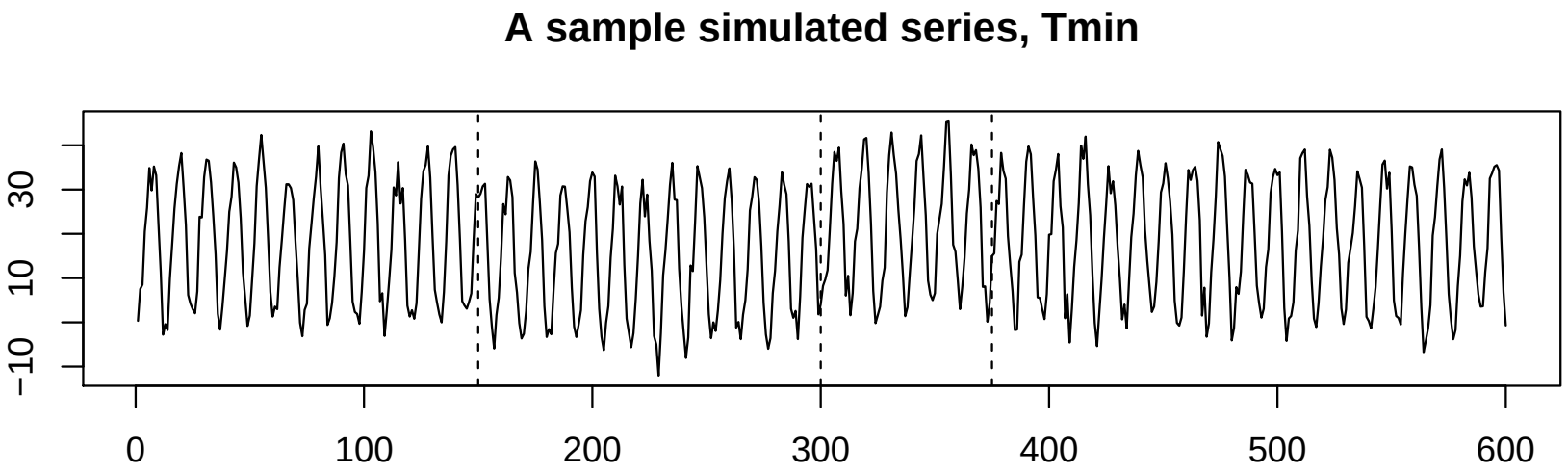
- Tmax (Series 1): shifts at 150, 300, 450. $\boldsymbol{\mu}_1 = (0, \Delta, 2\Delta, 3\Delta)'$
- Tmin (Series 2): shifts at 150, 300, 375. $\boldsymbol{\mu}_2 = (0, -\Delta, \Delta, 0)'$

VAR parameters

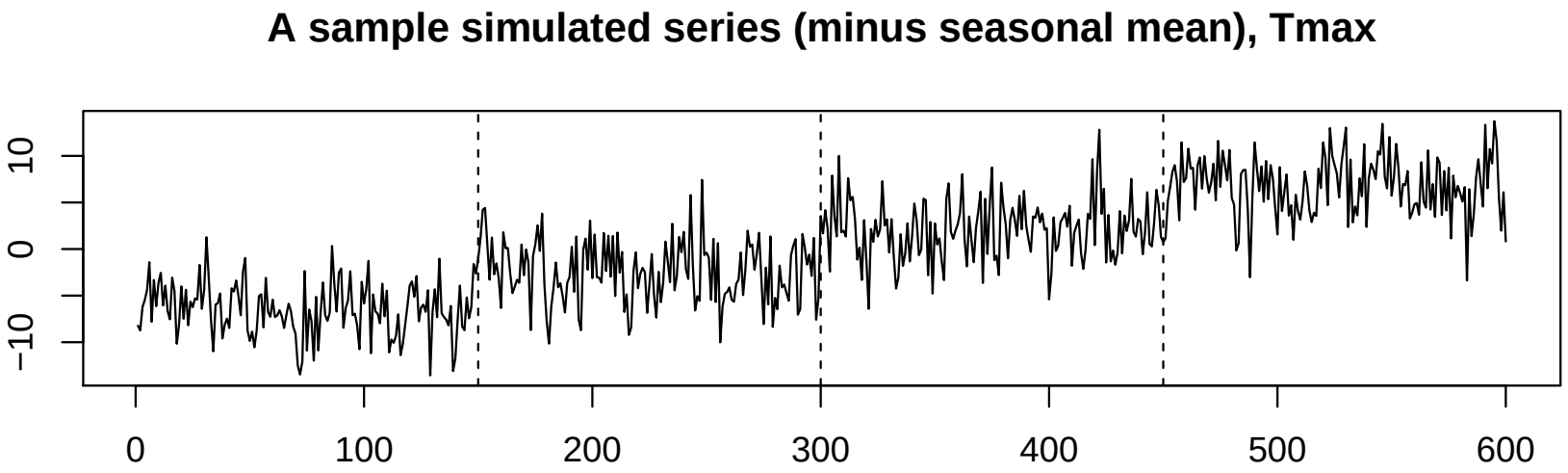
- $\boldsymbol{\Sigma} = \begin{pmatrix} 9 & 2 \\ 2 & 9 \end{pmatrix}, \Delta/3 = 1.5$
- $p = 3, \boldsymbol{\Phi}_1 = \begin{pmatrix} 0.2 & 0.02 \\ 0.02 & 0.2 \end{pmatrix}, \boldsymbol{\Phi}_2 = \begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix},$
  $\boldsymbol{\Phi}_3 = \begin{pmatrix} 0.05 & 0.005 \\ 0.005 & 0.05 \end{pmatrix}$

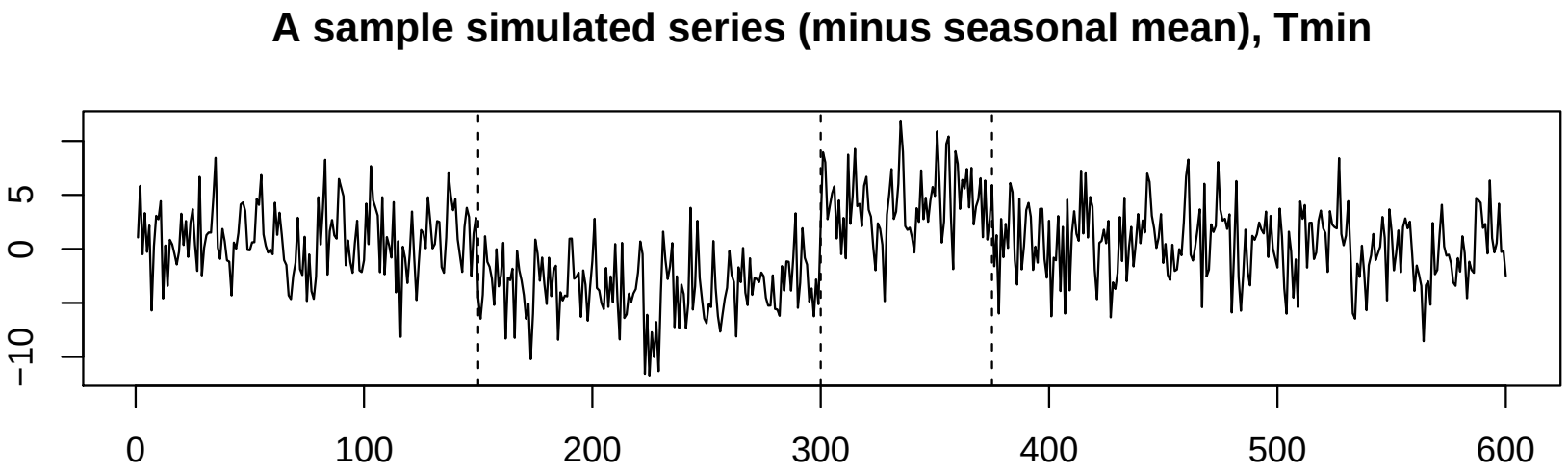
Combining simulated data Scenario 1 and 2



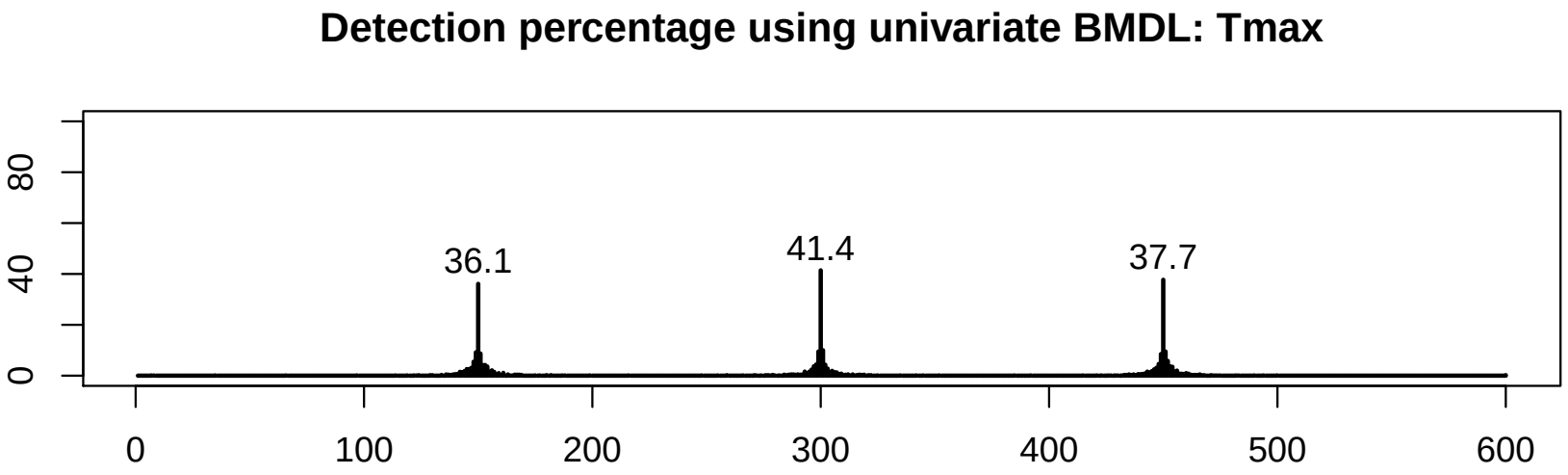


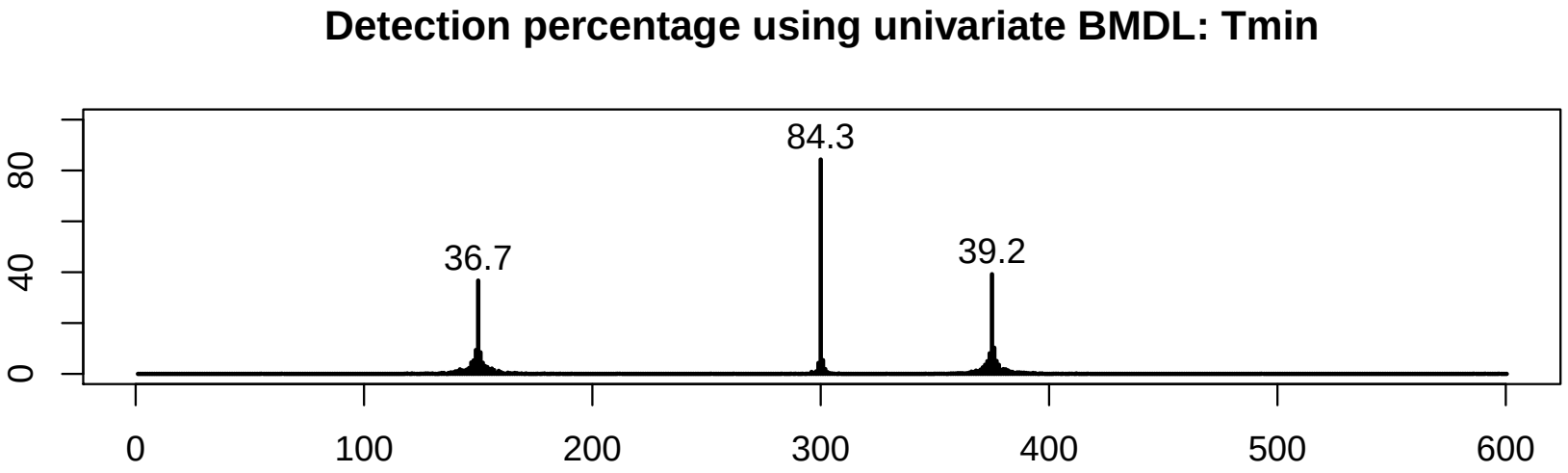
Combining simulated data Scenario 1 and 2



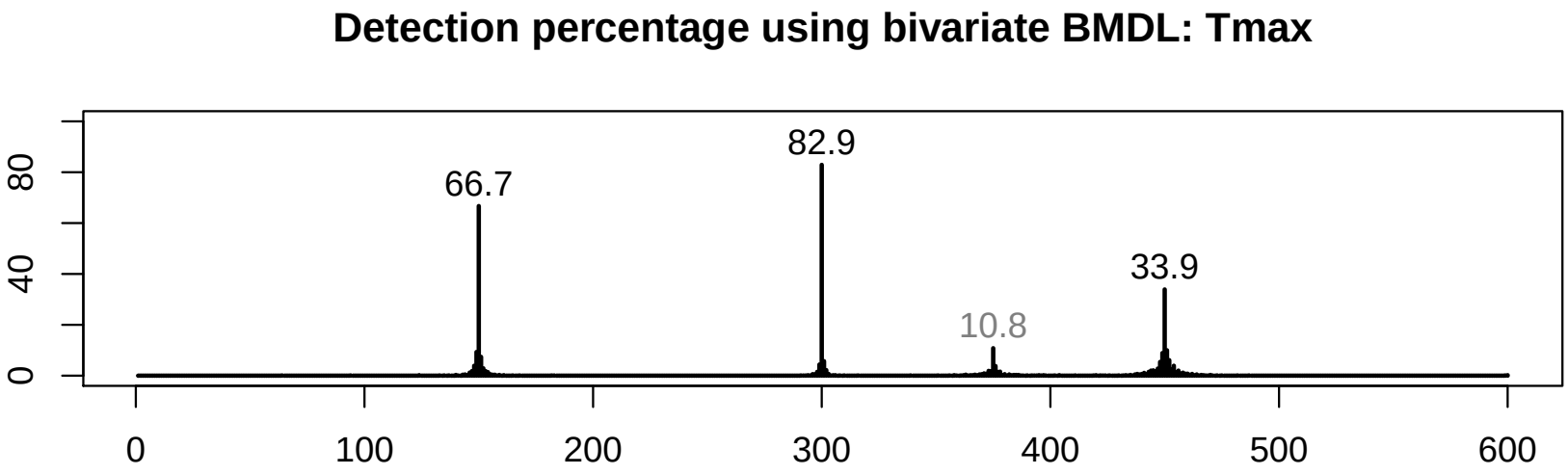


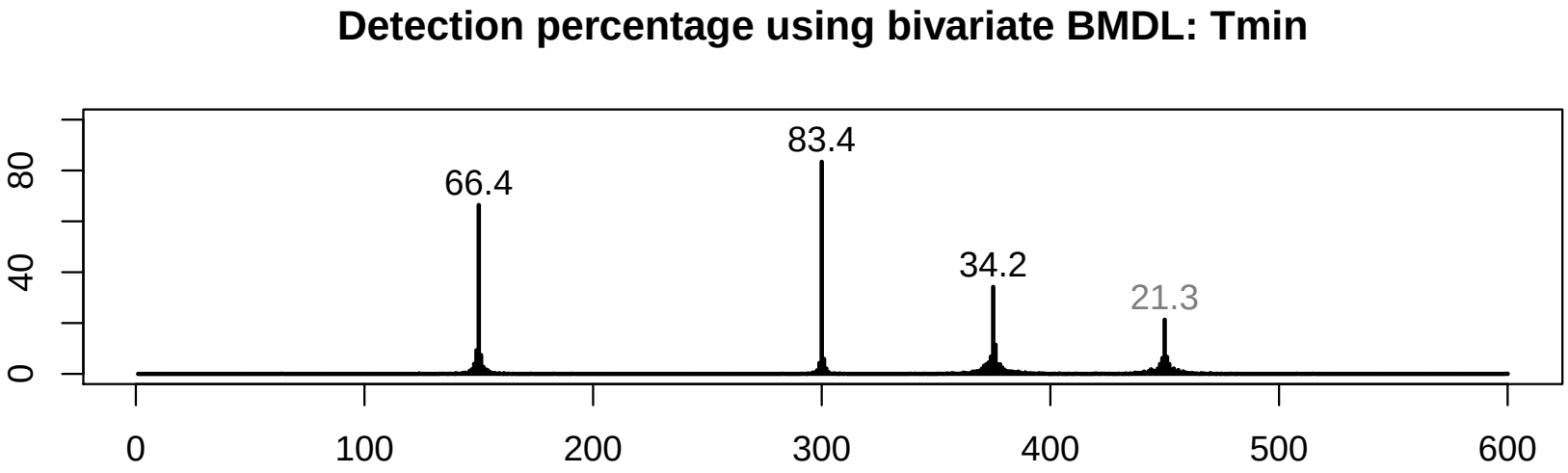
Combining simulated data Scenario 1 and 2





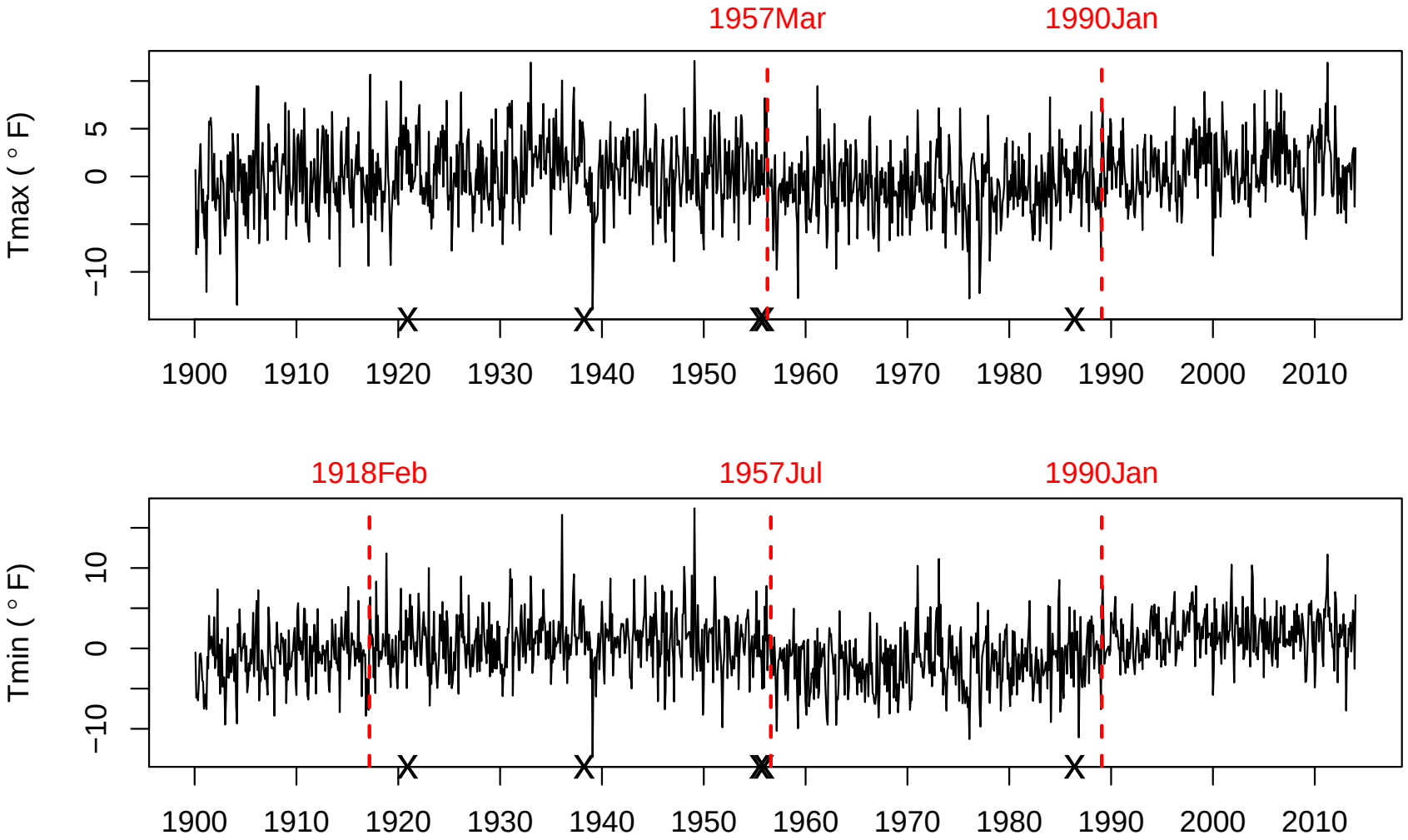
Combining simulated data Scenario 1 and 2



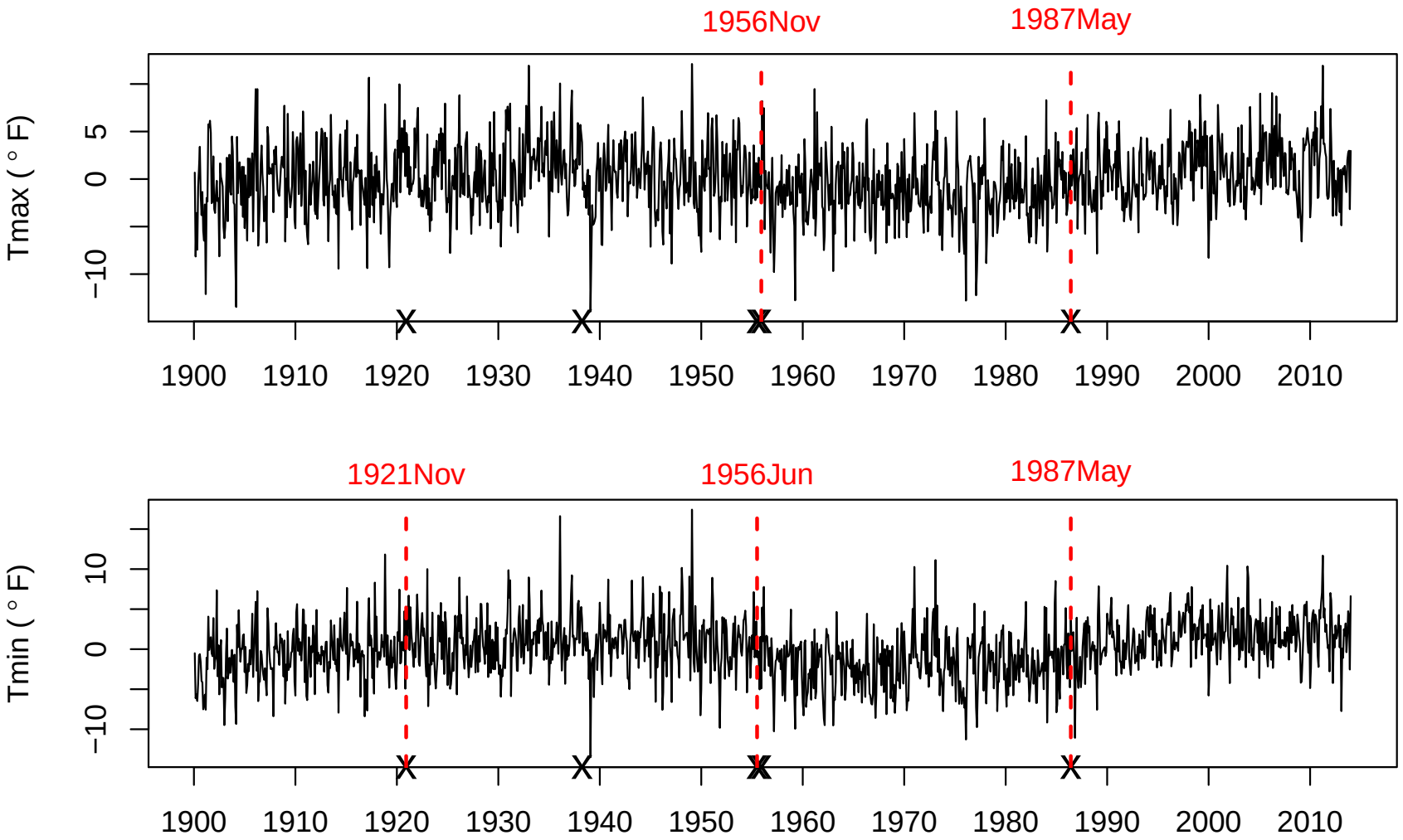


Outline

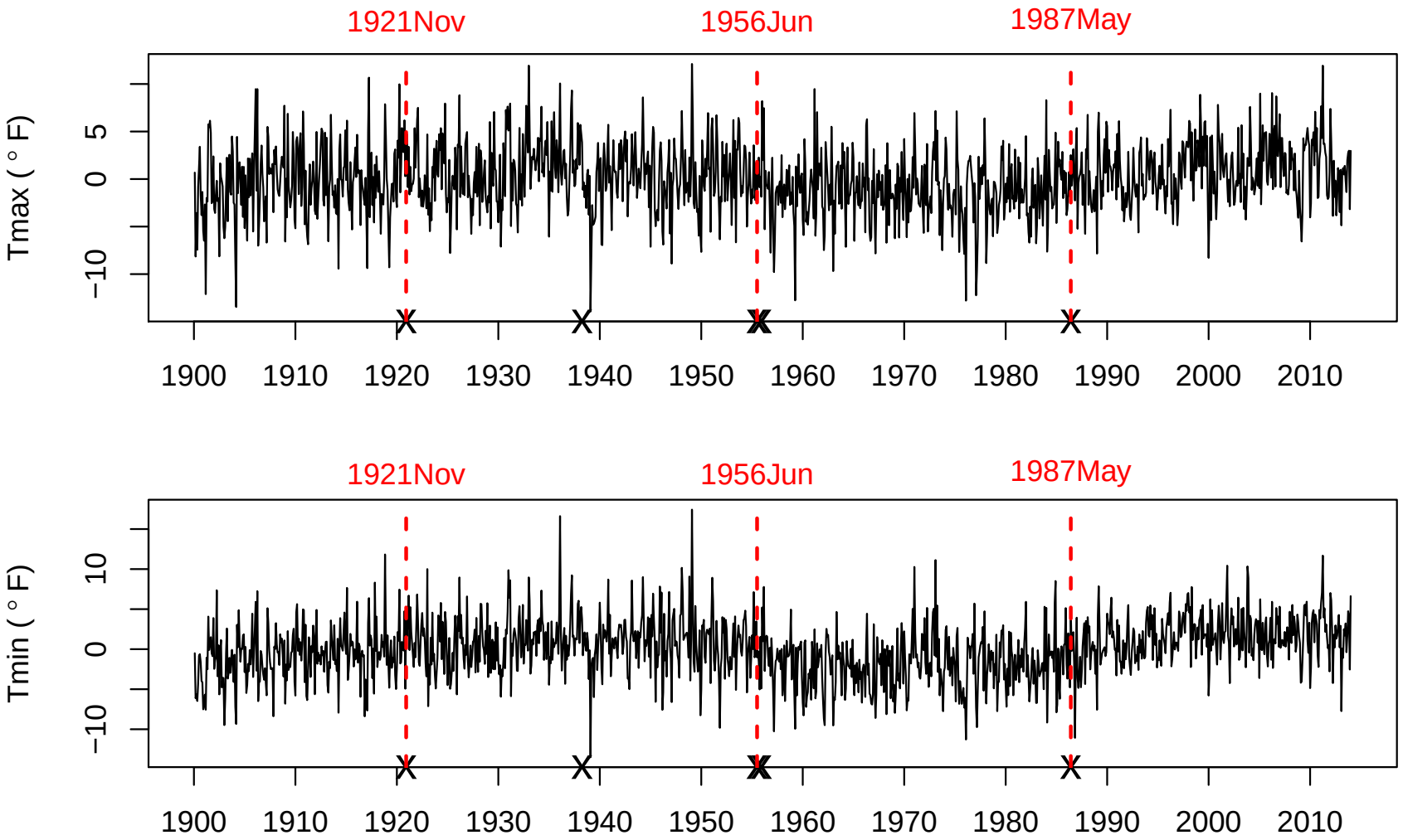
Univariate BMDL: no metadata



Univariate BMDL: with metadata



Bivariate BMDL: with metadata



Reference I

Richard A Davis and Chun Yip Yau. Consistency of minimum description length model selection for piecewise stationary time series models. *Electronic Journal of Statistics*, 7:381–411, 2013.

Richard A Davis, Thomas C. M. Lee, and Gabriel A Rodriguez-Yam. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239, 2006.

Chao Du, Chu-Lan Michael Kao, and S. C. Kou. Stepwise signal extraction via marginal likelihood. *Journal of the American Statistical Association*, 111(513):314–330, 2016.

Edward I. George and Robert E. McCulloch. Approaches for Bayesian variable selection. *Statistics Sinica*, 7:339–373, 1997.

Mark H. Hansen and Bin Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774, 2001.

J. Murray Mitchell. On the causes of instrumentally observed secular temperature trends. *Journal of Meteorology*, 10:244–261, 1953.

Jorma Risanen. *Stochastic Complexity in Statistical Inquiry*, volume 511. World Scientific, Singapore, 1989.

Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(143-157):623, 1948.