Spatio-temporal Models

Lecture 24

Dr. Colin Rundel

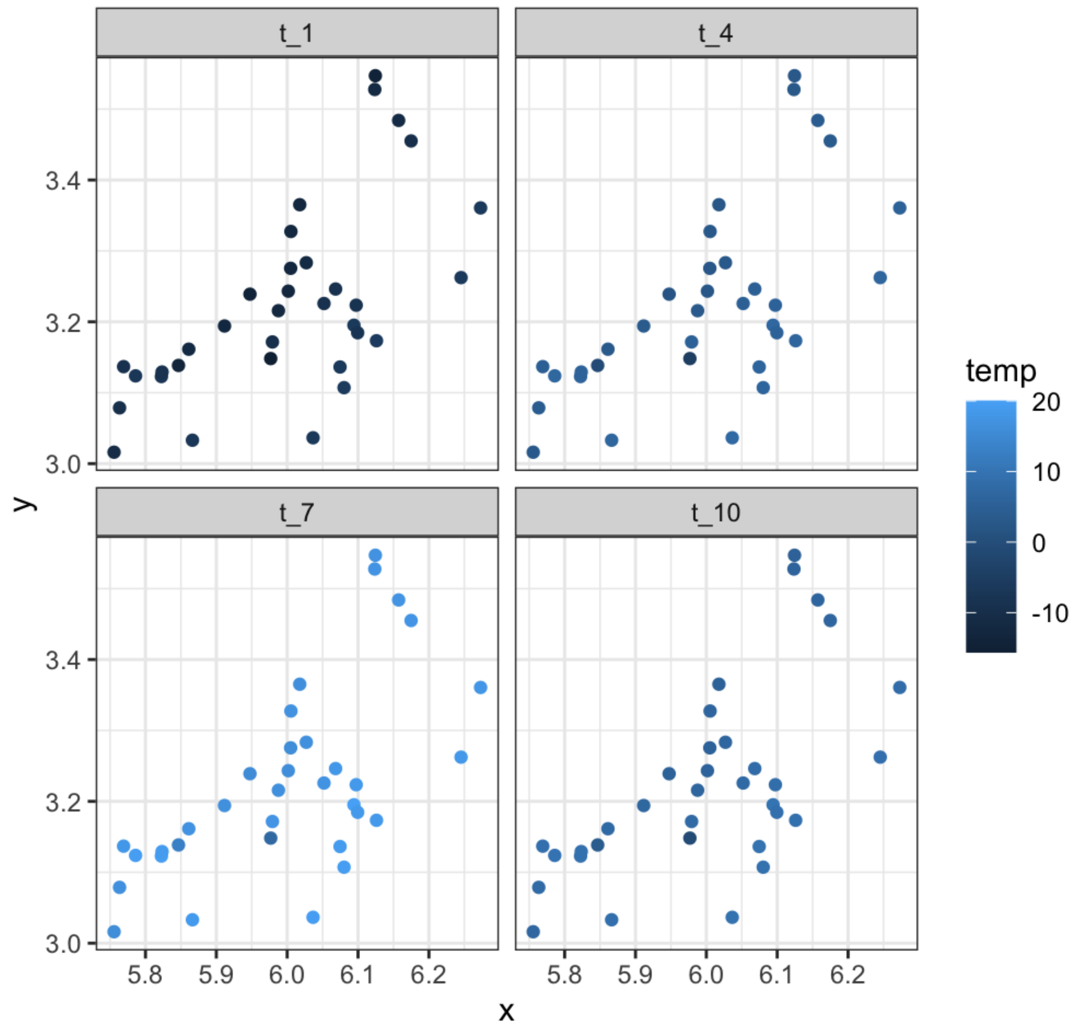
Spatial Models with AR time dependence

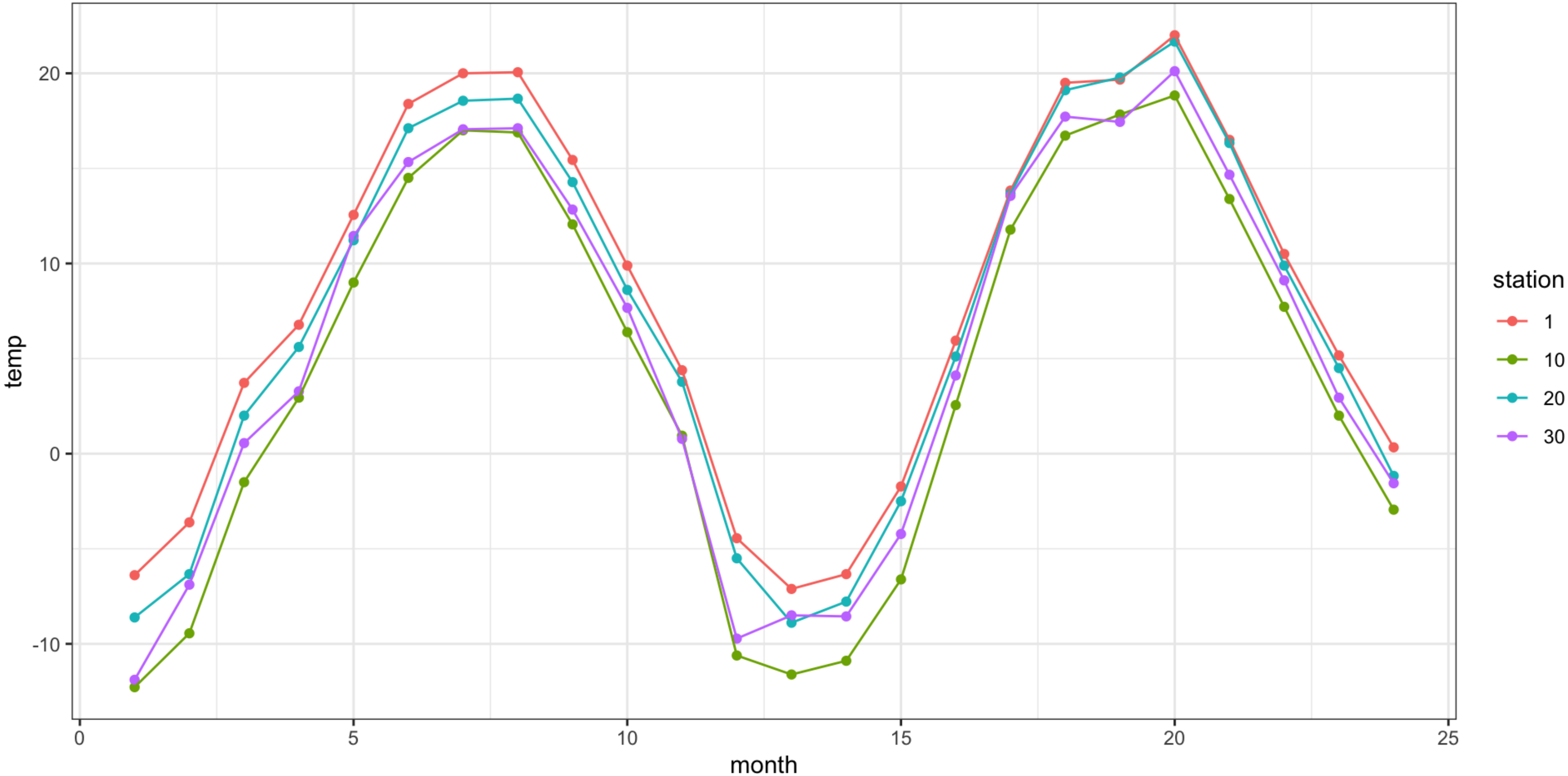
Example - Weather station data

`NETemp.dat` - Monthly temperature data (Celsius) recorded across the Northeastern US starting in January 2000.

```
# A tibble: 34 × 27
        x     y  elev    t_1   t_2    t_3   t_4   t_5   t_6   t_7   t_8
    <dbl> <dbl> <int>  <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
 1   6.09  3.20   102  -6.39 -3.61  3.72   6.78 12.6   18.4  20    20.1
 2   6.25  3.26     1  -6.28 -4.11  2.61   6.56 11.4   16.8  18.4  18.7
 3   6.16  3.48   157 -11.1  -9.44 -0.389  3.94  9.89  15.4  17.5  17.4
 4   6.12  3.53   176 -11.6  -9.72 -1.17   2.89  9.67  14.8  17.4  16.9
 5   6.00  3.28   400 -12.6  -9.06 -1.61   2.56  8.56  14.3  15.9  15.8
 6   6.05  3.23   133  -9.11 -6.39  1.22   4.94 10.9   15.9  17.3  17.6
 7   6.10  3.18    56  -7.94 -6.06  2.06   5.56 11.1   17    18.6  18.8
 8   6.07  3.14    59  -6.56 -3.5   3.17   6.17 11.5   17.4  19.1  19.4
 9   6.17  3.46   160  -9.94 -8.94 -0.278  3.56  9.61  15.3  17.7  17.3
10   6.01  3.33   360 -12.3  -9.44 -1.5    2.94  9     14.5  17    16.9
# … with 24 more rows, and 16 more variables: t_9 <dbl>, t_10 <dbl>,
#   t_11 <dbl>, t_12 <dbl>, t_13 <dbl>, t_14 <dbl>, t_15 <dbl>,
#   t_16 <dbl>, t_17 <dbl>, t_18 <dbl>, t_19 <dbl>, t_20 <dbl>,
```

Based on Andrew Finley and Sudipto Banerjee's notes from National Ecological Observatory Network


t_1
t_4
t_7
t_10
3.4
3.2
3.0
y
5.8
5.9
6.0
6.1
6.2
x
temp
20
10
0
-10



20
10
0
-10
temp
0
5
10
15
20
25
month
station
1
10
20
30

Dynamic Linear / State Space Models (time)

$$\underset{1\times 1}{y_t} = \underset{1\times p}{\boldsymbol{F}_t'} \; \underset{p\times 1}{\boldsymbol{\theta}_t} + v_t \qquad \text{observation equation}$$

$$\underset{p\times 1}{\boldsymbol{\theta}_t} = \underset{p\times p}{\boldsymbol{G}_t} \; \underset{p\times 1}{\boldsymbol{\theta}_{t-1}} + \underset{p\times 1}{\boldsymbol{\omega}_t} \qquad \text{evolution equation}$$

$$\boldsymbol{v}_t \sim (0, \boldsymbol{V}_t)$$
$$\boldsymbol{\omega}_t \sim (0, \boldsymbol{W}_t)$$

DLM vs ARMA

ARMA / ARIMA are special cases of the more general dynamic linear model framework, for example an $\text{AR}(p)$ can be written as

$$F_t' = (1, 0, \ldots, 0)$$

$$G_t = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

$$\omega_t = (\omega_1, 0, \ldots, 0), \qquad \omega_1 \sim \quad (0,\ \sigma^2)$$

$$y_t = \theta_t + v_t$$

$$\theta_t = \sum_{i=1}^{p} \phi_i\, \theta_{t-i} + \omega_1$$

$$v_t \sim \quad (0,\ \sigma_v^2)$$

$$\omega_1 \sim \quad (0,\ \sigma_\omega^2)$$

Dynamic spatio-temporal model

The observed temperature at time t and location s is given by $y_t(s)$ where,

$$
\begin{aligned}
y_t(\boldsymbol{s}) &= \boldsymbol{x}_t(\boldsymbol{s})\boldsymbol{\beta}_t + u_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}) \\
\epsilon_t(\boldsymbol{s}) &\overset{\text{ind}}{\sim} \quad (0, \tau_t^2)
\end{aligned}
$$

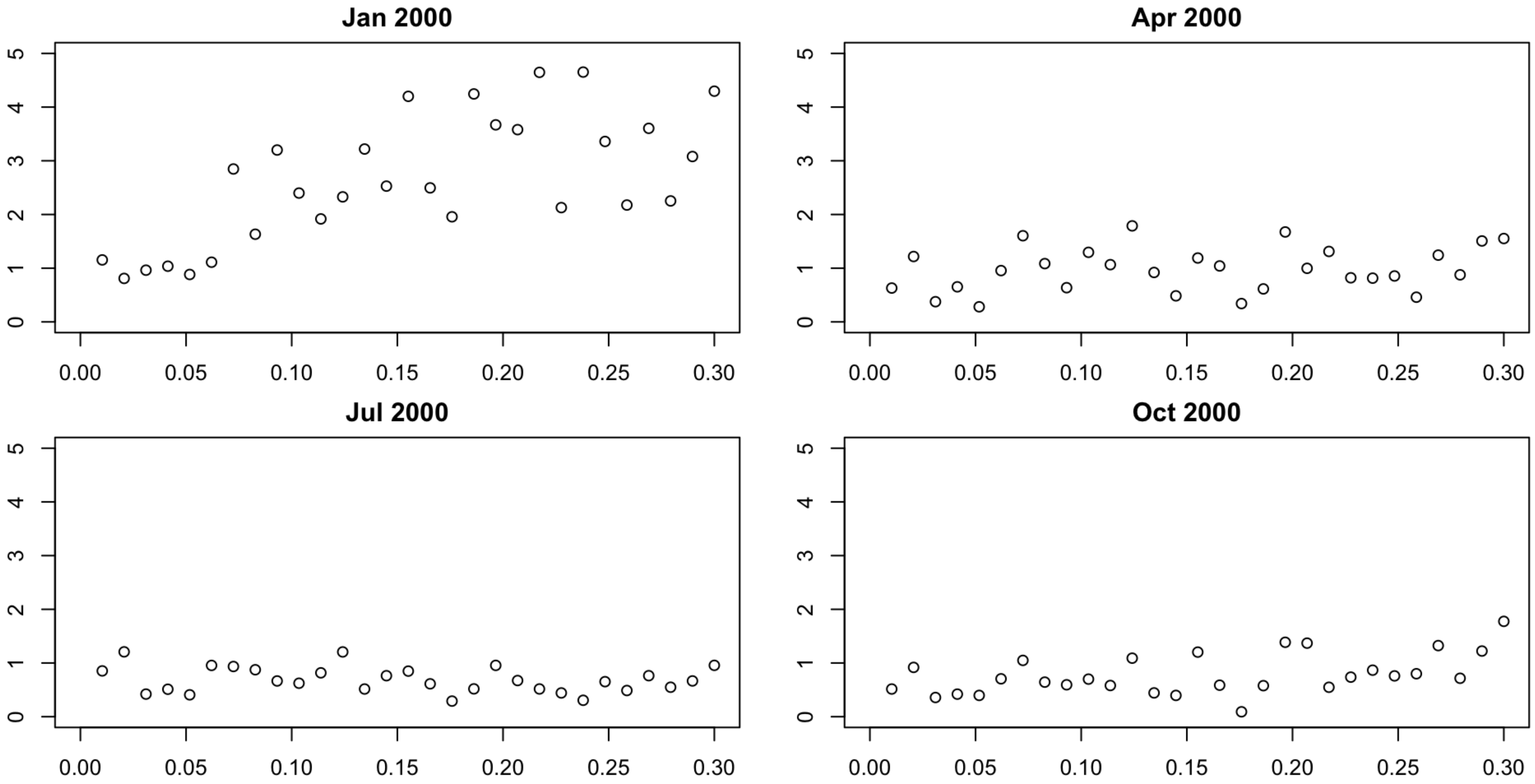
$$
\begin{aligned}
\boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t \\
\boldsymbol{\eta}_t &\overset{\text{i.i.d.}}{\sim} \quad (0, \boldsymbol{\Sigma}_\eta)
\end{aligned}
$$

$$
\begin{aligned}
u_t(\boldsymbol{s}) &= u_{t-1}(\boldsymbol{s}) + w_t(\boldsymbol{s}) \\
w_t(\boldsymbol{s}) &\overset{\text{ind.}}{\sim} \quad \left(\boldsymbol{0}, \Sigma_t(\phi_t, \sigma_t^2)\right)
\end{aligned}
$$

Additional assumptions for $t = 0$,

$$
\begin{aligned}
\boldsymbol{\beta}_0 &\sim \quad (\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\
u_0(\boldsymbol{s}) &= 0
\end{aligned}
$$

Variograms by time



Data and Model Parameters

Data:

```
max_d = coords %>% dist() %>% max()
n_t = 24
n_s = nrow(ne_temp)
```

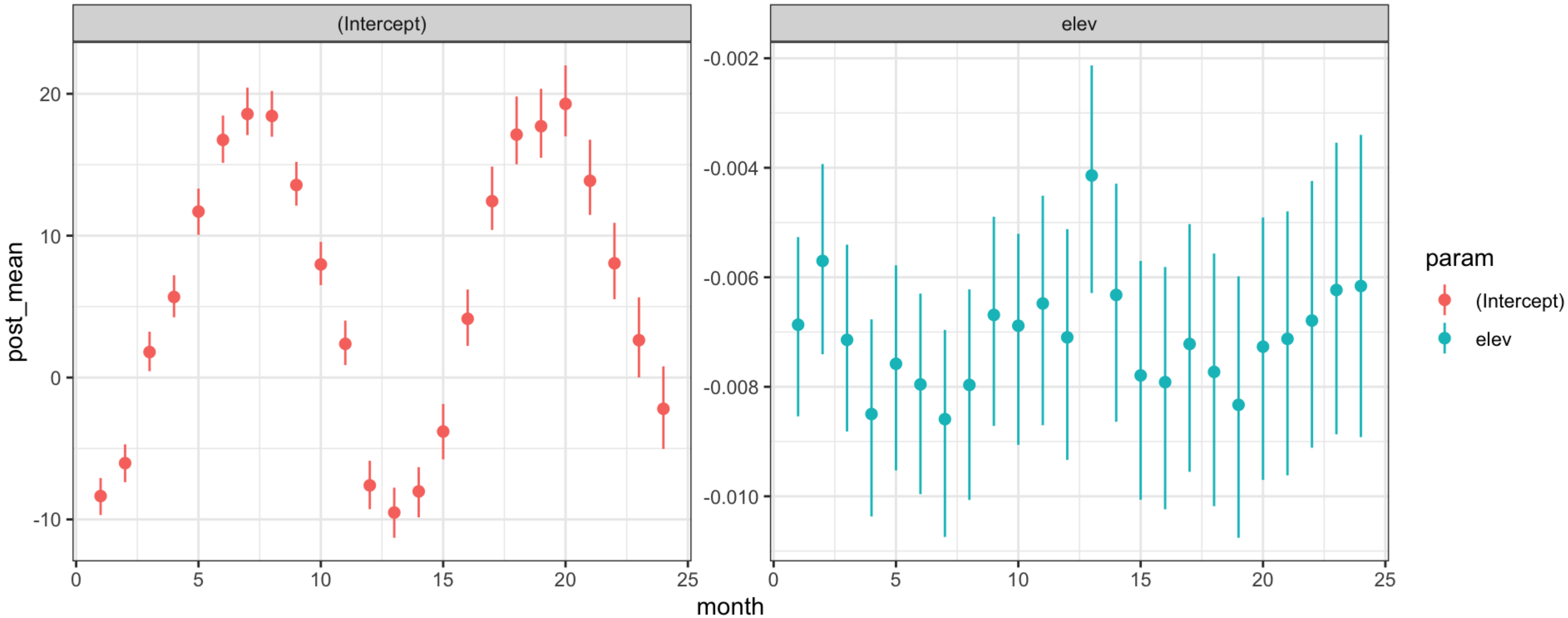
Parameters:

```
n_beta = 2
starting = list(
  beta = rep(0, n_t * n_beta), phi = rep(3/(max_d/4), n_t),
  sigma.sq = rep(1, n_t), tau.sq = rep(1, n_t),
  sigma.eta = diag(0.01, n_beta)
)
tuning = list(phi = rep(1, n_t))
priors = list(
  beta.0.Norm = list(rep(0, n_beta), diag(1000, n_beta)),
  phi.Unif = list(rep(3/(0.9 * max_d), n_t), rep(3/(0.05 * max_d), n_t)),
  sigma.sq.IG = list(rep(2, n_t), rep(2, n_t)),
  tau.sq.IG = list(rep(2, n_t), rep(2, n_t)),
  sigma.eta.IW = list(2, diag(0.001, n_beta))
)
```

Fitting with spDynLM from spBayes

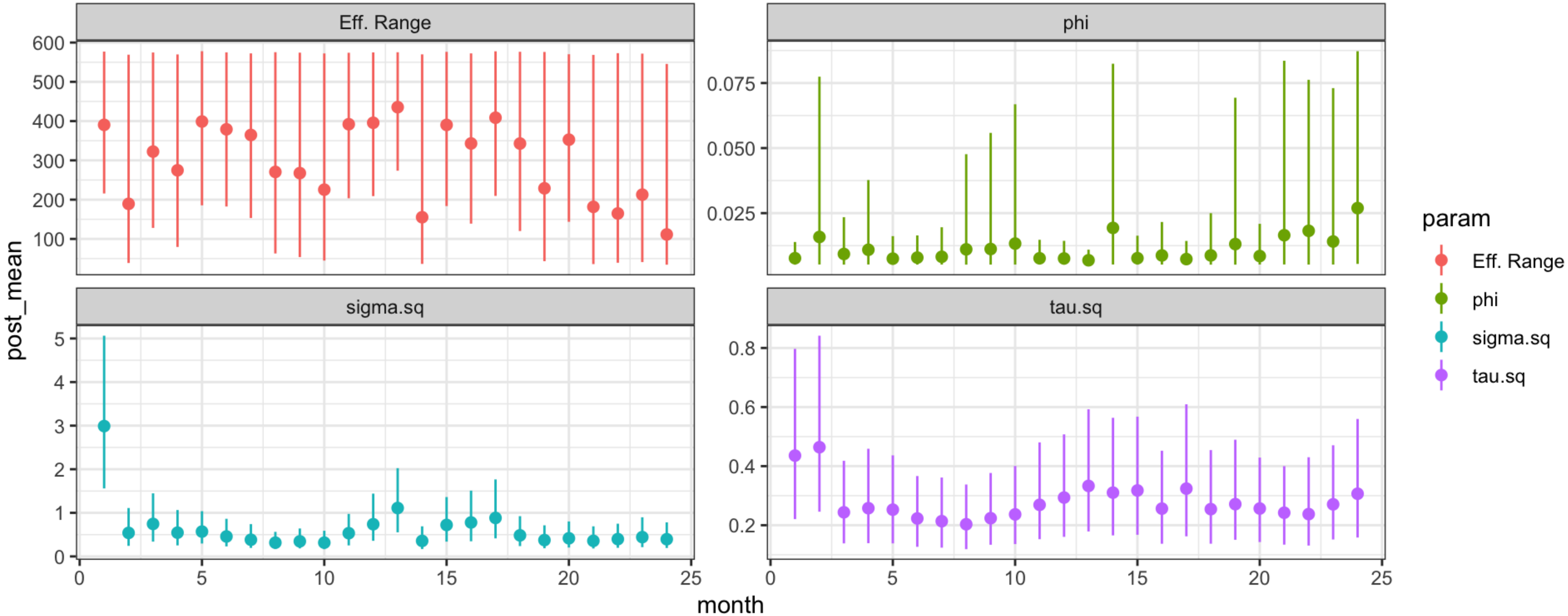
```
n_samples = 10000
models = lapply(paste0("t_",1:24, "~elev"), as.formula)

m = spBayes::spDynLM(
  models, data = ne_temp, coords = coords, get.fitted = TRUE,
  starting = starting, tuning = tuning, priors = priors,
  cov.model = "exponential", n.samples = n_samples, n.report = 1000
)

##  ----------------------------------------
##      General model description
##  ----------------------------------------
##  Model fit with 34 observations in 24 time steps.
##
##  Number of missing observations 0.
##
##  Number of covariates 2 (including intercept if specified).
##
##  Using the exponential spatial correlation model.
##
##  Number of MCMC samples 10000.
##
```

Posterior Inference - βs

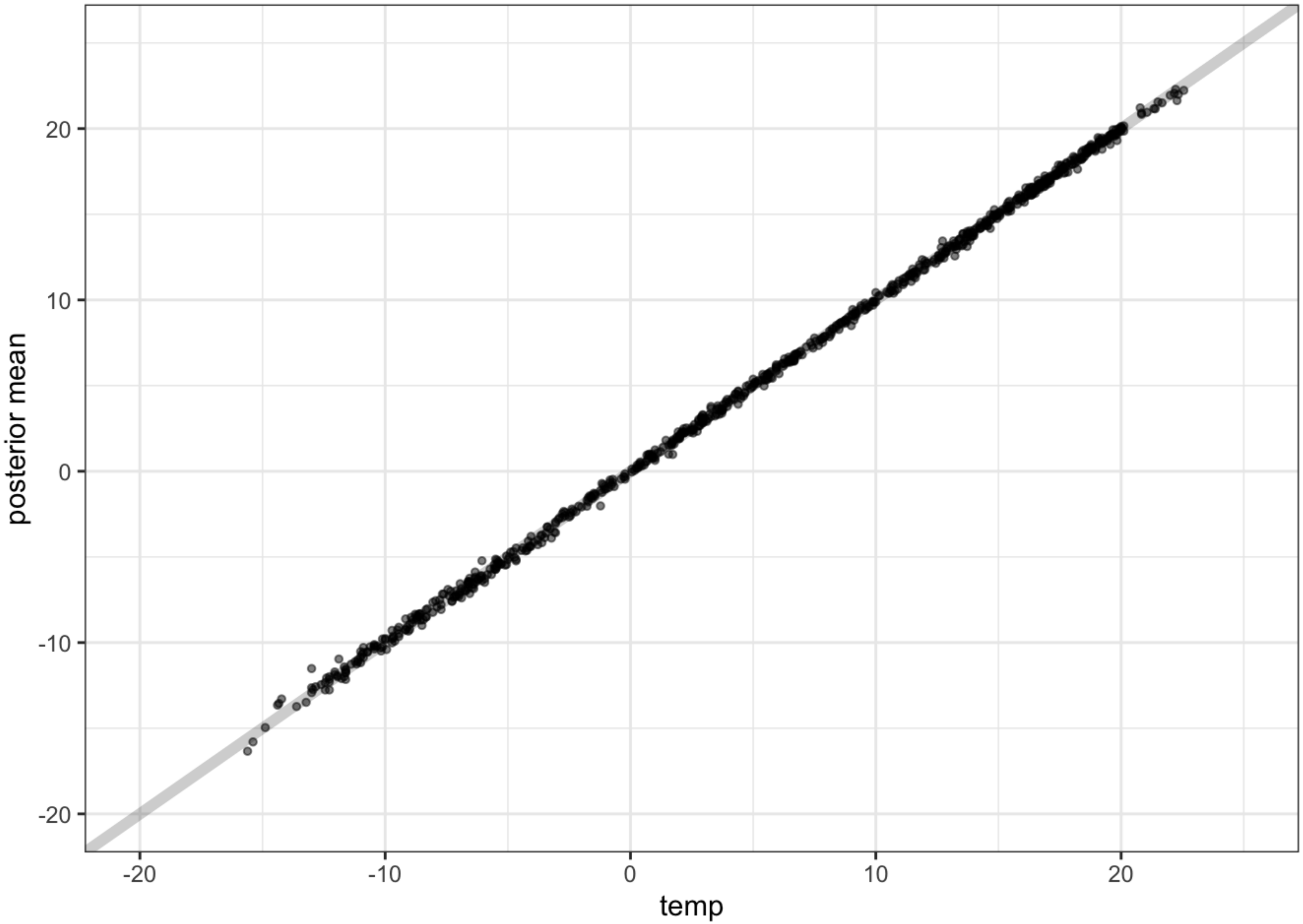


Lapse Rate $\approx -9.8\ ^{\circ}\mathrm{C/km}$.

Posterior Inference - θ



Posterior Inference - Observed vs. Predicted



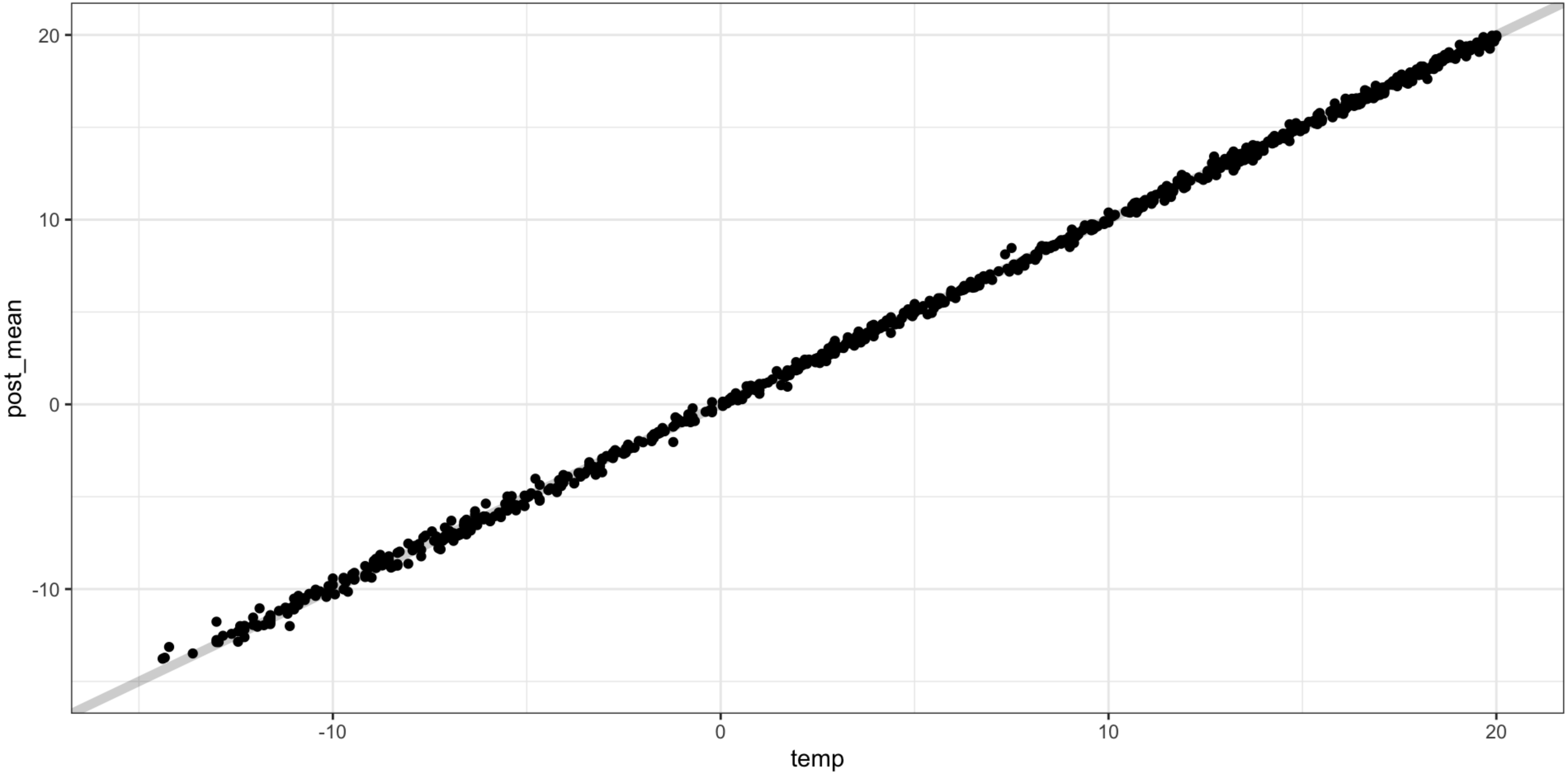
Prediction

`spPredict` does not support `spDynLM` objects but it will impute missing values.

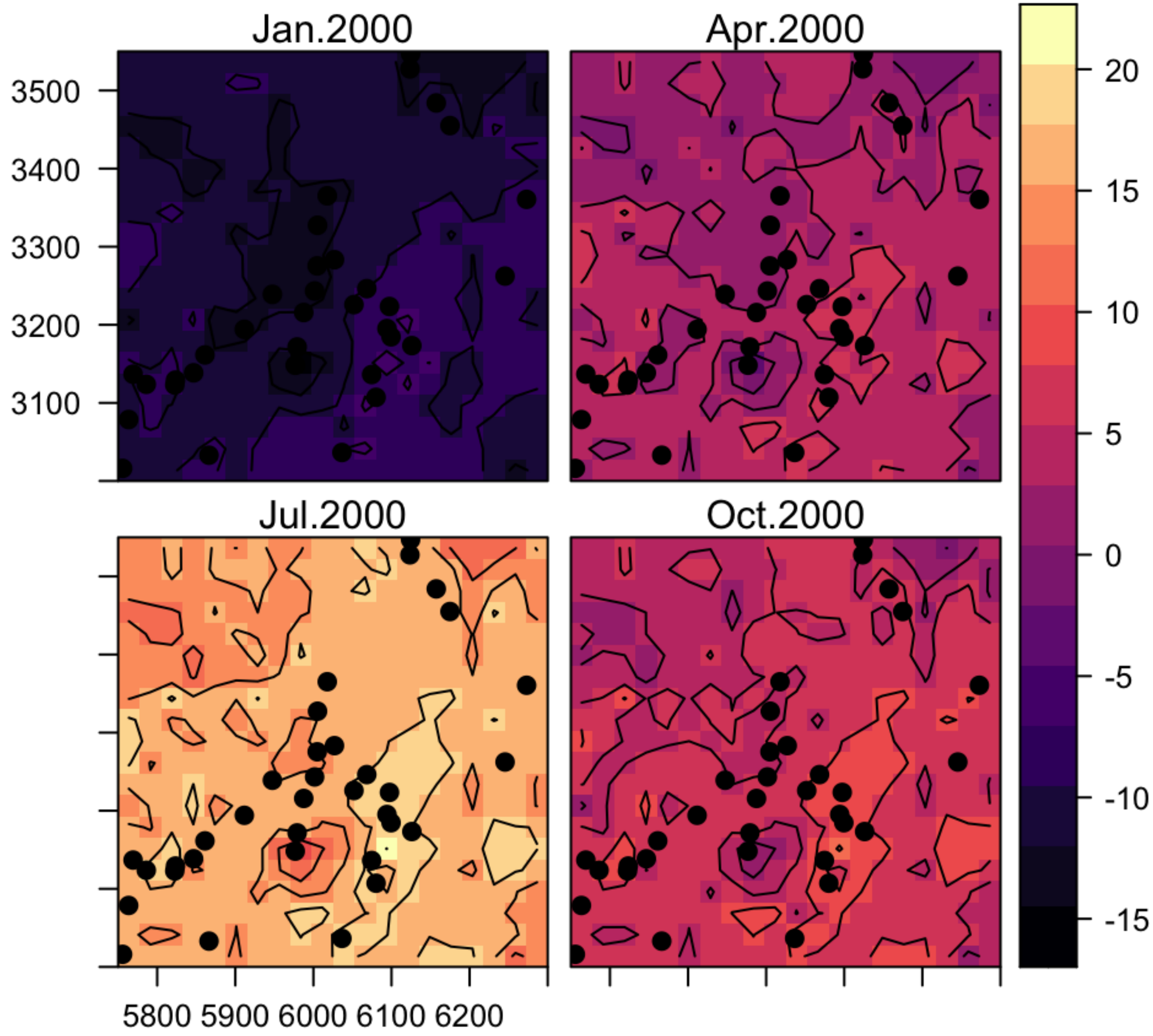
```
r = raster(xmn=5750, xmx=6300, ymn=3000, ymx=3550, nrow=20, ncol=20)

pred = xyFromCell(r, 1:length(r)) %>%
  as.data.frame() %>%
  mutate(type="pred") %>%
  bind_rows(
    ne_temp %>% mutate(type = "obs"),
    .
  )
```

```r
models_pred = lapply(paste0("t_",1:n_t, "~1"), as.formula)

n_samples = 5000
m_pred = spBayes::spDynLM(
  models_pred, data = pred, coords = coords_pred, get.fitted = TRUE,
  starting = starting, tuning = tuning, priors = priors,
  cov.model = "exponential", n.samples = n_samples, n.report = 1000)

## ----------------------------------------
##  General model description
## ----------------------------------------
## Model fit with 434 observations in 24 time steps.
##
## Number of missing observations 9600.
##
## Number of covariates 1 (including intercept if specified).
##
## Using the exponential spatial correlation model.
##
## Number of MCMC samples 5000.
```

Predictive performance

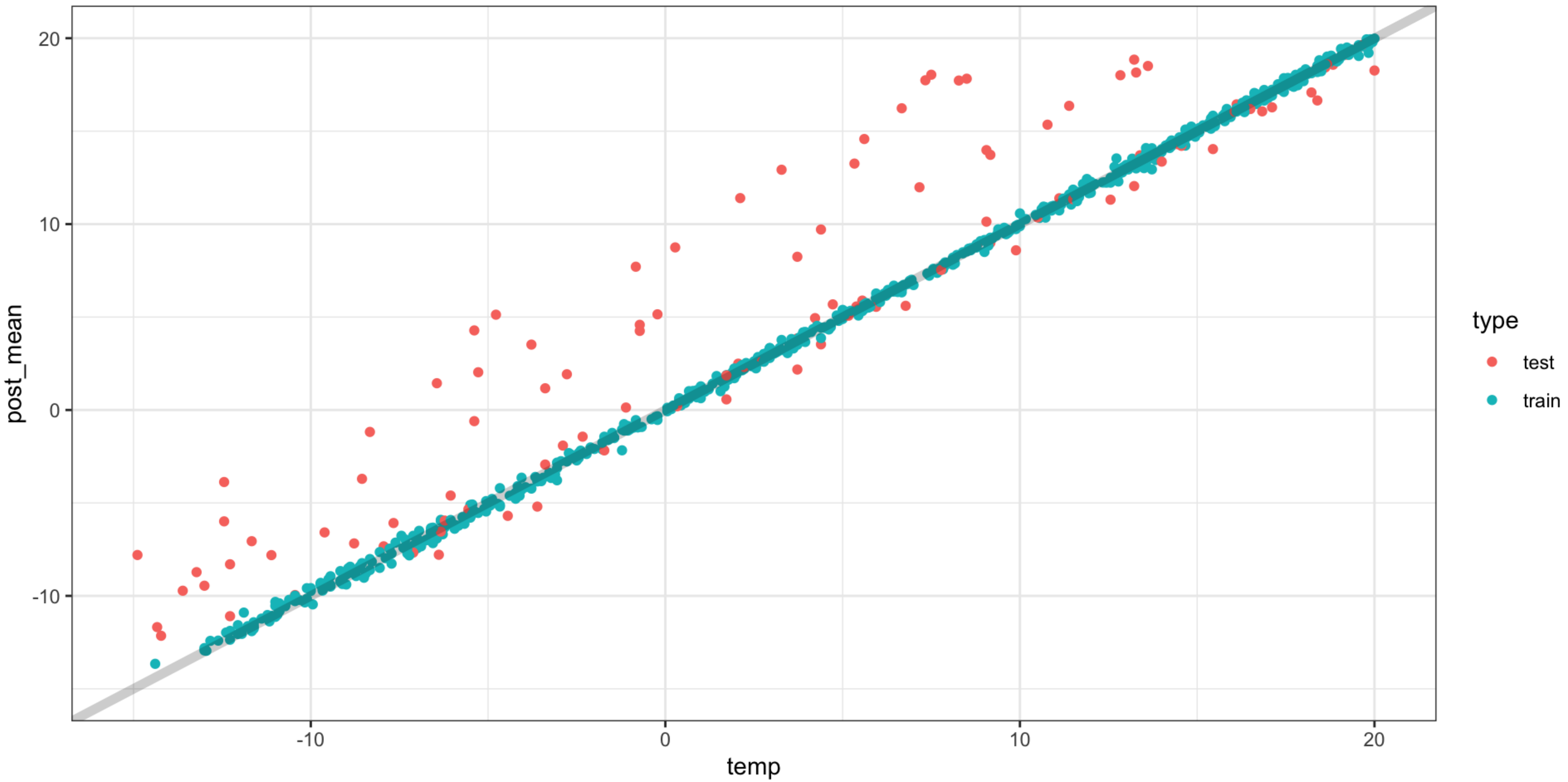


Predictive surfaces



Out-of-sample validation

```
# A tibble: 34 × 29
        x     y  elev type  station   t_1  t_10   t_11   t_12   t_13
    <dbl> <dbl> <int> <chr>   <int> <dbl> <dbl>  <dbl>  <dbl>  <dbl>
 1   6.09  3.20   102 test        1  NA    NA    NA      NA     NA
 2   6.25  3.26     1 train       2  -6.28  8.89  3.89   -4.22  -7.11
 3   6.16  3.48   157 train       3 -11.1   6.44  1.94   -8.72 -11.6
 4   6.12  3.53   176 train       4 -11.6   5.94  1.67   -9.17 -11.8
 5   6.00  3.28   400 train       5 -12.6   5.67  0.278 -10.7  -11.9
 6   6.05  3.23   133 train       6  -9.11  7.56  2.44   -7.11  -9.44
 7   6.10  3.18    56 test        7  NA    NA    NA      NA     NA
 8   6.07  3.14    59 train       8  -6.56  9.61  4.17   -4.89  -6.06
 9   6.17  3.46   160 train       9  -9.94  6.67  1.72   -8.44 -12.1
```



post_mean
temp
type
test
train
20
10
0
-10
-10
0
10
20

Spatio-temporal models for continuous time

Additive Models

In general, spatiotemporal models will have a form like the following,

$$
\begin{aligned}
y(\boldsymbol{s}, t) &= \underset{\text{mean structure}}{\mu(\boldsymbol{s}, t)} + \underset{\text{error structure}}{e(\boldsymbol{s}, t)} \\
&= \underset{\text{Regression}}{\boldsymbol{x}(\boldsymbol{s}, t)\,\boldsymbol{\beta}(\boldsymbol{s}, t)} + \underset{\text{Spatiotemporal RE}}{w(\boldsymbol{s}, t)} + \underset{\text{White Noise}}{\epsilon(\boldsymbol{s}, t)}
\end{aligned}
$$

The simplest possible spatiotemporal model is one where we assume there is no dependence between observations in space and time,

$$
w(\boldsymbol{s}, t) = \alpha(t) + \omega(\boldsymbol{s})
$$

these are straight forward to fit and interpret but are quite limiting (no shared information between space and time).

Spatiotemporal Covariance

Lets assume that we want to define our spatiotemporal random effect to be a single stationary Gaussian Process (in 3 dimensions*),

$$\boldsymbol{w}(\boldsymbol{s},\boldsymbol{t}) \sim \quad \left(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{s},\boldsymbol{t})\right)$$

where our covariance function depends on both $\|\mathrm{s}-\mathrm{s}'\|$ and $|\mathrm{t}-\mathrm{t}'|$,

$$\mathrm{cov}(\boldsymbol{w}(\boldsymbol{s},\boldsymbol{t}), \boldsymbol{w}(\boldsymbol{s}',\boldsymbol{t}')) = \mathrm{c}(\|\mathrm{s}-\mathrm{s}'\|, |\mathrm{t}-\mathrm{t}'|)$$

- Note that the resulting covariance matrix Σ will be of size $\mathrm{n_s} \cdot \mathrm{n_t} \times \mathrm{n_s} \cdot \mathrm{n_t}$.
  - Even for modest problems this gets very large (past the point of direct computability).
  - If $\mathrm{n_t} = 52$ and $\mathrm{n_s} = 100$ we have to work with a 5200×5200 covariance matrix

Separable Models

One solution is to use a seperable form, where the covariance is the product of a valid 2d spatial and a valid 1d temporal covariance / correlation function,

$$\text{cov}(\boldsymbol{w}(\boldsymbol{s}, \boldsymbol{t}), \boldsymbol{w}(\boldsymbol{s}', \boldsymbol{t}')) = \sigma^2 \, \rho_1(\|\boldsymbol{s} - \boldsymbol{s}'\|; \boldsymbol{\theta}) \, \rho_2(|\boldsymbol{t} - \boldsymbol{t}'|; \boldsymbol{\phi})$$

If we define our observations as follows (stacking time locations within spatial locations)

$$\boldsymbol{w}(\boldsymbol{s}, \boldsymbol{t}) = \Big(\mathrm{w}(\boldsymbol{s}_1, \mathrm{t}_1), \; \cdots, \; \mathrm{w}(\boldsymbol{s}_1, \mathrm{t}_{\mathrm{n_t}}), \; \cdots, \; \mathrm{w}(\boldsymbol{s}_{\mathrm{n_s}}, \mathrm{t}_1), \; \cdots, \; \mathrm{w}(\boldsymbol{s}_{\mathrm{n_s}}, \mathrm{t}_{\mathrm{n_t}})\Big)^t$$

then the covariance can be written as

$$\underset{\mathrm{n_s n_t} \times \mathrm{n_s n_t}}{\boldsymbol{\Sigma}_{\mathrm{w}}(\sigma^2, \theta, \phi)} = \sigma^2 \underset{\mathrm{n_s} \times \mathrm{n_s}}{\boldsymbol{H}_{\mathrm{s}}(\theta)} \otimes \underset{\mathrm{n_t} \times \mathrm{n_t}}{\boldsymbol{H}_{\mathrm{t}}(\phi)}$$

where $\boldsymbol{H}_{\mathrm{s}}(\theta)$ and $\boldsymbol{H}_{\mathrm{t}}(\theta)$ are correlation matrices defined by

$$\{\boldsymbol{H}_{\mathrm{s}}(\theta)\}_{\mathrm{ij}} = \rho_1(\|\boldsymbol{s}_{\mathrm{i}} - \boldsymbol{s}_{\mathrm{j}}\|; \theta)$$
$$\{\boldsymbol{H}_{\mathrm{t}}(\phi)\}_{\mathrm{ij}} = \rho_2(|\mathrm{t}_{\mathrm{i}} - \mathrm{t}_{\mathrm{j}}|; \phi)$$

Kronecker Product

Definition:

$$\underset{[m\times n]}{\boldsymbol{A}} \otimes \underset{[p\times q]}{\boldsymbol{B}} = \underset{[m\cdot p\times n\cdot q]}{\begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix}}$$

Properties:

$$\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A} \qquad \text{(usually)}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^t = \boldsymbol{A}^t \otimes \boldsymbol{B}^t$$

$$\begin{aligned} \det(\boldsymbol{A} \otimes \boldsymbol{B}) &= \det(\boldsymbol{B} \otimes \boldsymbol{A}) \\ &= \det(\boldsymbol{A})^{\text{rank}(\boldsymbol{B})} \det(\boldsymbol{B})^{\text{rank}(\boldsymbol{A})} \end{aligned}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$$

Kronecker Product and MVN Likelihoods

If we have a spatiotemporal random effect with a separable form,

$$\boldsymbol{w}(\boldsymbol{s},\boldsymbol{t}) \sim \quad (\boldsymbol{0},\ \boldsymbol{\Sigma}_{\mathrm{w}})$$

$$\boldsymbol{\Sigma}_{\mathrm{w}} = \sigma^2\, \boldsymbol{H}_{\mathrm{s}} \otimes \boldsymbol{H}_{\mathrm{t}}$$

then the likelihood for $\boldsymbol{w}$ is given by

$$\begin{aligned}
&-\frac{\mathrm{n}}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{w}}| - \frac{1}{2}\boldsymbol{w}^{\mathrm{t}}\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}\boldsymbol{w} \\
&= -\frac{\mathrm{n}}{2}\log 2\pi - \frac{1}{2}\log\left[(\sigma^2)^{\mathrm{n_t}\cdot\mathrm{n_s}}\,|\boldsymbol{H}_{\boldsymbol{s}}|^{\mathrm{n_t}}\,|\boldsymbol{H}_{\boldsymbol{t}}|^{\mathrm{n_s}}\right] - \frac{1}{2\sigma^2}\boldsymbol{w}^{\mathrm{t}}(\boldsymbol{H}_{\mathrm{s}}^{-1} \otimes \boldsymbol{H}_{\mathrm{t}}^{-1})\boldsymbol{w}
\end{aligned}$$

Non-seperable Models

- Additive and separable models are still somewhat limiting
- Cannot treat spatiotemporal covariances as 3d observations
- Possible alternatives:
  - Specialized spatiotemporal covariance functions, i.e.

$$\gamma(\boldsymbol{s}, \boldsymbol{s}', t, t') = \sigma^2(|t - t'| + 1)^{-1} \exp\Big(- \|\boldsymbol{s} - \boldsymbol{s}'\|(|t - t'| + 1)^{-\beta/2} \Big)$$

```
* Mixtures of separable covariances, i.e.
```

$$w(\boldsymbol{s}, t) = w_1(\boldsymbol{s}, t) + w_2(\boldsymbol{s}, t)$$