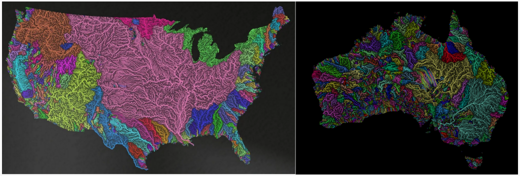
Time and tide wait for no one: spatio-temporal modelling in river networks

Edgar Santos-Fernández, PhD (santosfe@qut.edu.au)

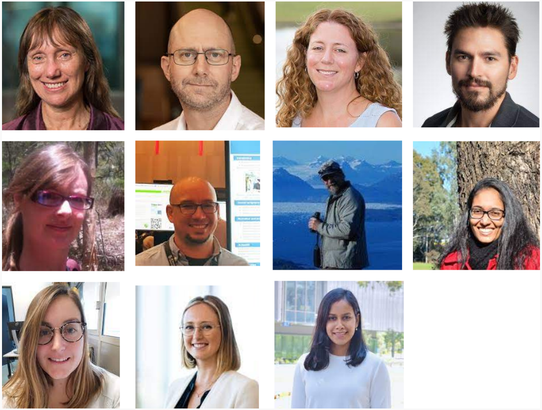
Jay M. Ver Hoef, Erin E. Peterson, James McGree, Daniel J. Isaak and Kerrie Mengersen

Queensland University of Technology

31 Oct 2022. StanConnect 2022: Stan through Space and Time



ARC Linkage Project: Revolutionising water-quality monitoring in the information age



QUT

MONASH University

Queensland Government
Department of Environment and Science

Healthy land & water

New sources of data 3: water quality monitoring

Computational Statistics and Data Analysis 170 (2022) 107446

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Bayesian spatio-temporal models for stream networks

Edgar Santos-Fernandez [a,b,*], Jay M. Ver Hoef [d], Erin E. Peterson [a,b,c], James McGree [a], Daniel J. Isaak [e], Kerrie Mengersen [a,b]

[a] *School of Mathematical Sciences, Queensland University of Technology, Australia*
[b] *Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), Australia*
[c] *Erin Peterson Consulting, Australia*
[d] *Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center, United States of America*
[e] *Rocky Mountain Research Station, US Forest Service, United States of America*

SSNbayes: An R package for Bayesian spatio-temporal modelling on stream networks

by Edgar Santos-Fernandez, Jay M. Ver Hoef, James M. McGree, Daniel J. Isaak, Kerrie Mengersen and Erin E. Peterson

Abstract Spatio-temporal models are widely used in many research areas from ecology to epidemiology. However, most covariance functions describe spatial relationships based on Euclidean distance only. In this paper, we introduce the R package **SSNbayes** for fitting Bayesian spatio-temporal models and making predictions on branching stream networks. **SSNbayes** provides a linear regression framework with multiple options for incorporating spatial and temporal autocorrelation. Spatial dependence is captured using stream distance and flow connectivity while temporal autocorrelation is modelled using vector autoregression approaches. **SSNbayes** provides the functionality to make predictions across the whole network, compute exceedance probabilities and other probabilistic estimates such as the proportion of suitable habitat. We illustrate the functionality of the package using a stream temperature dataset collected in Idaho, USA.

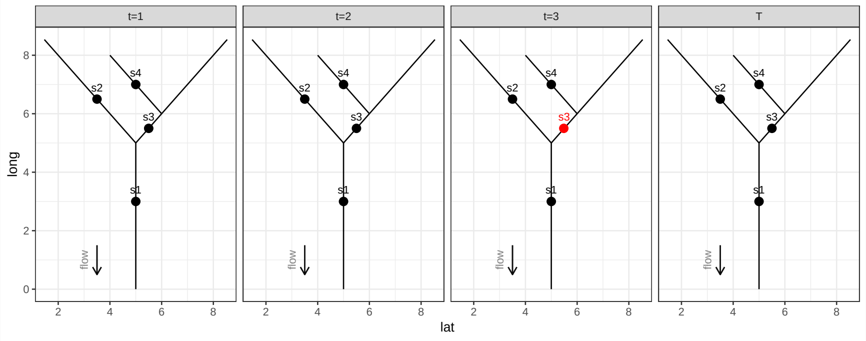
Background

Stream networks

- Increasing number of spatial locations.
- Observations are highly correlated.
- Spatial dependence determined by flow connectivity, stream distance, etc.
- Generally, traditional geostatistical models do not perform well.

Background

Repeated measures in a stream network



Bayesian space-time model (Santos-Fernandez et al., 2022)

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{v} + \boldsymbol{\epsilon} \tag{1}$$

$$[\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \beta, \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}] = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma} + \sigma_0^2 \boldsymbol{I}) \tag{2}$$

$$\boldsymbol{\mu}_t = \boldsymbol{X}_t\beta + \boldsymbol{\Phi}_1(\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1}\beta) \tag{3}$$

$$\boldsymbol{\Sigma} = COV(\boldsymbol{v}) = \boldsymbol{C}_{ED} + \boldsymbol{C}_{TU} + \boldsymbol{C}_{TD} \tag{4}$$

The autoregression or transition matrix $\boldsymbol{\Phi}$ is

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1S} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2S} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{S1} & \phi_{S2} & \cdots & \phi_{SS} \end{bmatrix}$$

The temporal part

Case 1 (AR)

$$\Phi_1 = \begin{bmatrix} \phi & 0 & 0 & 0 \\ 0 & \phi & 0 & 0 \\ 0 & 0 & \phi & 0 \\ 0 & 0 & 0 & \phi \end{bmatrix} \tag{5}$$

Case 2 (VAR) (a) and (b) $\mathbf{logit}(\phi_s) = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2$

$$\Phi_1 = \begin{bmatrix} \phi_1 & 0 & 0 & 0 \\ 0 & \phi_2 & 0 & 0 \\ 0 & 0 & \phi_3 & 0 \\ 0 & 0 & 0 & \phi_4 \end{bmatrix} \tag{6}$$

Case 3 (VAR2NN)

$$\Phi_1 = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} & 0 \\ \phi_{31} & 0 & \phi_{33} & \phi_{34} \\ \phi_{41} & 0 & \phi_{43} & \phi_{44} \end{bmatrix} \tag{7}$$

Incorporating spatial autocorrelation (Ver Hoef & Peterson, 2010; Peterson & Hoef, 2010)

Tail-up models (TU)

1- exponential:

$$C_u(h \mid \boldsymbol{\theta}) = \sigma_u^2 e^{-3h/\alpha_u}$$

2- linear-with-sill model:

$$C_u(h \mid \boldsymbol{\theta}) = \sigma_u^2 (1 - h/\alpha_u) \mathbb{1}(h/\alpha_u \leqslant 1)$$

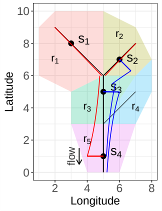
Tail-down models (TD)

1- exponential:

$$C_{TD}(a, b, h|\boldsymbol{\theta}) = \begin{cases} \sigma_d^2 e^{-3h/\alpha_d} & \text{if flow-connected,} \\ \sigma_d^2 e^{-3(a+b)/\alpha_d} & \text{if flow-unconnected,} \end{cases}$$

2- linear-with-sill model:

$$C_{TD}(a, b, h|\boldsymbol{\theta}) = \begin{cases} \sigma_d^2 (1 - \frac{h}{\alpha_d}) \mathbb{1}(\frac{h}{\alpha_d} \leqslant 1) & \text{if flow-connected,} \\ \sigma_d^2 (1 - \frac{b}{\alpha_d}) \mathbb{1}(\frac{b}{\alpha_d} \leqslant 1) & \text{if flow-unconnected,} \end{cases}$$



Hierarchical model and prior distributions:

$$
\begin{aligned}
[\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_T] &= \prod_{t=2}^{T} [\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \beta, \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}][\boldsymbol{y}_1] \\
[\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \beta, \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}] &= \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma} + \sigma_0^2 \boldsymbol{I}) \\
\boldsymbol{\mu}_t &= \boldsymbol{X}_t \beta + \boldsymbol{\Phi}_1 (\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1} \beta) \\
\boldsymbol{\Sigma} &= \sigma_u^2 \boldsymbol{R}(\alpha_u) + \sigma_d^2 \boldsymbol{R}(\alpha_d) + \sigma_e^2 \boldsymbol{R}(\alpha_e) \\
\text{Priors} & \\
\beta_0, \beta_1, \beta_2, \cdots, \beta_p &\sim \mathcal{N}(0, 100) \\
\sigma_0 &\sim \text{Uniform}(0, 50) \\
\sigma_u, \sigma_d, \sigma_e &\sim \text{Uniform}(0, 100) \\
\alpha_u, \alpha_d, \alpha_e &\sim \text{Uniform}(0, \alpha_{max}) \\
\alpha_{max} &= 4 \max(H) \\
\text{Elements of } \boldsymbol{\Phi}_1: & \\
\text{Case 1} & \\
\phi &\sim \text{Uniform}(-1, 1) \\
\text{Case 2a} & \\
\phi_s &\sim \mathcal{N}(0.5, 0.2) T[-1, 1] \\
\text{Case 2b} & \\
\text{logit}(\phi_s) &= \gamma_0 + \gamma_1 X_{1s} + \gamma_2 X_{2s} + \cdots + \gamma_J X_{JS} \\
\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_J &\sim \mathcal{N}(0, 100) \\
\text{Case 3} & \\
\phi_s, \phi_{sr} &\sim \text{Uniform}(-1, 1)
\end{aligned}
$$

SSNbayes R package

Based on Stan

remotes::install_github("EdgarSantos-Fernandez/SSNbayes", dependencies = T)

- Visualize stream network data in space and time.
- Fitting spatio-temporal linear models.
- Making predictions.

https://www.kaggle.com/edsans/ssnbayes-tutorial

SSNbayes computation

Method 1: Separable space-time model

$$[\boldsymbol{y}_t \mid \boldsymbol{\theta}, \boldsymbol{X}_t, \beta, \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}] = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma} + \sigma_0^2 \boldsymbol{I})$$

Separable full covariance matrix:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_S^{-1} \otimes \boldsymbol{\Sigma}_{var}^{-1}$$

$\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_{var}$: spatial and temporal covariance matrices respectivelly.

Method 2: Vector autoregression spatial model (Implemented in SSNbayes)

$$[\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{\theta}, \boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \beta, \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}] = \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma} + \sigma_0^2 \boldsymbol{I}) \tag{8}$$

$$\boldsymbol{\mu}_t = \boldsymbol{X}_t \beta + \boldsymbol{\Phi}_1 (\boldsymbol{y}_{t-1} - \boldsymbol{X}_{t-1} \beta) \tag{9}$$

SSNbayes functions

To fit a spatio-temporal stream network model:

```
fit <- ssnbayes(formula = temp ~ x1 + x2 + x3,
                data = df,
                path = path,
                space_method = list("use_ssn",
                                    "Exponential.taildown"),
                time_method = list("var", "date"),
                iter = 4000,
                warmup = 2000,
                chains = 3)
```

SSNbayes functions

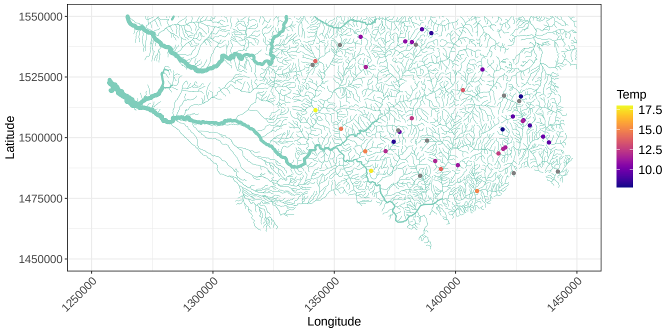
... and to make predictions using simple kriging:

$$\widehat{\boldsymbol{y}}_P = \boldsymbol{X}_P\beta + c_{OP}{}' C_{OO}^{-1}(\boldsymbol{y}_O - \boldsymbol{X}_O\beta), \tag{10}$$

```
pred <- predict(path = path,
                      obs_data = df,
                      stanfit = fit,
                      pred_data = df2,
                      nsamples = 4000,
                      chunk_size = 200)
```

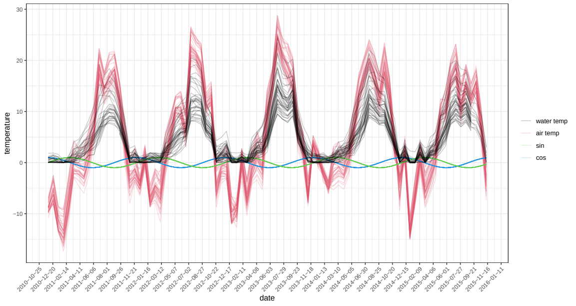
Case study

- mean daily water temperature in the Boise River Basin located in the Northwestern part of the United States.
- it includes 7,364 km of streams.
- 42 in-situ sensors



Covariates

- spatial: stream slope, elevation, cumulative drainage area
- temporal: harmonic covariates (sin and cos)
- spatio-temporal: air temperature



- time series of 5 years (2010-12-01 - 2015-12-01)
- out-of-sample prediction
- split: 80% for training and 20% for testing
- testing set (grey) and missing values(white)

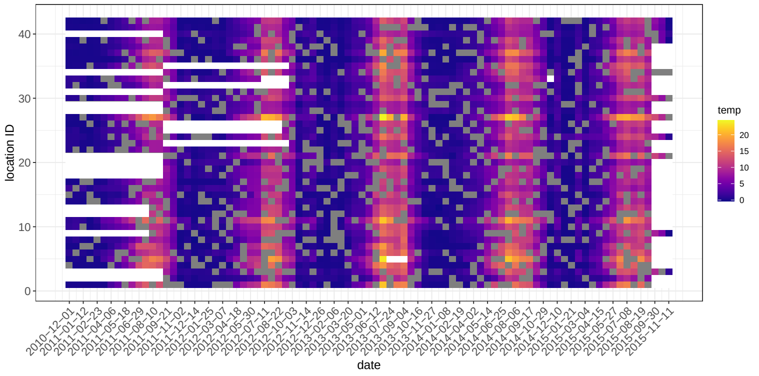
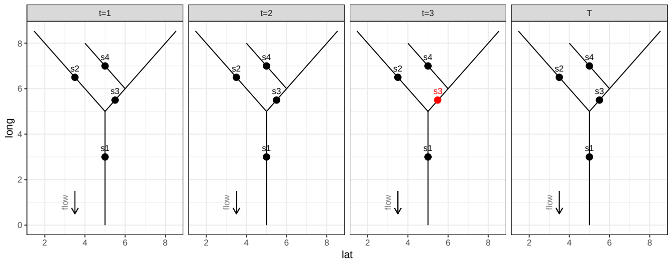


Table: Model combinations used in the case study.

	Spatial structure					
Temporal structure	Tail-Down	Tail-up	Euc dist	Tail-up/Tail-Down	Tail-up/Euc dist	Tail-down/Euc dist
AR (Case 1)	td_AR	tu_AR	ed_AR	tutd_AR	tued_AR	tded_AR
VAR(1) (Case 2a)	td_VAR	tu_VAR	ed_VAR	tutd_VAR	tued_VAR	tded_VAR
VAR(1) (Case 2b)	td_VAR_2b	tu_VAR_2b	ed_VAR_2b	tutd_VAR_2b	tued_VAR_2b	tded_VAR_2b
VAR(1) NN (Case 3)	td_VAR_2NN	tu_VAR_2NN	ed_VAR_2NN	tutd_VAR_2NN	tued_VAR_2NN	tded_VAR_2NN

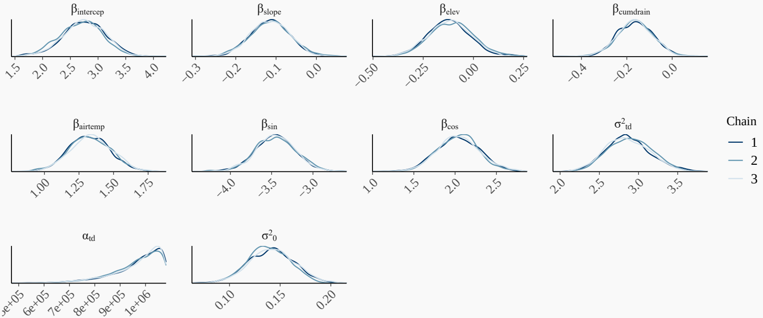


Results

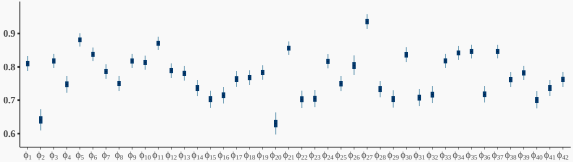
model	Prediction accuracy/Estimation						Comp	Qualitative factors		
	WAIC*	LOO*	CRPS*	RM-SPE*	SE rank*	cover	time*	ident	com-plex	interp
td_AR	9098	9170	0.289	0.549	2	0.942	hig	good	low	good
tu_AR	11874	11934	0.439	0.844	1	0.969	hig	good	low	good
ed_AR	10204	10355	0.314	0.576	5	0.963	low	good	low	good
tutd_AR	8840	8901	0.277	0.53	4	0.947	hig	mod	low	mod
tued_AR	64540	15371	0.602	1.241	13	0.985	low	mod	low	mod
tded_AR	9024	9210	0.299	0.565	8	0.947	med	mod	low	mod
td_VAR	10013	9518	0.286	0.54	15	0.957	hig	good	low	good
tu_VAR	11890	11959	0.447	0.854	2	0.965	hig	good	low	good
ed_VAR	27476	14838	0.304	2.446	19	0.977	low	good	low	good
tutd_VAR	73033	14225	0.52	1.109	16	0.971	low	mod	low	mod
tued_VAR	19829	13196	0.352	0.676	10	0.968	med	mod	low	mod
tded_VAR	10373	10145	0.28	0.527	14	0.965	med	mod	low	mod
td_VAR_2b	8832	8950	0.285	0.55	6	0.945	med	good	mod	good
tu_VAR_2b	11870	11997	0.439	0.843	4	0.968	hig	good	mod	good
ed_VAR_2b	10000	9977	0.309	0.567	17	0.957	med	good	mod	good
tutd_VAR_2b	8290	8399	0.278	0.537	7	0.939	med	mod	mod	good
tued_VAR_2b	9248	9132	0.282	0.534	17	0.96	low	mod	mod	good
tded_VAR_2b	9248	9132	0.281	0.534	18	0.942	low	mod	mod	good
td_VAR_2NN	8755	8914	0.282	0.518	5	0.933	hig	mod	mod	mod
tu_VAR_2NN	11426	11478	0.462	0.886	3	0.948	hig	mod	mod	mod
ed_VAR_2NN	11295	10364	0.325	0.595	9	0.957	low	mod	mod	mod
tutd_VAR_2NN	13212	9485	0.305	0.584	12	0.928	med	mod	mod	mod
tued_VAR_2NN	>90000	78866	0.28	2.593	20	0.937	low	mod	mod	mod
tded_VAR_2NN	82104	14774	0.615	1.327	11	0.969	med	mod	mod	mod

Results from the td_VAR_2b model

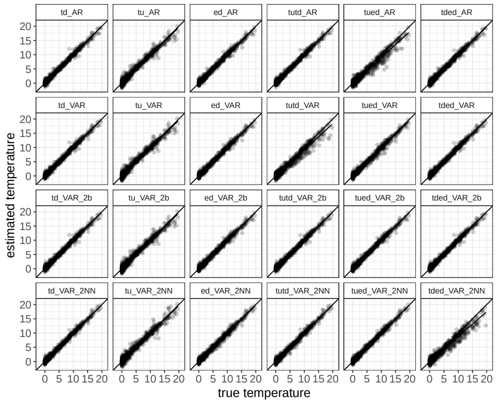
(a) Posterior distributions of β, σ^2_{td} and α_{td}



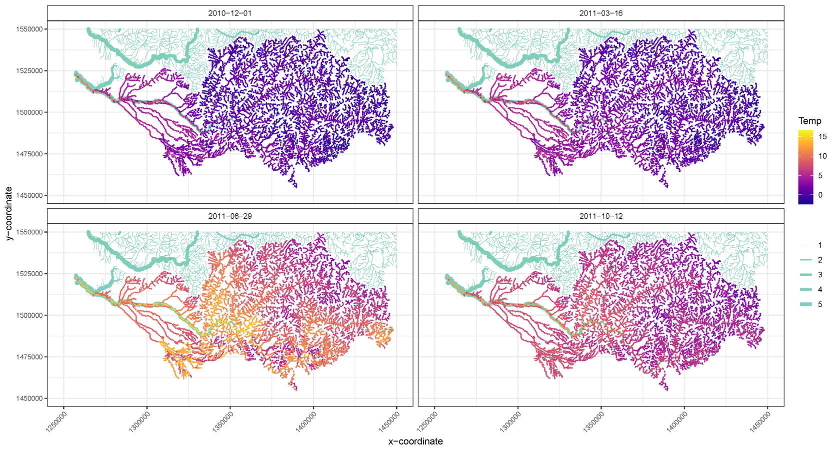
(b) Posterior boxplots of $\boldsymbol{\Phi}$

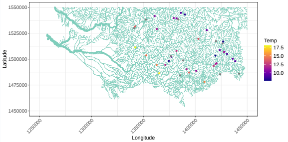


How good are the predictions?



Predictions on the rest of the network





Exceedance probability

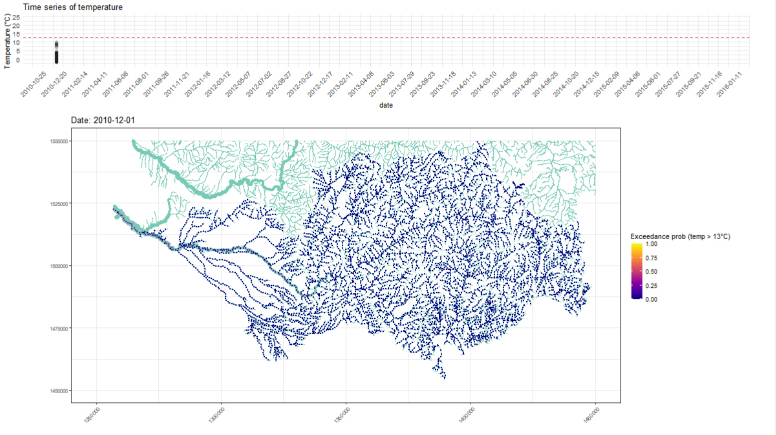
- Bull trout is a native threatened species.
- Temperature biological limit for juvenile = 13 °C

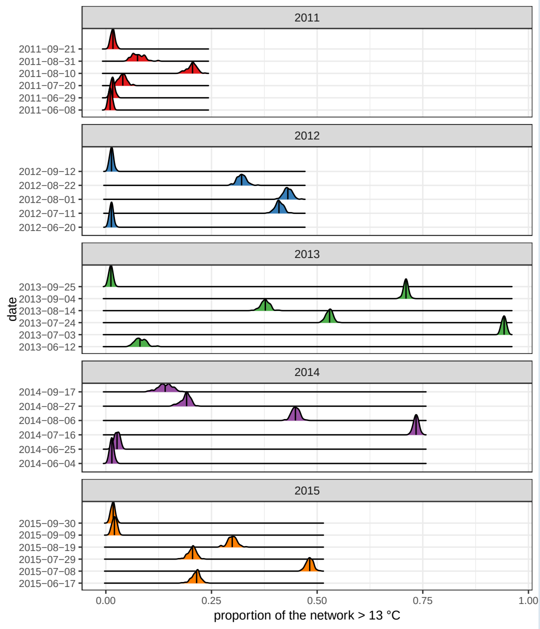
Interventions

(a) Riparian restoration. [Source: Wikipedia]

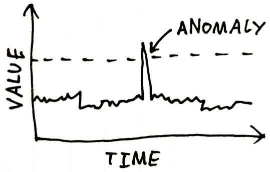
(b) Use of beavers. [Source: Getty]

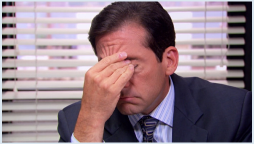
Exceedance probability




2011
2011−09−21
2011−08−31
2011−08−10
2011−07−20
2011−06−29
2011−06−08
2012
2012−09−12
2012−08−22
2012−08−01
2012−07−11
2012−06−20
2013
2013−09−25
2013−09−04
2013−08−14
2013−07−24
2013−07−03
2013−06−12
2014
2014−09−17
2014−08−27
2014−08−06
2014−07−16
2014−06−25
2014−06−04
2015
2015−09−30
2015−09−09
2015−08−19
2015−07−29
2015−07−08
2015−06−17
date
0.00
0.25
0.50
0.75
1.00
proportion of the network > 13 °C

Other applications and methods: Anomaly detection in spatio-temporal stream network data





- Anomalies: battery failure, sensor outside of the water, calibration issues, biofouling on optical sensors

Conclusions

- Fitting space-time models in Stan produce suitable estimates and predictions.
- Tend to outperform time series approaches existing in the literature.

Other/Future work

- Space-time anomaly detection in real time.
- Disentangle anomalies from water quality events.

Thanks

Acknowledgement: Collaborators: Jay Ver Hoef (NOAA Federal), Erin Peterson, Daniel Isaak (US Forest Service), James McGree (QUT), Department of Environment and Science, Queensland, et al.
Australian Research Council (ARC) Linkage Projects. "Revolutionising water-quality monitoring in the information age".
Other collaborators: Puwasala Gamakumara, Katie Buchhorn, Rob Hyndman, Catherine Leigh, et al.
Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).

Credits: Front image: `https://www.grasshoppergeography.com`
Video: Katie Buchhorn.

References

Peterson, E. E., & Hoef, J. M. V. (2010). A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*, *91*, 644–651.

Santos-Fernandez, E., Ver Hoef, J. M., Peterson, E. E., McGree, J., Isaak, D. J., & Mengersen, K. (2022). Bayesian spatio-temporal models for stream networks. *Computational Statistics & Data Analysis*, *170*, 107446. URL: https://www.sciencedirect.com/science/article/pii/S0167947322000263. doi:doi: https://doi.org/10.1016/j.csda.2022.107446.

Ver Hoef, J. M., & Peterson, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, *105*, 6–18.

How good are the predictions?