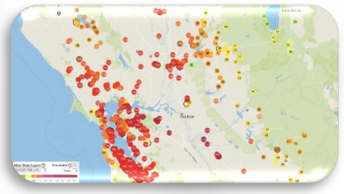
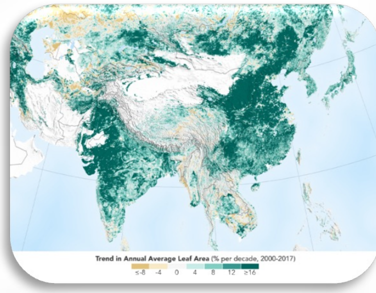
Latent **Meshed Gaussian Processes** for Scalable Bayesian Regression

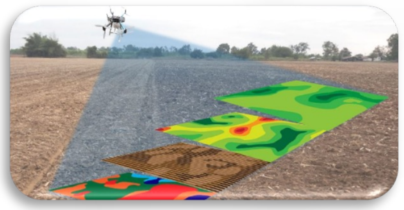
Michele Peruzzi
michele.peruzzi@duke.edu
github.com/mkln

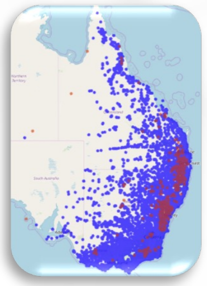
with
David Dunson
Sudipto Banerjee
Andrew Finley

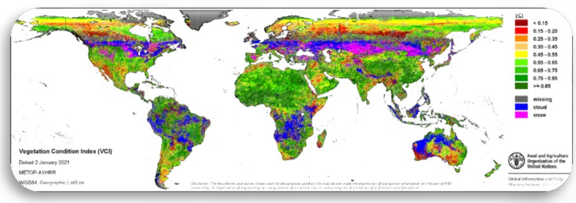
References
Meshed GPs (2020). JASA in press
doi:10.1080/01621459.2020.1833889
Better MGPs via GriPS (2021)
arXiv:2101.03579
Spamtrees (2021). JMLR
arxiv.org/abs/2012.00943











Crowd-sourced
Satellites
Remote sensing
SPATIAL BIG DATA
Multivariate
Misaligned
Multi-type
Huge dimension

R PACKAGE meshed

```r
meshout <- meshed::spmeshed(
  y = Y,
  x = X,
  coords = coords,
  family = c("poisson",
             "gaussian",
             "binomial"),
  k = 2,
  grid_size = c(30, 30),
  block_size = 20,
  n_samples = 5000,
  n_burn = 2000,
  n_thin = 5,
  n_threads = 4,
  prior = list(phi=c(2, 20))
```

Y matrix of outcomes
q columns
OK with NA (will predict)

X covariates

coords spatial coordinates
OK with spacetime

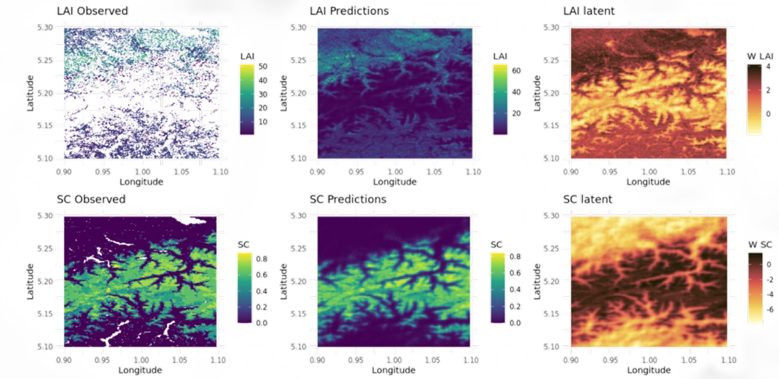
family list outcome types
OK with multi-type

k number of spatial factors
linear coregionalization: OK k<q

grid_size to build $\mathcal{S}$.
(optional. spmeshed will try to figure out what's best for speed)

parallel computing with **OpenMP**
...if meshed was compiled with OpenMP support

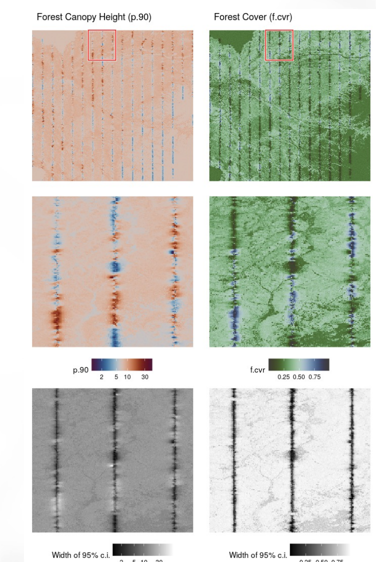
works best with R linked to **OpenBLAS/Intel MKL**



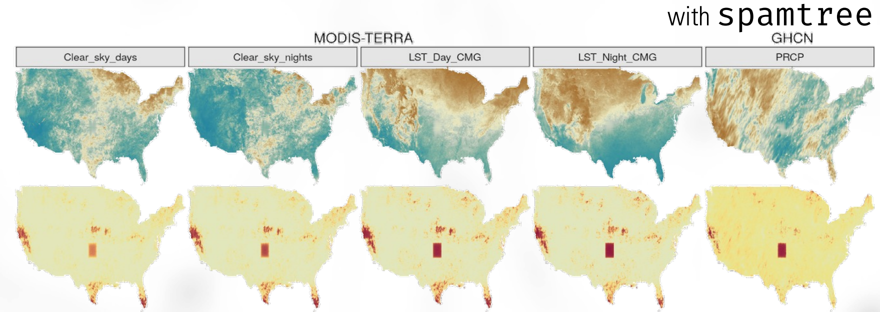
count and discrete outcomes

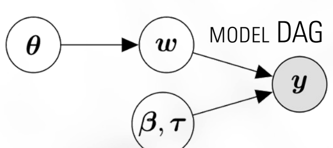
multivariate big data
with meshed

continuous outcomes
modeled as Gaussians



with spamtree

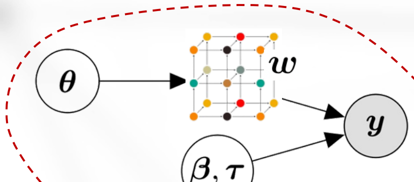




flexible Bayesian
hierarchical models for
MULTIVARIATE OUTCOMES

SCALABLE MCMC FOR MGPs

- Gridded $\mathcal{S}$
- Parameter expansion & parametrization (GriPS)
- Gibbs sampler for Gaussian outcomes
- Meshed Langevin on Riemann Manif. otherwise (MELANGE)



CUBIC MESHED GP
R PACKAGE meshed

modeling
Gaussian outcomes

$$\underbrace{y_j(\boldsymbol{\ell})}_{j\text{th outcome}} = \underbrace{\boldsymbol{x}_j(\boldsymbol{\ell})^\top \boldsymbol{\beta}_j}_{\text{covariates}} + \underbrace{w_j(\boldsymbol{\ell})}_{\text{latent effects}} + \varepsilon_j(\boldsymbol{\ell})$$

Gaussian measurement error $\varepsilon(\boldsymbol{\ell}) \sim N(0, \tau_j^2)$

characterize
spatial associations
across outcomes
across locations

multivariate GP with cross-covariance $\boldsymbol{C}_{\boldsymbol{\theta}}$

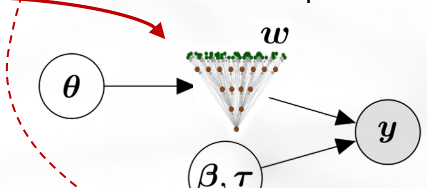
$$\boldsymbol{w}(\cdot) \sim GP(\boldsymbol{0}, \underbrace{\boldsymbol{C}_{\boldsymbol{\theta}}}_{\text{dimension } (nq,\, nq)})$$

does not scale
REPLACE WITH

MESHED GP

$$\boldsymbol{w}(\cdot) \sim MGP_{\mathcal{S},\mathcal{G}}(\boldsymbol{0}, \boldsymbol{C}_{\boldsymbol{\theta}})$$

SPATIAL MULTIVARIATE TREES
R PACKAGE spamtree



DIY MESHED SPATIAL PROCESS

- reference set $\mathcal{S}$ of knots
- fix a DAG with patterns
- partition of $\mathcal{S}$ linked to DAG nodes
- a rule to determine what happens at other locations

modeling
multi-type outcomes

$$\underbrace{y_j(\boldsymbol{\ell})}_{j\text{th outcome}} \mid \eta_j(\boldsymbol{\ell}), \tau_j \sim P_j(\eta_j(\boldsymbol{\ell}), \tau_j)$$

$$\underbrace{\eta_j(\boldsymbol{\ell})}_{\text{linear predictor}} = \underbrace{\boldsymbol{x}_j(\boldsymbol{\ell})^\top \boldsymbol{\beta}_j}_{\text{covariates}} + \underbrace{w_j(\boldsymbol{\ell})}_{\text{latent effects}}$$

outcome index $j = 1, \ldots, q$
spatial coordinate $\boldsymbol{\ell} \in \mathcal{D} \subset \Re^d$