University of Zurich[UZH]

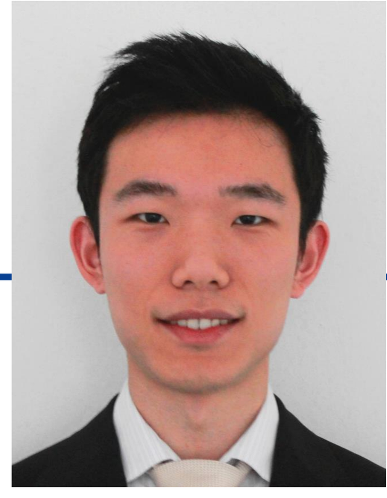
Flexible Spatial Fusion Model Framework

Joint Analysis of Point and Areal Data

Craig Wang[1], Reinhard Furrer[1,2]

[1]Department of Mathematics; [2]Department of Computational Science; University of Zurich, Switzerland

Contact: craig.wang@uzh.ch



Motivation

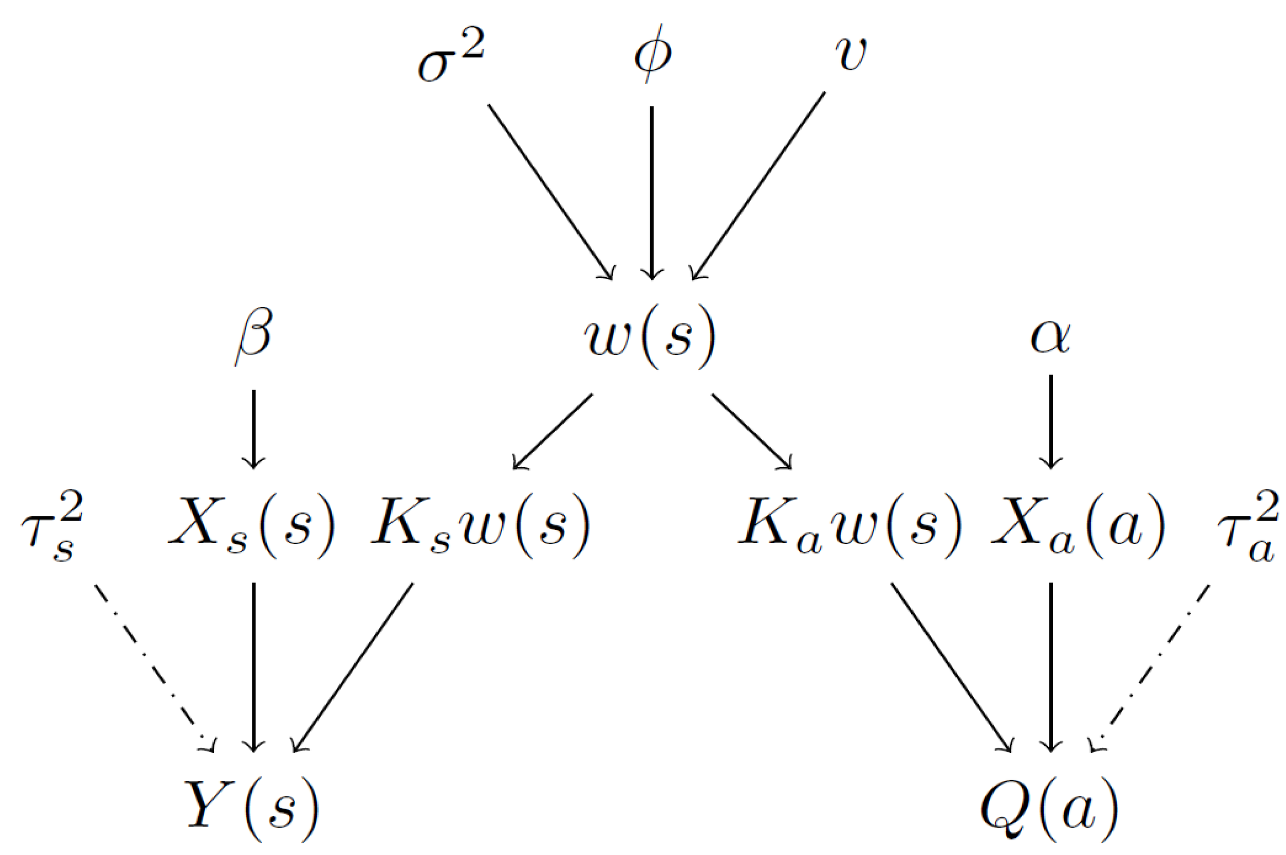
- What is the underlying spatial risk surface for lung diseases, given lung function measurements at individual level and cause-specific mortality at zip-code level?
- A single source of data may not be suitable for modelling due to data missing, measurement bias, and weakly identified model parameters.

Summary

We develop flexible spatial fusion model framework which incorporates data collected from different spatial support by assuming a common latent Gaussian process.

The framework allows different distribution assumptions from point and areal data, and they provide smaller prediction error without adding significant computational burden.

Model Formulation



Notation	Description
σ^2, ϕ, v	Partial sill, range and smoothness parameter
$w(s)$	Continuous latent spatial process
K_s, K_a	Design matrices
$X_s(s), X_a(a)$	Point and area covariates
β, α	Coefficient for point and areal covariates
τ_s^2, τ_a^2	Measurement errors[1]
$Y(s), Q(a)$	Point and areal response

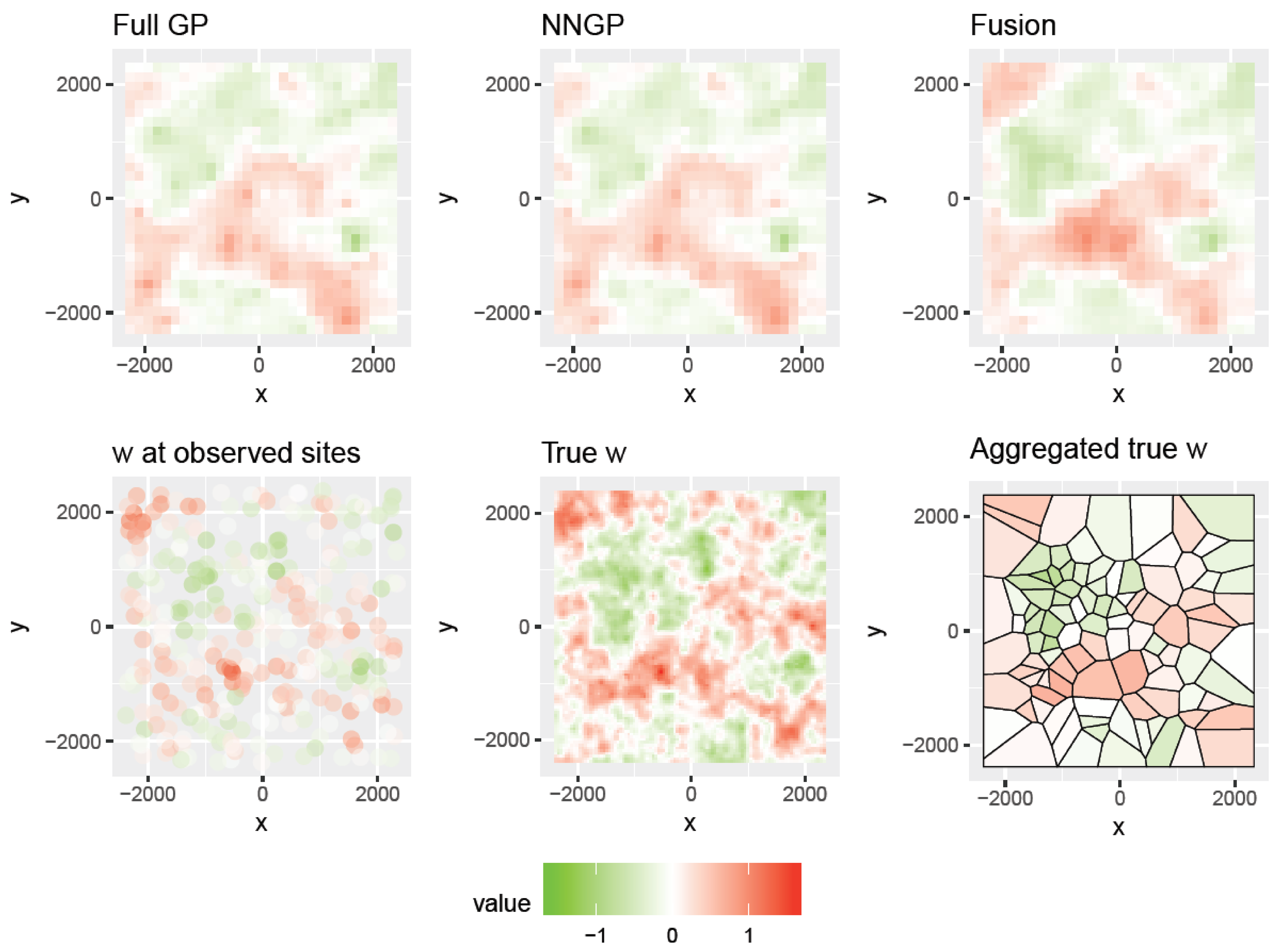
Model Implementation

- Nearest neighbor Gaussian process[2] (NNGP) with m = 5 neighbors
- Stan modelling language[3]
  - No-U-Turn sampler (NUTS)
  - Tuning: adapt_delta = 0.95, max_treedepth = 12
  - 4'000 iterations with 2'000 warm-up samples, 4 chains
- Approximation to stochastic integrals with 4 sampling points[4]

Simulation Study and Results

Simulation settings:

- observation: 300 sites and 100 areas in a square domain
- prediction: 900 sites
- $Y \sim N(X_s^T\beta + K_s w, \tau^2);\ Q \sim \text{Poisson}(X_a^T\alpha + K_a w);\ X_s, X_a \sim N(0,1);$
- exponential covariance: $v = 0.5$; effective range: 870
- priors: $\alpha, \beta \sim N(0, 5^2);\ \sigma^2, \tau^2 \sim IG(2,1);\ \phi \sim N(300, 100^2)$



Model	Full GP	NNGP	Fusion
$\sigma^2 = 0.2$	0.35 (0.18, 0.68)	0.34 (0.17, 0.64)	0.28 (0.17, 0.47)
$\tau^2 = 1.0$	0.95 (0.63, 1.20)	0.97 (0.73, 1.21)	1.02 (0.83, 1.24)
$\phi = 290$	250 (111, 460)	322 (152, 499)	389 (236, 553)
$\beta_0 = 1$	1.00 (0.79, 1.20)	0.99 (0.77, 1.20)	0.99 (0.77, 1.21)
$\beta_1 = 5$	5.08 (4.95, 5.20)	5.08 (4.95, 5.20)	5.06 (4.94, 5.19)
$\alpha_0 = 1$	–	–	0.93 (0.68, 1.17)
$\alpha_1 = 2$	–	–	2.01 (1.85, 2.18)
Predictive MSE	0.1814	0.1813	**0.1551**
95% CI Width	2.17	2.05	**1.70**
95% CI Coverage	98.6	97.7	**96.3**

Future Work

- Conduct further simulation studies to evaluate model performance under different scenarios
- Improve computational efficiency
- Apply spatial fusion model framework to an epidemiological dataset

References

[1] Banerjee, S., Carlin, B.P., Gelfand, A.E., (2014). p.136-139, Hierarchical modeling and analysis for spatial data. CRC Press.
[2] Datta, A., Banerjee S., Finley A.O., and Gelfand A.E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. Journal of the American Statistical Association 111 (514),
[3] Carpenter, B., Gelman, A., Homan, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., (2017). Stan: A probabilistic programming language. Journal of Statistical Software 76.
[4] Fuentes, M., Raftery, A.E., (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 61, 36 – 45.