

Supplementary Material: Spatially Aggregated Gaussian Processes with Multivariate Areal Outputs

A Derivation of the multivariate GP $f(\mathbf{x})$

In this appendix, we show that the process $f(\mathbf{x})$ defined via (2) is itself a multivariate GP with mean function $\mathbf{m}(\mathbf{x}) = \mathbf{W}\boldsymbol{\nu}(\mathbf{x})$ and covariance function $\mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathbf{W}\boldsymbol{\Gamma}(\mathbf{x}, \mathbf{x}')\mathbf{W}^\top + \boldsymbol{\Lambda}(\mathbf{x}, \mathbf{x}')$. To prove that $f(\mathbf{x})$ is indeed a multivariate GP, one has only to show that, for an arbitrary $k \in \{1, 2, \dots\}$ and an arbitrary set of k points $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X}$, $\bar{\mathbf{f}} = (\mathbf{f}(\mathbf{x}_1), \dots, \mathbf{f}(\mathbf{x}_k))^\top$ is a multivariate Gaussian random variable. By the definition (2) of $f(\mathbf{x})$, one has

$$\bar{\mathbf{f}} = (\mathbf{I} \otimes \mathbf{W})\bar{\mathbf{g}} + \bar{\mathbf{n}}, \quad (19)$$

where we let $\bar{\mathbf{g}} = (\mathbf{g}(\mathbf{x}_1), \dots, \mathbf{g}(\mathbf{x}_k))^\top$ and $\bar{\mathbf{n}} = (\mathbf{n}(\mathbf{x}_1), \dots, \mathbf{n}(\mathbf{x}_k))^\top$, and where \otimes denotes the Kronecker product. By the definition of Gaussian processes, since $\mathbf{g}(\mathbf{x})$ and $\mathbf{n}(\mathbf{x})$ are Gaussian processes, $\bar{\mathbf{g}}$ and $\bar{\mathbf{n}}$ are multivariate Gaussian random variables. Since (19) shows that $\bar{\mathbf{f}}$ is a linear combination of the multivariate Gaussian random variables $\bar{\mathbf{g}}$ and $\bar{\mathbf{n}}$, it is itself multivariate Gaussian, irrespective of the choice of $\mathbf{x}_1, \dots, \mathbf{x}_k$. This in turn shows that $f(\mathbf{x})$ is again a multivariate Gaussian process.

Mean of $f(\mathbf{x}_i)$ is given by

$$\mathbb{E}(f(\mathbf{x}_i)) = \mathbf{W}\mathbb{E}(\mathbf{g}(\mathbf{x}_i)) = \mathbf{W}\boldsymbol{\nu}(\mathbf{x}_i). \quad (20)$$

Covariance of $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ is given by

$$\begin{aligned} \text{Cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) &= \mathbb{E}((f(\mathbf{x}_i) - \mathbf{W}\boldsymbol{\nu}(\mathbf{x}_i))(f(\mathbf{x}_j) - \mathbf{W}\boldsymbol{\nu}(\mathbf{x}_j))^\top) \\ &= \mathbb{E}((\mathbf{W}(\mathbf{g}(\mathbf{x}_i) - \boldsymbol{\nu}(\mathbf{x}_i)) + \mathbf{n}(\mathbf{x}_i))(\mathbf{W}(\mathbf{g}(\mathbf{x}_j) - \boldsymbol{\nu}(\mathbf{x}_j)) + \mathbf{n}(\mathbf{x}_j))^\top) \\ &= \mathbf{W}\boldsymbol{\Gamma}(\mathbf{x}_i, \mathbf{x}_j)\mathbf{W}^\top + \boldsymbol{\Lambda}(\mathbf{x}_i, \mathbf{x}_j). \end{aligned} \quad (21)$$

These show that the mean function $\mathbf{m}(\mathbf{x})$ and the covariance function $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ of the multivariate Gaussian process $f(\mathbf{x})$ are given by $\mathbf{m}(\mathbf{x}) = \mathbf{W}\boldsymbol{\nu}(\mathbf{x})$ and $\mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathbf{W}\boldsymbol{\Gamma}(\mathbf{x}, \mathbf{x}')\mathbf{W}^\top + \boldsymbol{\Lambda}(\mathbf{x}, \mathbf{x}')$, respectively.

B Derivation of the posterior GP $f^*(\mathbf{x})$

In this appendix, we derive the posterior Gaussian process $f^*(\mathbf{x})$ shown in Section 4. We here assume that the integral appearing in the definition of the observation model (5) is well-defined, and defer discussion on conditions for its well-definedness to Appendix C. Let $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'))$ be a multivariate GP defined on $\mathcal{X} \subset \mathbb{R}^2$ taking values in \mathbb{R}^S . For an arbitrary $k, k' \in \mathbb{N}$ and an arbitrary set of $(k + k')$ points $\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{x}'_1, \dots, \mathbf{x}'_{k'} \in \mathcal{X}$, let

$$\hat{\mathbf{f}} = \begin{pmatrix} \bar{\mathbf{f}} \\ \bar{\mathbf{f}}' \end{pmatrix}, \quad \bar{\mathbf{f}} = \begin{pmatrix} \mathbf{f}(\mathbf{x}_1) \\ \vdots \\ \mathbf{f}(\mathbf{x}_k) \end{pmatrix}, \quad \bar{\mathbf{f}}' = \begin{pmatrix} \mathbf{f}(\mathbf{x}'_1) \\ \vdots \\ \mathbf{f}(\mathbf{x}'_{k'}) \end{pmatrix}. \quad (22)$$

By the definition of GP, $\hat{\mathbf{f}}$ is a $(k + k')S$ -dimensional Gaussian vector. Let

$$\hat{\mathbf{m}} = \begin{pmatrix} \bar{\mathbf{m}} \\ \bar{\mathbf{m}}' \end{pmatrix}, \quad \bar{\mathbf{m}} = \begin{pmatrix} \mathbf{m}(\mathbf{x}_1) \\ \vdots \\ \mathbf{m}(\mathbf{x}_k) \end{pmatrix}, \quad \bar{\mathbf{m}}' = \begin{pmatrix} \mathbf{m}(\mathbf{x}'_1) \\ \vdots \\ \mathbf{m}(\mathbf{x}'_{k'}) \end{pmatrix} \quad (23)$$

and

$$\hat{\mathbf{K}} = \begin{pmatrix} \bar{\mathbf{K}} & \bar{\mathbf{K}}'^\top \\ \bar{\mathbf{K}}' & \bar{\mathbf{K}}'' \end{pmatrix}, \quad (\bar{\mathbf{K}})_{ij} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j), \quad (\bar{\mathbf{K}}')_{ij} = \mathbf{K}(\mathbf{x}'_i, \mathbf{x}_j), \quad (\bar{\mathbf{K}}'')_{ij} = \mathbf{K}(\mathbf{x}'_i, \mathbf{x}'_j) \quad (24)$$

be the mean vector and the covariance matrix of $\hat{\mathbf{f}}$. In the following, we specifically assume that $\mathbf{x}'_1, \dots, \mathbf{x}'_{k'}$ are taken to be grid points of a regular grid covering \mathcal{X} and with the grid cell volume

Δ , and consider Riemann sums to approximate those integrals on \mathcal{X} appearing in the formulation of SAGP. We then take the limit $\Delta \rightarrow 0$ to derive the posterior GP given areal observations on $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'))$.

Consider the observation process yielding observations \mathbf{y} , defined by

$$\mathbf{y} = \bar{\mathbf{A}}\bar{\mathbf{f}}' + \mathbf{w}, \quad (25)$$

where

$$\bar{\mathbf{A}} = (\mathbf{A}(\mathbf{x}'_1) \ \cdots \ \mathbf{A}(\mathbf{x}'_{k'})) \Delta, \quad (26)$$

and where \mathbf{w} is an S -dimensional Gaussian noise vector with mean zero and covariance Σ . One has

$$\bar{\boldsymbol{\mu}} = \mathbb{E}(\mathbf{y}) = \bar{\mathbf{A}}\bar{\mathbf{m}}' \quad (27)$$

and

$$\bar{\mathbf{C}} = \text{Cov}(\mathbf{y}) = \bar{\mathbf{A}}\bar{\mathbf{K}}''\bar{\mathbf{A}}^\top + \Sigma, \quad (28)$$

respectively. The posterior of $\bar{\mathbf{f}}$ given \mathbf{y} is known to be a multivariate Gaussian with mean

$$\bar{\mathbf{m}}^* = \bar{\mathbf{m}} + \bar{\mathbf{H}}^\top \bar{\mathbf{C}}^{-1}(\mathbf{y} - \bar{\boldsymbol{\mu}}) \quad (29)$$

and covariance

$$\bar{\mathbf{K}}^* = \bar{\mathbf{K}} - \bar{\mathbf{H}}^\top \bar{\mathbf{C}}^{-1} \bar{\mathbf{H}}, \quad (30)$$

respectively, where $\bar{\mathbf{H}} = \bar{\mathbf{A}}\bar{\mathbf{K}}'$.

By regarding sums over the k' terms as Riemann sums approximating the corresponding integrals over \mathcal{X} , in the limit $\Delta \rightarrow 0$, one can replace those sums over k' terms with the corresponding integrals over \mathcal{X} . Specifically, one has

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^{k'} \mathbf{A}(\mathbf{x}'_i) \mathbf{f}(\mathbf{x}'_i) \Delta + \mathbf{w} \\ &\rightarrow \int_{\mathcal{X}} \mathbf{A}(\mathbf{x}) \mathbf{f}(\mathbf{x}) d\mathbf{x} + \mathbf{w}, \end{aligned} \quad (31)$$

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \sum_{i=1}^{k'} \mathbf{A}(\mathbf{x}'_i) \mathbf{m}(\mathbf{x}'_i) \Delta \\ &\rightarrow \int_{\mathcal{X}} \mathbf{A}(\mathbf{x}) \mathbf{m}(\mathbf{x}) d\mathbf{x} = \boldsymbol{\mu}, \end{aligned} \quad (32)$$

$$\begin{aligned} \bar{\mathbf{C}} &= \sum_{i,j=1}^{k'} \mathbf{A}(\mathbf{x}'_i) \mathbf{K}(\mathbf{x}'_i, \mathbf{x}'_j) \mathbf{A}(\mathbf{x}'_j)^\top \Delta^2 + \Sigma \\ &\rightarrow \iint_{\mathcal{X} \times \mathcal{X}} \mathbf{A}(\mathbf{x}) \mathbf{K}(\mathbf{x}, \mathbf{x}') \mathbf{A}(\mathbf{x}')^\top d\mathbf{x} d\mathbf{x}' + \Sigma = \mathbf{C}, \end{aligned} \quad (33)$$

showing that the mean vector $\bar{\boldsymbol{\mu}}$ and the covariance matrix $\bar{\mathbf{C}}$ of \mathbf{y} are reduced in this limit to the vector $\boldsymbol{\mu}$ and the matrix \mathbf{C} defined in (9) and (10), respectively. One also has

$$\begin{aligned} \bar{\mathbf{H}} &= \left(\sum_{i=1}^{k'} \mathbf{A}(\mathbf{x}'_i) \mathbf{K}(\mathbf{x}'_i, \mathbf{x}_1) \Delta \ \cdots \ \sum_{i=1}^{k'} \mathbf{A}(\mathbf{x}'_i) \mathbf{K}(\mathbf{x}'_i, \mathbf{x}_k) \Delta \right) \\ &\rightarrow \left(\int_{\mathcal{X}} \mathbf{A}(\mathbf{x}') \mathbf{K}(\mathbf{x}', \mathbf{x}_1) d\mathbf{x}' \ \cdots \ \int_{\mathcal{X}} \mathbf{A}(\mathbf{x}') \mathbf{K}(\mathbf{x}', \mathbf{x}_k) d\mathbf{x}' \right) \\ &= \left(\mathbf{H}(\mathbf{x}_1) \ \cdots \ \mathbf{H}(\mathbf{x}_k) \right), \end{aligned} \quad (34)$$

where $\mathbf{H}(\mathbf{x})$ is defined in (14). The above calculation shows that in the limit $\Delta \rightarrow 0$ the posterior process $\mathbf{f}^*(\mathbf{x})$ is a multivariate GP with mean function $\mathbf{m}^*(\mathbf{x})$ and covariance function $\mathbf{K}^*(\mathbf{x}, \mathbf{x}')$ given by (15) and (16), respectively.

C On integrability

In this appendix, we discuss conditions for the observation model (5) to be well defined. Assume that $\mathcal{X} \subset \mathbb{R}^2$ is a bounded Jordan-measurable set, and that elements $a_{s,n}(\mathbf{x})$ of $\mathbf{A}(\mathbf{x})$ are Riemann integrable on \mathcal{X} . (The latter assumption is satisfied when $a_{s,n}(\mathbf{x})$ is defined as in (7) with

Jordan-measurable regions $\mathcal{R}_{s,n}$.) Since it is known that a continuous function on \mathcal{X} is Riemann integrable on \mathcal{X} , and that a product of Riemann integrable functions is again Riemann integrable, a sufficient condition for the observation model (5) to be well defined is that the prior process $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'))$ is sample-path continuous.

The assumption made in Section 3 of the integrability of $\nu_l(\mathbf{x})$, $l = 1, \dots, L$, assures integrability of the mean function $\mathbf{m}(\mathbf{x}) = \mathbf{W}\boldsymbol{\nu}(\mathbf{x})$, which allows us to reduce integrability of the prior process $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'))$ to sample-path continuity of the zero-mean process $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}(\mathbf{x}, \mathbf{x}'))$ on \mathcal{X} . A sufficient condition [1, Theorem 1.4.1] for the sample-path continuity of the zero-mean Gaussian process is that for some $0 < C < \infty$ and $\alpha, \eta > 0$

$$k_{s,s}(\mathbf{x}, \mathbf{x}) + k_{s,s}(\mathbf{x}', \mathbf{x}') - 2k_{s,s}(\mathbf{x}, \mathbf{x}') \leq \frac{C}{|\log \|\mathbf{x} - \mathbf{x}'\||^{1+\alpha}} \quad (35)$$

holds for all $s \in \{1, \dots, S\}$ and for all \mathbf{x}, \mathbf{x}' with $\|\mathbf{x} - \mathbf{x}'\| < \eta$. If one uses the squared-exponential kernels for $\{\gamma_l\}$, then one can confirm that the above condition is satisfied, and consequently the observation model (5) is well defined.

It should be noted that the sample-path continuity discussed above is different from the mean-square (MS) continuity. A process $\mathbf{f}(\mathbf{x})$ is said to be MS continuous at $\mathbf{x} = \mathbf{x}_*$ if for any sequence \mathbf{x}_k converging to \mathbf{x}_* as $k \rightarrow \infty$ it holds that $\mathbb{E}[\|\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{x}_*)\|^2] \rightarrow 0$ as $k \rightarrow \infty$. A necessary and sufficient condition for a random field to be MS continuous at \mathbf{x}_* is that its covariance function $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ is continuous at the point $\mathbf{x} = \mathbf{x}' = \mathbf{x}_*$ [2, Appendix 10A], which in the case of Gaussian processes is weaker than the above sufficient condition for the sample-path continuity.

D Description of real-world areal data sets

We used the real-world areal data sets from NYC Open Data⁴ and Chicago Data Portal⁵ to evaluate the proposed model. These data sets are collected and released for improving city environments, and consist of a variety of categories including social indicators, land use, and air quality. Details of the areal data sets we used in the experiments are listed in Table 3. The number of data sets in New York City and Chicago are 10 and 3, respectively. Each data set is associated with one of the predefined geographical partitions. The number of partition types in New York City and Chicago are 4 and 2, respectively. Table 3 shows the respective partition names and the number of regions in the corresponding partition. These data sets are gathered once a year at the time ranges shown in Table 3; the values of data were divided by the number of observation times. Then, the data were normalized so that each variable in each city has zero mean and unit variance.

E Results

Table 4 shows MAPE and standard errors for GPR, 2-stage GP, SLFM, and SAGP, where the experiments for Crime data set in Chicago have not been conducted because the coarser version for training is not available online. For all data sets, SAGP achieved the comparable or better performance than the other methods.

References

- [1] Adler, R. J., and Taylor, J. E. 2007. *Random Fields and Geometry*. Springer.
- [2] Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd edition.

⁴<https://opendata.cityofnewyork.us>

⁵<https://data.cityofchicago.org/>

Table 3: Real-world areal data sets.

(a) New York City			
Data	Partition	#regions	Time range
PM2.5	UHF42	42	2009 – 2010
Poverty rate	Community district	59	2009 – 2013
Unemployment rate	Community district	59	2009 – 2013
Mean commute	Community district	59	2009 – 2013
Population	Community district	59	2009 – 2013
Recycle diversion rate	Community district	59	2009 – 2013
Crime	Police precinct	77	2010 – 2016
Fire incident	Zip code	186	2010 – 2016
311 call	Zip code	186	2010 – 2016
Public telephone	Zip code	186	2016

(b) Chicago			
Data	Partition	#regions	Time range
Crime	Police Precinct	25	2012
Poverty rate	Community district	77	2008 – 2012
Unemployment rate	Community district	77	2008 – 2012

Table 4: MAPE and standard errors for the prediction of fine-grained areal data in New York City and Chicago. The numbers in parentheses denote the number L of the latent GPs estimated by the validation procedure. The single star (\star) and the double star ($\star\star$) indicate significant difference between SAGP and other models at the levels of P values of < 0.05 and < 0.01 , respectively.

(a) New York City				
	GPR	2-stage GP	SLFM	SAGP
PM2.5	0.072 ± 0.010 (–)	0.042 ± 0.005 (–)	0.036 ± 0.005 (6)	$0.030 \pm 0.005^*$ (5)
Poverty rate	0.344 ± 0.046 (–)	0.210 ± 0.022 (–)	0.207 ± 0.025 (4)	$0.177 \pm 0.019^{\star\star}$ (3)
Unemployment rate	0.319 ± 0.036 (–)	0.193 ± 0.021 (–)	0.195 ± 0.024 (3)	$0.165 \pm 0.020^*$ (3)
Mean commute	0.131 ± 0.020 (–)	0.068 ± 0.009 (–)	0.057 ± 0.007 (4)	0.050 ± 0.007 (6)
Population	0.577 ± 0.104 (–)	0.389 ± 0.033 (–)	0.337 ± 0.039 (3)	$0.295 \pm 0.033^*$ (3)
Recycle diversion rate	0.353 ± 0.049 (–)	0.236 ± 0.034 (–)	0.222 ± 0.032 (4)	0.211 ± 0.029 (4)
Crime	0.860 ± 0.102 (–)	0.454 ± 0.075 (–)	0.401 ± 0.053 (2)	$0.379 \pm 0.055^{\star\star}$ (3)
Fire incident	1.097 ± 0.097 (–)	0.746 ± 0.084 (–)	0.500 ± 0.052 (4)	$0.396 \pm 0.038^{\star\star}$ (3)
311 call	0.083 ± 0.004 (–)	0.070 ± 0.004 (–)	0.061 ± 0.004 (6)	$0.052 \pm 0.003^{\star\star}$ (3)
Public telephone	0.131 ± 0.008 (–)	0.083 ± 0.008 (–)	0.086 ± 0.008 (4)	0.080 ± 0.007 (6)

(b) Chicago				
	GPR	2-stage GP	SLFM	SAGP
Poverty rate	0.599 ± 0.099 (–)	0.380 ± 0.060 (–)	0.335 ± 0.052 (2)	$0.278 \pm 0.032^{\star\star}$ (2)
Unemployment rate	0.478 ± 0.047 (–)	0.318 ± 0.032 (–)	0.278 ± 0.025 (2)	$0.231 \pm 0.021^*$ (2)