FactorGP: A conditional-independence-aware approximate Gaussian process model for mapping

Jennifer Wakulicz*, Teresa Vidal-Calleja†, Thierry Peynot‡,
Tom Drummond§, Stefan Williams* and Donald G. Dansereau*
*Australian Centre for Robotics, School of Aerospace, Mechanical and Mechatronic Engineering, University of Sydney
Email: {firstname.lastname}@sydney.edu.au
†QUT Centre for Robotics, Queensland University of Technology (QUT)
Email: t.peynot@qut.edu.au
‡Robotics Institute, University of Technology Sydney
Email: teresa.vidalcalleja@uts.edu.au
§School of Computing and Information Systems, University of Melbourne
Email: tom.drummond@unimelb.edu.au

***Abstract*—Gaussian processes are widely used for robotic mapping of spatial fields such as gas concentration, terrain height and ocean currents. These physical phenomena often display non-stationarity and discontinuities ill-captured by stationary kernels currently standardised in literature. We present *FactorGP*, an approximate Gaussian process factor graph model that represents data non-stationarity and discontinuities with spatially varying factor parameters and conditional independence in a learned prior. The hierarchical model structure separates data uncertainty from model uncertainty, enabling new active mapping capabilities unique to FactorGP that can target each source independently. We evaluate FactorGP against established non-stationary kernel methods for active mapping of discontinuous gas fields and non-stationary terrains. FactorGP exhibits more faithful reconstruction of high-frequency structure, having positive implications for future use in downstream tasks such as robotic infrastructure inspection and environmental monitoring.**

I. Introduction

Mapping is a fundamental capability in robotics. Different map representations – sparse, dense, semantic, probabilistic – can help or hinder downstream robotics tasks. The continuous nature of Gaussian processes makes them well-suited to mapping continuous physical phenomena such as ocean currents [1], [2], terrain [3], [4], [5], gas [6] or temperature fields [7], [8]. Their probabilistic nature allows a robot to query any spatial location for a field prediction with associated uncertainty, enabling *active* mapping, where a robot chooses sensing locations that minimise uncertainty in the map.

However, active mapping frameworks that naively employ GPs suffer several limitations. First, the cost of exact inference with GPs scales cubically with the number of observations, resulting in ever-growing compute requirements that become prohibitive in online active mapping. Second, commonly used stationary kernels assume data are correlated uniformly throughout the domain, giving map reconstructions that smooth over discontinuities and high variance regions. Finally, kernel hyperparameters are typically fixed throughout active mapping missions, ignoring uncertainty arising from these modelling choices.

In this paper, we address these limitations through a single approximate GP representation. Inspired by the link between Matérn GPs and Gaussian Markov random fields (GMRFs) [9], we present *FactorGP*[1]. FactorGP represents a Matérn-approximating GMRF as a factor graph, exposing the conditional independence structure of the GMRF as a set of independently parametrised factors, or graph edges. This way, per-edge correlation and connectivity can be learned from data individually to build an informed prior that faithfully captures non-stationarity and discontinuities in the underlying field (see Fig. 1). Observations are then incorporated via simple additive operations on the global prior, building a posterior that retains the GMRF sparsity for efficient inference.

Since sparse, noisy observations induce uncertainty in the learned GMRF, FactorGP infers a distribution over factor graph priors via Stein variational gradient descent (SVGD), yielding a hierarchical probabilistic model. At the higher layer is the distribution over factor graphs, each defining a GMRF prior over the field at the lower layer, thus separating uncertainty in the field given a prior model from uncertainty over the prior model itself. Excitingly, this permits the design of new active mapping strategies that can independently target reduction in model uncertainty versus reconstruction uncertainty, or balance the two goals.

Our contributions are summarised as follows:

- FactorGP, a conditional-independence-aware approximate GP for mapping discontinuous, non-stationary fields,
- an active mapping paradigm unique to FactorGP in which model and data uncertainty are separated,
- evaluation against existing stationary and non-stationary baselines on simulated gas and real-world terrain datasets.

FactorGP map reconstructions capture structural features more accurately than existing methods and enable targeted information gathering to deduce map topology, benefiting downstream tasks such as environmental monitoring and infrastructure inspection.

[1]Code and datasets released here.

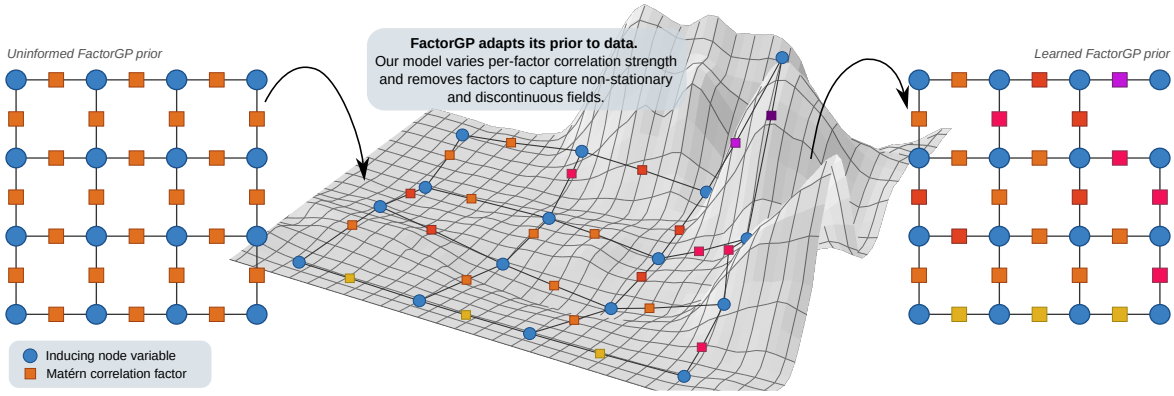

Fig. 1: We introduce FactorGP, a model that represents a Matérn-GP-approximating GMRF as a factor graph, where inducing nodes (circles) are connected by correlation factors (squares). Starting from an uninformed prior (left), factor parameters and inclusions are inferred from data (centre), giving an informed prior (right) with spatially varying correlation strengths (represented by factor colour) and conditional independence between nodes (represented by absent factors), allowing FactorGP to model non-stationarity and discontinuities respectively.

II. Related Work

The cubic cost of exact inference using GPs has motivated extensive research into sparse approximations. Sparse GPs introduce a fixed number of inducing variables which act to summarise data, reducing inference cost when the inducing set is smaller than the number of observations. Many sparse GP models used in active mapping [10], [11] employ the fully independent training conditional assumption [12], treating all observations as mutually independent given the inducing set, leaving an all-to-all dependence assumption between the inducing variables.

Other sparse GP models make tighter conditional independence assumptions to further reduce computation. KISS-GP [13] places inducing points on a uniform grid and assumes only nearest neighbours are correlated, reducing inference to local grid interpolation. This grid-and-interpolation structure mirrors Gaussian Markov random field (GMRF) approximations to GPs, where dependencies between variables in a joint Gaussian distribution are captured in a graph with edges between dependent variables. [9] make the connection between GMRFs and GPs explicit by constructing GMRFs that solve the finite-difference discretisation of the stochastic partial differential equation for which the Matérn GP is a solution.

More generally the view of sparse GP mapping as inference over a factor graph – a graph of latent variables with edges encoding their dependencies – has been leveraged for scalability in [14], where inducing points are arranged into a tree to achieve linear-time inference via Gaussian belief propagation (GBP). Factor graph representations further lend themselves to submap fusion for GP mapping in larger domains. In [15], the domain is partitioned into conditionally independent submaps connected in a GMRF framework, updating each individually and propagating information through edge connections. [16] performs a similar partitioning, but updates the global map through loopy GBP. However, in all cases the conditional independences and thus factor graph structure are designed to reduce computational demand, relegating the implications of these modelling assumptions for reconstruction to a secondary effect. FactorGP instead exploits this as a feature, inferring the factor graph distribution from data to faithfully reconstruct discontinuities and non-stationarity in the data.

The stationary assumption is common in GP regression but can be violated by physical data, particularly at discontinuities. This motivates the development of non-stationary GP models. Mixture-of-expert models [17] combine several GPs with independently optimised hyperparameters with a gating function that assigns experts based on inputs. However, this gating function is typically expensive to learn. Input warping models such as Deep Kernel Learning (DKL) [18] instead use small neural networks to map inputs to a latent space in which the stationary assumption applies, but are known to overfit [19], and cannot capture discontinuities. The Attentive Kernel (AK) [4] learns spatially-dependent weighted sums of a *finite* set of lengthscales, reducing overfitting relative to DKL. AK additionally addresses discontinuities by learning visibility scores between spatial inputs. Both AK and DKL are learning-based approaches, capturing non-stationarity implicitly and working best when pre-trained on pilot data. Meanwhile, FactorGP encodes non-stationarity and discontinuities explicitly in a GMRF factor graph – a representation that is both interpretable and amenable to the scalability improvements demonstrated by the mapping methods discussed above.

III. Method

Consider a latent function of interest $f : \mathcal{X} \to \mathbb{R}$ over a given spatial domain $\mathcal{X}$. A mobile robot at position $\mathbf{x}_i \in \mathcal{X}$ at time i takes sparse, noisy measurements z_i of the function with measurement model

$$z_i = f(\mathbf{x}_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2), \tag{1}$$

where ε_i is zero-mean Gaussian distributed sensor noise with noise variance σ_n^2. Given a measurement set $\mathbf{z} = \{z_1, \ldots, z_N\}$

we wish to reconstruct a probabilistic estimate of f and use uncertainty to guide future data collection.

To this end, we introduce FactorGP (Fig. 1), an approximate GP regression model that exploits the connection between GMRFs and Matérn GPs [9]. At its core, FactorGP places a GMRF factor graph over the domain, where each edge carries a local factor encoding correlation strength between inducing nodes. Rather than fixing edge factors to those implied by a stationary kernel, FactorGP infers them from data, allowing the model to vary correlation across the domain and capture discontinuities via absent edges.

A. The FactorGP prior

We model the function prior as a GMRF on a regular grid of M inducing nodes $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]^\mathsf{T}$ with uniform spacing h. Letting $\mathbf{f}_\mathbf{U} = [f(\mathbf{u}_1), \ldots, f(\mathbf{u}_M)]^\mathsf{T}$ denote the vector of function values at the inducing nodes, the prior is a zero-mean multivariate Gaussian written in *information form* as

$$p(\mathbf{f}_\mathbf{U} \mid \mathbf{Q}) = \mathcal{N}_I(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{Q}}), \tag{2}$$

where $\mathbf{Q} = (\mathbf{G}, \boldsymbol{\kappa})$ flags the prior precision's dependence on two variables: $\mathbf{G}$, an adjacency matrix detailing edge connections between inducing nodes and $\boldsymbol{\kappa}$, a corresponding vector of per-edge correlation parameters.

The sparse connectivity captured by $\mathbf{G}$ implies conditional independence assumptions between inducing nodes, reducing the prior to a product over pairwise factors

$$p(\mathbf{f}_\mathbf{U} \mid \mathbf{Q}) = \prod_{(i,j)\in\mathbf{G}} p(f(\mathbf{u}_i), f(\mathbf{u}_j)), \tag{3}$$

where each $p(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \mathcal{N}_I(\mathbf{0}, \boldsymbol{\Lambda}_{\boldsymbol{Q}}^{ij})$. This decomposition admits a factor graph representation of the GMRF prior exemplified in Fig. 1 and gives a natural, intuitive method for constructing the joint global precision $\boldsymbol{\Lambda}_\mathbf{Q}$ induced by $\mathbf{Q}$ via local factors only. That is, each edge $\mathbf{G}_{ij} = 1$ in $\mathbf{G}$ contributes a local 2×2 factor with zero mean and precision

$$\boldsymbol{\Lambda}_\mathbf{Q}^{ij} = \frac{1}{h^2}\begin{bmatrix} a_{ij}/2^d & -1 \\ -1 & a_{ij}/2^d \end{bmatrix}, \qquad a_{ij} = \kappa_{ij}^2 h^2 + 2d, \tag{4}$$

where d is spatial dimension and $\kappa_{ij} \in \boldsymbol{\kappa}$ is the per-edge stiffness dictating correlation between nodes going as $1/\kappa_{ij}$.

Notably, this formulation allows one to vary the per-edge correlation parameters between nodes to capture *non-stationary data*. Further, binary edge inclusions $\mathbf{G}_{ij} \in \{0, 1\}$ model zero correlation between nodes, allowing the prior capture strict *discontinuities*. FactorGP infers $\mathbf{Q} = (\mathbf{G}, \boldsymbol{\kappa})$ from data by approximating $p(\mathbf{Q} \mid \mathbf{z})$, as detailed in Section III-C.

B. Posterior inference with FactorGP

Conditioning the FactorGP prior on observations yields a Gaussian posterior. Since observations of the latent function f can be taken at arbitrary locations, we employ bilinear interpolation to calculate a weight vector $\mathbf{w}_i \in \mathbb{R}^M$ with at most four non-zero entries summing to one. These entries correspond to the relative distance of observation z_i to its four surrounding inducing nodes. Thus, the measurement model in Eq. (1) becomes

$$z_i = \mathbf{w}_i^\mathsf{T} \mathbf{f}_\mathbf{U} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2). \tag{5}$$

To condition the prior on an observation, we again leverage the factor graph representation. Here, each observation attaches to the graph as at most a quaternary factor connecting surrounding inducing nodes identified by $\mathbf{w}_i$,

$$p(z_i \mid f(\mathbf{u}_j), f(\mathbf{u}_k), f(\mathbf{u}_l), f(\mathbf{u}_m)) = \mathcal{N}_I(\boldsymbol{\eta}_{z_i}, \boldsymbol{\Lambda}_{z_i}), \tag{6}$$

where $f(\mathbf{u}_{(\cdot)})$ are the surrounding inducing node values, $\boldsymbol{\eta}_{z_i} = \frac{1}{\sigma_n^2}\mathbf{w}_i z_i$ and $\boldsymbol{\Lambda}_{z_i} = \frac{1}{\sigma_n^2}\mathbf{w}_i\mathbf{w}_i^\mathsf{T}$. This data factor then contributes additively to the prior to give the posterior

$$p(\mathbf{f}_\mathbf{U} \mid \mathbf{Q}, z_i) = \mathcal{N}_I(\boldsymbol{\eta}_{\text{post}} = \mathbf{0} + \boldsymbol{\eta}_{z_i}^{(M)}, \boldsymbol{\Lambda}_{\text{post}} = \boldsymbol{\Lambda}_\mathbf{Q} + \boldsymbol{\Lambda}_{z_i}^{(M)}), \tag{7}$$

where the notation $(\cdot)^{(M)}$ denotes the M-dimensional vector or square matrix with entries of $(\cdot)$ placed in positions associated with the surrounding inducing nodes and filling all remaining entries with zero.

The posterior mean at inducing nodes is recovered as $\boldsymbol{\mu} = \boldsymbol{\Lambda}_{\text{post}}^{-1}\boldsymbol{\eta}_{\text{post}}$, where inversion of $\boldsymbol{\Lambda}_{\text{post}}$ is efficient due to its sparsity. Then, predictions f_* at arbitrary locations $\mathbf{x}_* \in \mathcal{X}$ are calculated with interpolation weights again, analogous to [13],

$$\begin{aligned} \mathbb{E}[f_* \mid z_i, \mathbf{Q}] &= \mathbf{w}_*^\mathsf{T}\boldsymbol{\mu}, \\ \text{Var}[f_* \mid z_i, \mathbf{Q}] &= \mathbf{w}_*^\mathsf{T}\boldsymbol{\Lambda}_{\text{post}}^{-1}\mathbf{w}_*. \end{aligned} \tag{8}$$

When handling streamed data, the prior in Eq. (7) should be replaced by the previous posterior at each iteration in a Bayesian fashion. When performing batch updates with more than one observation, we stack observations into vector $\mathbf{z}$ and corresponding weight vectors into matrix $\mathbf{W}$. For batch queries, we similarly formulate weight matrix $\mathbf{W}_*$ for use in place of the weight vector in Eq. (8).

C. Factor learning with Stein variational gradient descent

We infer per-edge inclusions $\mathbf{G}_{ij}$ and correlation parameters κ_{ij} of each factor in Eq. (4) from data by approximating the posterior $p(\mathbf{Q} \mid \mathbf{z})$ with Stein variational gradient descent (SVGD) [20] for its parallelism and ability to capture the full posterior, thereby allowing model uncertainty to be quantified. To formulate the posterior $p(\mathbf{Q} \mid \mathbf{z}) \propto p(\mathbf{z} \mid \mathbf{Q})p(\mathbf{Q})$ we first place a prior over the graph structure $\mathbf{G}$ and correlation $\boldsymbol{\kappa}$,

$$p(\mathbf{Q}) = \prod_{(i,j)\in\mathcal{E}} \text{Bernoulli}(\mathbf{G}_{ij}; q) \log\mathcal{N}(\kappa_{ij}; \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}^2}), \tag{9}$$

where $\mathcal{E}$ is a prior candidate edge set, q is the factor inclusion prior probability and $\log\mathcal{N}$ is a log-normal distribution with prior mean $\mu_{\boldsymbol{\kappa}}$ and variance $\sigma_{\boldsymbol{\kappa}}^2$.

The GMRF log marginal likelihood term $\log p(\mathbf{z} \mid \mathbf{Q})$ is

$$\log p(\mathbf{z} \mid \mathbf{Q}) = -\frac{1}{2}\left[\mathbf{z}^\mathsf{T}\boldsymbol{\Sigma}_z^{-1}\mathbf{z} + \log|\boldsymbol{\Sigma}_z| + N\log 2\pi\right], \tag{10}$$

where $\boldsymbol{\Sigma}_z = \mathbf{W}\boldsymbol{\Lambda}_\mathbf{Q}^{-1}\mathbf{W}^\mathsf{T} + \sigma_n^2\mathbf{I}$, and $\mathbf{W}$ is the bilinear interpolation matrix for vector of stacked observations $\mathbf{z}$ detailed in Section III-B.

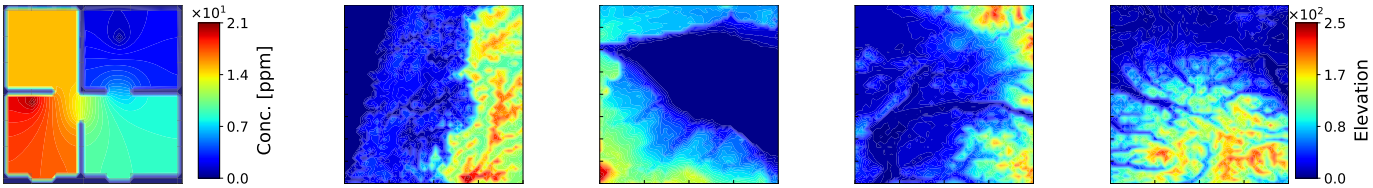

Fig. 2: Ground truth datasets. **Left:** The ground truth gas field with source (bottom-left quadrant) flowing to a sink (top-right quadrant) and distinct field discontinuities at walls. **Centre-left to right:** terrain height maps from the SRTM [23] dataset: the `N17E073`, `N43W080`, `N45W123`, `N47W124` maps, each showing varying degrees of non-stationarity and discontinuities.

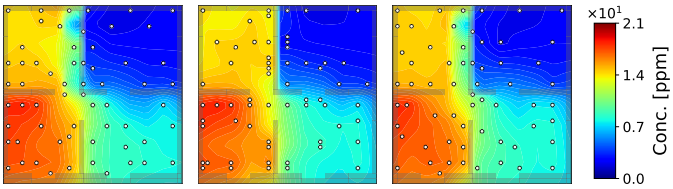

Fig. 3: Reconstructed gas field after $N = 60$ samples (shown as circles). **Left-to-right:** Data-only $(\alpha, \beta) = (1, 0)$, model-only $(\alpha, \beta) = (0, 1)$ and balanced $(\alpha, \beta) = (1, 1)$ acquisition functions. The model-only acquisition prioritises sampling near boundaries to identify discontinuities in data.

With prior and likelihood in place, we initialise K SVGD particles encoding $(\mathbf{G}_{ij}, \kappa_{ij})$ for each edge (i, j) in a set of pre-defined candidate edges $\mathcal{E}$. The particles are iteratively transported in parallel according to

$$\mathbf{g}^{(l)} \leftarrow \mathbf{g}^{(l)} + \alpha \hat{\phi}^*(\mathbf{g}^{(l)}), \tag{11}$$

where $\hat{\phi}^*$ is the vector field that optimally reduces the Kullback-Leibler divergence between the particle distribution and the posterior $p(\mathbf{Q} \mid \mathbf{z})$. That is,

$$\hat{\phi}^*(\mathbf{g}) = \frac{1}{K} \sum_{i=1}^{K} \left[k_{\mathbf{g}^{(i)}} \nabla_{\mathbf{g}^{(i)}} \log p(\mathbf{Q}^{(i)} \mid \mathbf{z}) + \nabla_{\mathbf{g}^{(i)}} k_{\mathbf{g}^{(i)}} \right], \tag{12}$$

where $k_{\mathbf{g}^{(i)}} = k(\mathbf{g}^{(i)}, \mathbf{g})$ is the RBF kernel with lengthscale set according to the median heuristic used in [20]. Since $\mathbf{G}_{ij} \in \{0, 1\}$ is a binary variable, we use the Gumbel-Softmax straight-through estimator [21], [22] to obtain differentiable samples during the SVGD optimisation.

The culmination of the above defines the hierarchical FactorGP model where $\mathbf{Q}$ is inferred at the higher level, shaping the GMRF prior over the latent function f at the lower level. The full predictive model marginalises over both levels,

$$p(f_*|\mathbf{z}) = \int \underbrace{p(f_* \mid \mathbf{f}_\mathbf{U}, \mathbf{Q}) p(\mathbf{f}_\mathbf{U} \mid \mathbf{Q}, \mathbf{z})}_{\substack{\text{estimated with GMRF} \\ \text{(Eq. (8))}}} \underbrace{p(\mathbf{Q} \mid \mathbf{z})}_{\substack{\text{estimated} \\ \text{with SVGD}}} d\mathbf{Q} d\mathbf{f}_\mathbf{U}. \tag{13}$$

IV. FactorGP for Active Mapping

The hierarchical nature of FactorGP separates two sources of uncertainty: uncertainty due to data, and uncertainty in the model. Concretely, the GMRF posterior captures data uncertainty while the posterior over $\mathbf{Q} = (\mathbf{G}, \boldsymbol{\kappa})$ approximated by SVGD captures model uncertainty. This admits bespoke active perception to target reduction in one or the other.

The information provided by a measurement at candidate sensing location $\mathbf{x}_*$ can be quantified by the total variance,

$$\mathrm{Var}[f_*] = \underbrace{\mathbb{E}_\mathbf{Q}[\mathrm{Var}[f_* \mid \mathbf{Q}, \mathbf{z}]]}_{\text{data uncertainty}} + \underbrace{\mathrm{Var}_\mathbf{Q}[\mathbb{E}[f_* \mid \mathbf{Q}, \mathbf{z}]]}_{\text{model uncertainty}}, \tag{14}$$

requiring field predictions (Eq. (8)) from all SVGD particles to compute. As this can be expensive for high K, we propose the lightweight proxy

$$h(\mathbf{x}_*) = \alpha \cdot \mathrm{Var}\left[f_* \mid \hat{\mathbf{Q}}, \mathbf{z} \right] + \beta \cdot \mathbf{w}_*^\mathsf{T} \boldsymbol{\sigma}_\mathbf{Q}, \tag{15}$$

where the first term is the predictive variance with $\mathbf{Q} = \hat{\mathbf{Q}}$, the maximum a posteriori model found by SVGD and the second term is the SVGD particle spread $\boldsymbol{\sigma}_\mathbf{Q}$ averaged over local edges identified by weights $\mathbf{w}_*$. Scalars $\alpha, \beta \geq 0$ serve to weigh data and model uncertainty respectively, allowing either term to be targeted independently by setting the other to zero.

V. Evaluation

We evaluate FactorGP on three fronts across a suite of environments shown in Fig. 2. First, we explore the unique information gathering heuristic admitted by the model. Second, we evaluate the model's ability to capture discontinuities. Third, we evaluate its ability to capture non-stationary data.

A. Exploiting decomposed uncertainty for active perception

To illustrate the value of the decoupled uncertainty heuristic Eq. (15), we study active mapping of gas in a building. The ground truth field, shown in Fig. 2, is generated by simulating gas diffusion between a source in the bottom-left quadrant and a sink in the top-right, with walls inducing discontinuities in the field. We run active sampling experiments under three heuristic weightings to examine sampling behaviour when reducing data uncertainty $(\alpha, \beta) = (1, 0)$, model uncertainty $(\alpha, \beta) = (0, 1)$ and balancing the two $(\alpha, \beta) = (1, 1)$.

The fields reconstructed from $N = 60$ samples are compared in Fig. 3. The model-only heuristic exhibits distinct sampling behaviour focused along either side of the building walls as it attempts to disambiguate competing model priors that best describe the data, consequently producing the most faithful field reconstruction of the discontinuities. Meanwhile, the data-only and balanced heuristics sample more broadly across the domain, yielding similar reconstructions. Still, the

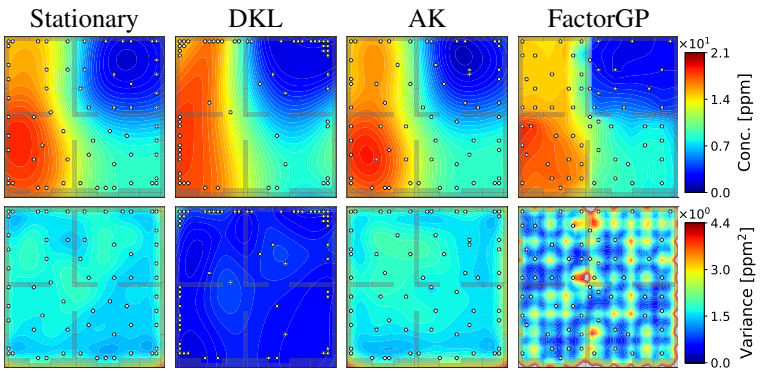

Fig. 4: Gas concentration mapping results. **Left to right:** The Matérn-3/2 stationary GP, DKL, AK, and FactorGP models. **Row 1:** predictive mean, **Row 2:** predictive variance. Sampling locations are shown in white. FactorGP most faithfully reconstructs discontinuities. The FactorGP variance reflects model uncertainty, peaking around the building structure.

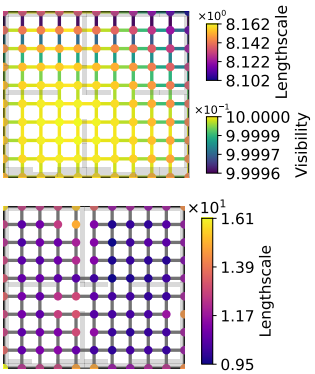

Fig. 5: AK (top) and FactorGP (bottom) lengthscales (node colour) and AK visibility (edge colour) for the gas dataset. FactorGP graph shows exclusion of factors bridging walls, while AK parameters are uniform across the domain.

TABLE I: Gas concentration active sampling results. Best and second best results per dataset are indicated by shading.

Model	RMSE↓	GRMSE↓	SMSE↓	SSIM↑
Stationary	1.53	4.2	0.07	0.47
DKL	1.61	4.19	0.08	0.43
AK	1.52	4.17	0.07	0.44
FactorGP	1.35	4.17	0.06	0.55

balanced heuristic recovers slightly sharper reconstruction of the discontinuities between vertical walls due to inclusion of model uncertainty.

B. Mapping discontinuous and non-stationary functions

We evaluate FactorGP's performance on mapping the discontinuous gas dataset and the non-stationary NASA Shuttle Radar Topography Mission (SRTM) terrain maps [23] (Fig. 2). As in [4], SRTM elevation maps are discretised to 8-bit grayscale bitmaps, mapping elevation to pixel intensities in the range $[0, 255]$. We evaluate FactorGP against a stationary Matérn-3/2 sparse GP and two non-stationary baselines: the Attentive Kernel (AK) [4] and the Deep Kernel Learning (DKL) [18] method. We use the default implementation of AK and DKL released by the authors of [4], implementing both in a sparse GP regime and replacing the RBF kernel with Matérn-3/2 for more direct comparison with FactorGP.

For both datasets, all methods are initialised with only a single initial sample. At each step, baselines select sensing locations according to predictive entropy maximisation while FactorGP uses Eq. (15) with $(\alpha, \beta) = (1, 1)$. For gas field mapping, we study unconstrained active sampling (the robot can 'teleport'), while for the SRTM terrain mapping experiments we model a sensor-equipped mobile robot that takes measurements continuously along a Dubins path between selected sensing locations.

We evaluate all methods with four metrics: 1) root mean squared error (RMSE) to measure overall reconstruction accuracy, 2) standardised mean square error (SMSE), which normalises the mean squared error by test target variance, penalising methods that fail to capture scale, 3) structural similarity (SSIM), which penalises the over-smoothing rewarded by RMSE by measuring structural similarity between predicted and ground truth maps, 4) the RMSE of predicted versus ground truth gradient magnitudes (GRMSE) to specifically measure the recovery of sharp transitions and discontinuities in reconstructions.

1) Gas regression results: Results for an active sampling budget of $N = 60$ are shown in Fig. 4. On this dataset, AK is given 10 lengthscale primitives between $[0.01, 0.5]$ and the FactorGP prior Eq. (9) is initialised with $(q, \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}}) = (0.55, 0.01, 1.0)$. All models re-optimise after each sample. Here the stationary and DKL models severely smooth the gas field, ignoring discontinuities in gas concentration across walls. The AK model produces more physically plausible reconstructions, particularly around the wall separating top quadrants, though discontinuities are still poorly resolved. FactorGP most faithfully reconstructs the full field and focuses sampling around the discontinuities, capturing them most accurately, as supported by the quantitative results in Table I. Unlike baseline approaches, the uncertainty captured by FactorGP peaks around room boundaries, a direct result of its capacity to capture model uncertainty. Quantitatively, FactorGP improves upon all baselines, demonstrating the structural advantage of explicit discontinuity modelling with FactorGP. The final AK and FactorGP model structures are illustrated in Fig. 5. Notably, the AK visibility scores remain near-uniform throughout the domain, failing to suppress correlations across walls. Meanwhile, FactorGP correctly removes factors that bridge room boundaries where the field discontinuity is most pronounced, inferring building topology.

2) SRTM terrain results: Quantitative results for the SRTM active terrain mapping experiments under a sampling budget of $N = 700$ are given in Table II. Since each measurement contributes only a local update to FactorGP via Eq. (6), we

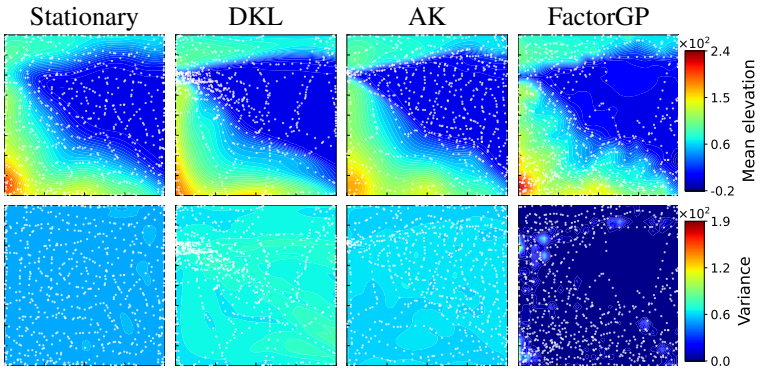

Fig. 6: N43W080 results. **Left to right:** The stationary Matérn-3/2, DKL, AK, and FactorGP models. **Row 1:** predictive mean. **Row 2:** predictive variance. Samples collected along trajectories are shown in white. FactorGP prioritises sampling rough terrain where model uncertainty is high, while AK and DKL priorities the lake bed. The stationary GP exhibits uniform coverage.

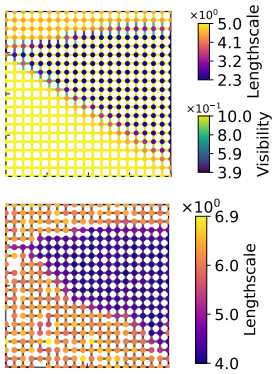

Fig. 7: AK (top) and FactorGP (bottom) learned lengthscales (node colour) and AK visibility (edge colour) for N43W080. Both identify discontinuities, but AK does not learn zero cross-border correlations, while FactorGP does.

TABLE II: Active terrain mapping results. Best and second best results are indicated by shading. FactorGP prior parameters are listed below dataset labels.

	N17E073 $(q, \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}}) = (0.65, 1.5, 1.0)$				N43W080 $(q, \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}}) = (0.69, -0.8, 1.0)$				N45W123 $(q, \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}}) = (0.65, 2.2, 1.0)$				N47W124 $(q, \mu_{\boldsymbol{\kappa}}, \sigma_{\boldsymbol{\kappa}}) = (0.65, 1.7, 1.0)$			
Model	RMSE↓	GRMSE↓	SMSE↓	SSIM↑	RMSE↓	GRMSE↓	SMSE↓	SSIM↑	RMSE↓	GRMSE↓	SMSE↓	SSIM↑	RMSE↓	GRMSE↓	SMSE↓	SSIM↑
Stationary	19.01	14.91	0.09	0.49	11.67	8.43	0.06	0.72	14.88	10.45	0.09	0.67	24.91	19.99	0.18	0.39
DKL	26.21	14.40	0.17	0.40	10.33	6.81	0.04	0.81	27.97	11.11	0.33	0.39	32.37	16.58	0.31	0.36
AK	23.43	14.45	0.13	0.42	10.04	7.37	0.04	0.81	17.20	10.62	0.12	0.59	26.98	17.27	0.21	0.39
FactorGP	20.57	13.41	0.10	0.42	7.93	6.50	0.03	0.82	16.91	10.10	0.12	0.61	25.90	15.72	0.20	0.4

accumulate 100 measurements before re-running SVGD to ensure a meaningful model update. Measurements have global impact in baseline methods, which we therefore optimise every sample for best performance. AK is initialised with 10 lengthscale primitives evenly spaced in $[0.01, 0.5]$. FactorGP initial params are listed below dataset titles in Table II.

FactorGP achieves the lowest GRMSE for all maps, indicating consistently superior recovery of sharp terrain transitions. Interestingly, the stationary model outperforms non-stationary baselines in most metrics, a stark contrast to evaluation presented in [4] where these methods use the radial basis function (RBF) kernel. We attribute this to the reduced smoothness of Matérn-3/2 better suiting the high spatial variance of physical processes than RBF as argued in [24], [25], reducing the margin for non-stationary methods to improve upon. Nevertheless, FactorGP is the best performing non-stationary model overall.

The abrupt lake bed boundary in N43W080 presents a challenging test for the stationary Matérn-3/2 baseline. Here the non-stationary methods excel, with FactorGP dominating across all metrics. The robot trajectories, map reconstructions and variances are shown in Fig. 6, where we see divergent active sensing behaviours from each model. The stationary baseline takes a coverage-like trajectory while DKL and AK incorrectly focus sensing efforts on the lake bed. FactorGP interrogates more informative high elevation terrain, giving a more accurate reconstruction. The final AK and FactorGP models are shown in Fig. 7, where lengthscales of both models reflect the terrain features. The visibility scores learned by AK mask correlations by at most 0.39. FactorGP instead makes hard binary exclusions, particularly along the top lake edge.

VI. Conclusion

In this paper we introduced FactorGP, a conditional-independence-aware approximate GP model for mapping non-stationary and discontinuous spatial functions. By separating model and data uncertainty, FactorGP admits unique active mapping in which robots can gather information to reduce either source independently. This framework opens many exciting avenues for future research. Most immediately come extensions to larger domains with conditional-independence-aware submapping and incremental SVGD updates that exploit model structure. More ambitiously, we wish to incorporate directed edges in the factor graph prior for causal inference. This capability could enable *interactive* perception – where a robot can plan interventions to determine cause and effect embedded in the graph structure – with significant potential to advance infrastructure and environmental monitoring.

Acknowledgments

This research is funded by the Australian Research Council discovery project DP250103135. T.P. acknowledges continued support from the Queensland University of Technology (QUT) through the Centre for Robotics.

References

[1] R. Berlinghieri, B. L. Trippe, D. R. Burt, R. J. Giordano, K. Srinivasan, T. Özgökmen, J. Xia, and T. Broderick, "Gaussian processes at the Helm(holtz): A more fluid model for ocean currents," in *40th International Conference on Machine Learning (ICML)*, vol. 202, 2023, pp. 2113–2163.

[2] K. M. B. Lee, C. Yoo, B. Hollings, S. Anstee, S. Huang, and R. Fitch, "Online Estimation of Ocean Current from Sparse GPS Data for Underwater Vehicles," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 3443–3449.

[3] M. Popović, T. Vidal-Calleja, G. Hitz, I. Sa, R. Siegwart, and J. Nieto, "Multiresolution mapping and informative path planning for UAV-based terrain monitoring," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1382–1388.

[4] W. Chen, R. Khardon, and L. Liu, "Adaptive Robotic Information Gathering via non-stationary Gaussian processes," *The International Journal of Robotics Research*, vol. 43, no. 4, pp. 405–436, 2024. eprint: https://doi.org/10.1177/02783649231184498.

[5] A. Leininger, M. Ali, H. Jardali, and L. Liu, "Gaussian Process-based Traversability Analysis for Terrain Mapless Navigation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 10 925–10 931.

[6] C. Stachniss, C. Plagemann, and W. Burgard, "Gas distribution modeling using sparse Gaussian process mixture models," *Robotics: Science and Systems IV*, p. 310, 2009.

[7] B. Yang, R. Wang, J. Li, and T. Li, "Hierarchical Informative Path Planning for Multi-variable Monitoring in 3D Environmental Fields," in *2025 International Conference on Machine Learning, Computational Intelligence and Pattern Recognition (MLCIPR)*, 2025, pp. 379–384.

[8] M. Popović, T. Vidal-Calleja, J. J. Chung, J. Nieto, and R. Siegwart, "Informative path planning for active field mapping under localization uncertainty," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 10 751–10 757.

[9] F. Lindgren, H. Rue, and J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 4, pp. 423–498, 2011.

[10] K. Jakkala and S. Akella, "Multi-Robot Informative Path Planning from Regression with Sparse Gaussian Processes," in *2024 IEEE International Conference on Robotics and Automation, ICRA*, 2024.

[11] J. Wakulicz, K. M. Brian Lee, C. Yoo, T. Vidal-Calleja, and R. Fitch, "Informative Planning for Worst-Case Error Minimisation in Sparse Gaussian Process Regression," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 11 066–11 072, ISBN: 978-1-7281-9681-7.

[12] J. Quiñonero-Candela and C. E. Rasmussen, "A Unifying View of Sparse Approximate Gaussian Process Regression," *Journal of Machine Learning Research*, vol. 6, no. 65, pp. 1939–1959, 2005.

[13] A. G. Wilson and H. Nickisch, "Kernel interpolation for scalable structured Gaussian processes (KISS-GP)," *32nd International Conference on Machine Learning (ICML)*, vol. 3, pp. 1775–1784, 2015.

[14] T. Bui and R. Turner, "Tree-structured Gaussian Process Approximations," in *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[15] L. Sun, T. Vidal-Calleja, and J. V. Miro, "Bayesian fusion using conditionally independent submaps for high resolution 2.5D mapping," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 3394–3400.

[16] S. Nabarro, M. van der Wilk, and A. J. Davison, "A distributed Gaussian process model for multi-robot mapping," in *2026 IEEE International Conference on Robotics and Automation, ICRA*, 2026.

[17] C. Rasmussen and Z. Ghahramani, "Infinite Mixtures of Gaussian Process Experts," in *Advances in Neural Information Processing Systems*, vol. 14, 2001.

[18] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, "Deep Kernel Learning," in *19th International Conference on Artificial Intelligence and Statistics*, vol. 51, 2016, pp. 370–378.

[19] S. W. Ober, C. E. Rasmussen, and M. van der Wilk, "The promises and pitfalls of deep kernel learning," in *37th Conference on Uncertainty in Artificial Intelligence*, vol. 161, 2021, pp. 1206–1216.

[20] Q. Liu and D. Wang, "Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm," in *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[21] E. Jang, S. Gu, and B. Poole, "Categorical Reparameterization with Gumbel-Softmax," in *International Conference on Learning Representations (ICLR)*, 2017.

[22] C. J. Maddison, A. Mnih, and Y. W. Teh, "The Concrete Distribution: A continuous relaxation of discrete random variables," in *International Conference on Learning Representations (ICLR)*, 2017.

[23] NASA JPL, *NASA Shuttle Radar Topography Mission Global 1 arc second number*, 2013.

[24] M. L. Stein, *Interpolation of Spatial Data* (Springer Series in Statistics). Springer New York, 1999. Accessed: Jun. 4, 2026.

[25] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. 2006.