

© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Coupling Conditionally Independent Submaps for Large-Scale 2.5D Mapping with Gaussian Markov Random Fields

Liye Sun, Teresa Vidal-Calleja and Jaime Valls Miro

Abstract—Building large-scale 2.5D maps when spatial correlations are considered can be quite expensive, but there are clear advantages when fusing data. While optimal submapping strategies have been explored previously in covariance-form using Gaussian Process for large-scale mapping, this paper focuses on transferring such concepts into information form. By exploiting the conditional independence property of the Gaussian Markov Random Field (GMRF) models, we propose a submapping approach to build a nearly optimal global 2.5D map. In the proposed approach data is fused by first fitting a GMRF to one sensor dataset; then conditional independent submaps are inferred using this model and updated individually with new data arrives. Finally, the information is propagated from submap to submap to later recover the fully updated map. This is efficiently achieved by exploiting the inherent structure of the GMRF, fusion and propagation all in information form. The key contribution of this paper is the derivation of the algorithm to optimally propagate information through submaps by only updating the common parts between submaps. Our results show the proposed method reduces the computational complexity of the full mapping process while maintaining the accuracy. The performance is evaluated on synthetic data from the Canadian Digital Elevation Data.

I. INTRODUCTION

2.5D maps (occupancy, elevation, etc.) are compact representations of the environment. Although 3D maps can represent the environment more accurately, 2.5D maps are faster to access, require much less storage space and are suitable for navigation. This paper focuses on large-scale/high-resolution 2.5D mapping. One of the main challenges is how to fuse data at variable resolutions while incorporating sensor noise and data correlation in a statistically sound and computationally efficient way during both inference and fusion process.

The inherent structure of the environment captured by the sensor data could either be modelled in a covariance based approach - Gaussian Process (GP), as per our previous work [1], or an information matrix (inverse covariance) based method - Gaussian Markov Random Fields (GMRF), as per [2]. Spatial correlation is then encoded either in a semi-dense covariance matrix or in a sparse information matrix.

GP [3] is a powerful non-parametric Bayesian method. Unfortunately, its non-parametric nature causes high computations for large data sets. To build high-resolution maps, previously we split the whole map into conditionally inde-

pendent (CI) submaps based on data correlation, as per [4]. However, some correlation needs to be thrown away.

GMRF [5] has both the excellent analytical properties of the Gaussian distribution and the computational superiority due to its close link with numerical methods for sparse matrices. In [2], we combine GMRF model with Bayesian fusion to achieve a computation gain, by using numerical methods induced by the sparse information matrix. However, recovery of mean estimates and variances of the full map can incur in high computational complexity.

This paper speeds up [2] by adapting the covariance-form CI submapping strategy of [4] to information form. The motivation comes from that for the Gaussian distributed variables, there is a direct mapping between the sparse structure of the information matrix and the CI elements. The main contribution of this paper is, therefore, the closed-form solution to propagate information between CI submaps after fusion, all in information form. In this way there is a substantial reduction in computation and memory complexity for scalability. Moreover, recovering the mean and covariance for the whole map can be done independently per submap without incurring in any approximation. The proposed approach is applied to the scenario of high-resolution elevation mapping using synthetic data from Canadian Digital Elevation Data [6]. We evaluate the ability to fuse multi-resolution noisy data and the computational advantages.

The paper is organised as follows. Section II discusses some related works. Section III provides an overview of the proposed approach. Section IV reviews our previous work about GMRF for global mapping, and Section V introduces the information-form CI submapping approach. The experimental validation is presented in Section VI. The conclusion and plans are discussed in Section VII.

II. RELATED WORK

The problem of estimating probabilistic 2.5D maps have been extensively studied in the literature [7]. For example, [8] builds elevation maps and uses a scatter matrix to represent the local geometric uncertainty. In [9] multiple sensors' measurements are fused by using the sum of weighted variances of the uncorrelated data.

Instead of modelling each sample individually, GP characterises the spatial correlation via the covariance function [3]. Specifically, [10] adopts GP regression for terrain mapping and shows GP could handle the incompleteness of data by considering the correlation. In [11], data fusion in a GP framework is performed by considering data from different sensors as noisy samples of a regular GP. However,

All authors are with the Centre for Autonomous Systems at Faculty of Engineering and IT, University of Technology, Sydney NSW 2007, Australia. Liye.Sun@student.uts.edu.au, Teresa.VidalCalleja@uts.edu.au and Jaime.VallsMiro@uts.edu.au.

such approach scales cubically in the total amount of data. These works show that GP could generate a multi-resolution representation that incorporates and manages uncertainty in a statistically sound way and handles spatial correlation between data in an appropriate manner.

GP's high cost, $O(N^3)$ for training and $O(N^2M)$ for testing¹, has hindered exact inference for large-scale data. Various approximations have been developed to accommodate big N [12]. As an illustration, [10] uses a kd -tree to search for nearest training data of a query position but requires to build a new covariance matrix for each test position. Bayesian Committee Machine (BCM) [13] improves on independent local GPs by averaging the predictive inference of multiple GP experts, assuming the training subsets are conditionally independent given the test points. However BCM is a purely predictive approximation. Meanwhile, to account for big M it is common to use divide-and-conquer strategy. For instance, [14] divides both the training and test sets into subsets before applying local GPs. Some disadvantages include that the correlations between submaps are ignored, and information are used far more than once.

GMRF is discrete domain GP equipped with the Markov property. When the Markov model is fully connected, it is the Gaussian Process Random Field (GPRF) [15], which approximates the global GP by coupling local GPs via pairwise potential functions. But it takes in total $O(N^2m)$ for training, with local subset of maximal size of m . In the spatial statistic literature, [16] links the Matérn GP with the GMRF in a stochastic partial differential equation (SPDE) approach. The information matrix, generally used to parametrise GMRF, could be built in such a way that closely approximates the covariance matrix. The idea is to model the continuous domain as a Matérn GP, while doing all computations using the corresponding GMRF efficiently owing to the sparse information matrix.

Bayesian fusion lies in the core of probabilistic fusion. It computes the posterior probability distribution/density of the state given the set of measurements and the prior distribution. New sensor observations could be fused each time and update the probability density of the state estimates. When using Gaussian distributions and linear models, Bayesian fusion has an analytical solution that is optimal. Instead of regarding the individual variance of each sample, we show in [1] that incorporating the data correlation within Bayesian fusion facilitates to handle the incompleteness and inconsistency of multi-modal data. Nevertheless, the correlated Bayesian fusion quickly becomes intractable for large data sets, due to factorising the dense covariance matrix. One of the known solutions for the Simultaneous Mapping and Localisation (SLAM) problem is based on submapping algorithms, where in most cases independent submaps are considered. However, independent submaps ignore the correlations between each other producing a loose approximate solution. In [4], we borrow the idea of CI submaps from [17] and partition the GP predicted prior into CI submaps based on the length-scale

hyperparameter of the GP. We then conduct the covariance-form Bayesian fusion at the submap level. After building all submaps, the cross-correlation propagates through all submaps to obtain an updated global map.

Noteworthy, Bayesian fusion is "cheap" in information form [18]. Extensive examples in SLAM [19] and multi-sensor data fusion [20] show that fusion in information form is more efficient than its dual as only requires an addition operation. However, it is often costly to recover the state estimate (mean and variance), unless the information matrix is sparse. In [19], the sparse information matrix is maintained by using an undirected graph with local relative constraints (edges) between pairs of nearby points, which is a GMRF [21].

III. APPROACH OVERVIEW

The approach presented here considers information-form Bayesian fusion for CI submapping. First, we learn a GMRF model from a single source. Then to build the high-resolution map, we use the trained GMRF model to predict submaps that are conditionally independent, *i.e.*, an overlapping region between consecutive submaps exists. Each submap is represented by an information vector and an information matrix that depicts the dependencies amongst grid locations. The fusion step then occurs at the submap level. Finally, cross-correlation information needs to be propagated to the information vector and matrix.

Specifically, let $\Psi_1 = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$ be a dataset which contains n point-referenced measurements $y_i \in Y$ observed at the location $\mathbf{x}_i \in X$ ². Ψ_1 is taken from a latent process $y_i = \xi_i + \varepsilon_i$, where the underlying state $\xi_i = \xi(\mathbf{x}_i) \in \xi$, and ξ is Gaussian distributed. $\varepsilon = \{\varepsilon_i\}_{i=1}^N$, and $\varepsilon_i = \mathcal{N}(0, \sigma_\varepsilon^2)$. GMRF is used to infer the distribution of ξ from Ψ_1 and we get $p(\xi|X, \Psi_1)$. To handle a big amount of X^* , we divide X^* into CI regions $\{X_{s_i}^*\}_{i=1}^r$. $X_{s_1}^* \cap X_{s_2}^* = \emptyset$, where r is the total number of submaps. $X_{s_1}^* \cap X_{s_2}^* \neq \emptyset$, *i.e.* consecutive regions have an overlapping part. Note that the size of each region is directly derived based on the blocks of the information matrix of the GMRF. Then GMRF is used to predict the distribution of each submap, denoted as $p(\xi_{s_i}|X_{s_i}^*) = \mathcal{N}^{-1}(\boldsymbol{\eta}_{s_i}, \mathbf{Q}_{s_i})$ ³. This prior map gives us the ability to increase or decrease map resolution.

Given another dataset $\Psi_2 = \{(\mathbf{x}_1, z_1), (\mathbf{x}_2, z_2), \dots, (\mathbf{x}_k, z_k)\}$, with k sensor readings z_i measured from $\mathbf{x}_i \in X^*$ locations. $\mathbf{z} = H\xi + R$, H is the observation matrix and R is a diagonal matrix representing the iid Gaussian noise $\boldsymbol{\omega} = \mathcal{N}(0, \sigma_z^2)$. Each submap density $p(\mathbf{z}_{s_i}|\xi_{s_i}, X_{s_i}^*)$ could be obtained via marginalisation from the whole map.

It is then straightforward to use Ψ_2 to update the prior submap. By applying a maximum *a posteriori* estimator we get the posterior submap $\{\mathcal{N}^{-1}(\boldsymbol{\eta}_{s_i}^+, \mathbf{Q}_{s_i}^+)\}_{i=1}^r$ based on Bayes' rule $p(\xi_{s_i}|\mathbf{z}_{s_i}, X_{s_i}^*) \propto p(\xi_{s_i}|X_{s_i}^*)p(\mathbf{z}_{s_i}|\xi_{s_i}, X_{s_i}^*)$. This is

²We assume measurements' global locations are known and accurate.

³ $p(\xi_{s_i}|X_{s_i}^*)$ denotes $p(\xi_{s_i}|X_{s_i}^*, \Psi_1)$ for short. \mathcal{N}^{-1} shows the Gaussian in information form; let $p(\xi_{s_i}|X_{s_i}^*) = \mathcal{N}(\boldsymbol{\mu}_{s_i}, \boldsymbol{\Sigma}_{s_i})$, then $\mathbf{Q}_{s_i} = \boldsymbol{\Sigma}_{s_i}^{-1}$, $\boldsymbol{\eta}_{s_i} = \mathbf{Q}_{s_i}\boldsymbol{\mu}_{s_i}$.

¹ N and M are the number of training and test points

correlated Bayesian fusion because by incorporating the dependencies encoded in Q_{s_i} during fusion, one single measurement could update all its neighbours, compared with standard point-to-point Bayesian fusion where Q_{s_i} is diagonal.

The next step is *correlation propagation*, i.e., to propagate the influence of new observations backwards to all submaps. This method is based on the CI property between submaps.

Finally, the global estimates of the mean state vector and the variances can be efficiently recovered within each submap. Note that there is no need to factorise the global information matrix for recovery.

IV. GMRF FOR FUSION IN INFORMATION FORM

A GMRF is defined on a set of discrete sites connected by a graph \mathcal{G} . ξ is a GMRF with reference to \mathcal{G} with mean μ and information matrix Q if its pdf is in the below form:

$$p(\xi) = (2\pi)^{-n/2} |Q|^{1/2} \exp \left\{ -\frac{1}{2} (\xi - \mu)^\top Q (\xi - \mu) \right\} \quad (1)$$

where Q encodes the CI property: $Q_{ij} = 0$ if and only if ξ_i and ξ_j are conditionally independent [22]. The continuously indexed GMRF [16] closely approximates the Matérn GP [3]. Q is parameterised based on its direct link to the stochastic partial differential equation (SPDE).

We train the GMRF model on Ψ_1 , and then predict the map $p(\xi|X^*) = \mathcal{N}^{-1}(\eta, Q)$ for query points X^* , where

$$\eta = Q(X^*, X) \mathbf{y} \quad (2a)$$

$$Q = Q(X^*, X^*) + Q_\varepsilon. \quad (2b)$$

Q_ε is a diagonal matrix with diagonal values to be $(\sigma_\varepsilon^2)^{-1}$.

For fusion at the global level, we update the prior map (2) with observations from Ψ_2 based on Bayes rule, and the posterior map $p(\xi|\mathbf{z}, X^*) = \mathcal{N}^{-1}(\eta^+, Q^+)$ is

$$\eta^+ = \eta + H^\top Q_z \mathbf{z} \quad (3a)$$

$$Q^+ = Q + H^\top Q_z H. \quad (3b)$$

H is the observation matrix, which selects part of state ξ that is observed by \mathbf{z} . Q_z is a diagonal matrix with diagonal values to be $(\sigma_z^2)^{-1}$.

We then solve a sparse linear system $Q^+ \mu = \eta^+$ to get μ . Q^+ will be re-ordered before Cholesky factorisation $Q^+ = LL^\top$. Then we solve two triangular systems. For $\mathbf{x}_i \in \mathbb{R}^2$, it takes at least $O(M^{3/2})$ to re-order the $M \times M$ matrix Q^+ , and the re-ordered Cholesky factor at least has $O(M \log M)$ fill-ins [2]. Therefore, considering all the correlations within one sparse information matrix will become impractical when we aim at high-resolution maps. This is the reason why in this work we up to split into submaps as explained in Section V.

V. CI SUBMAPPING

A. Global dependencies

We now explain the global dependence using the example in Fig. 1. For $i \neq j$, submap s_i is correlated with submap s_j ; accordingly, the covariance matrix is dense. By contrast, for $i \neq j$, s_i and s_j are *conditionally independent* given the

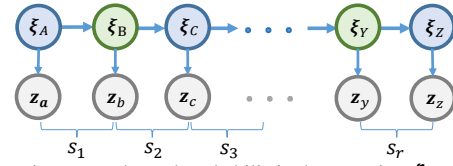


Fig. 1: Bayesian networks and probabilistic dependencies. ξ_A, \dots, ξ_Z represent a set of components A, \dots, Z of the state at some positions in the 2.5D map, respectively. $\mathbf{z}_a, \dots, \mathbf{z}_z$ denotes the set of sensor measurements related to these positions. s_1, \dots, s_r denote submaps. Two consecutive submaps contain two sets of components that share a common part.

rest of submaps⁴; hence, $Q_{s_i, s_j} = 0$ in the information matrix Q . Consequently, Q has most of the elements being zero. The only non-zero elements are $j = i - 1, i, i + 1$, namely Q is a block tridiagonal matrix. This shows that while the covariance matrix models the independence of variables, the information matrix models the conditional independence; indeed, the CI property is naturally reflected in the sparse pattern of the information matrix.

Fig.2(a) gives a schematic view of Q . Surely, there is no need to store or compute elements known to be zero. Although Q is sparse, computations involving the whole Q can be expensive (see Section IV). Ideally, we would do all computations at submap-level, while being able to recover the optimal global map. However, we could not directly adapt the approach in Section IV to submaps for two reasons: (1) submap-level fusion confines new information \mathbf{z}_{s_i} within one submap s_i ; (2) spatial correlations between X_{s_i} and X_{s_j} , where $i \neq j$, are lost during GMRF prediction. These problems also emerge in the covariance-form CI submapping method [4], nevertheless, they can be solved via the correlation propagation algorithm presented in the following section.

B. Local dependencies and CI submapping

Without loss of generality, we will take two submaps s_1 and s_2 as an illustration to explain the proposed correlation propagation algorithm. We assume s_1 and s_2 together form the full map. Let the state vector for each submap be $\xi_{s_1} = [\xi_a, \xi_b]$ and $\xi_{s_2} = [\xi_b, \xi_c]$.

First, the prior submaps $p(\xi_{s_1}|X_{s_1}^*) = \mathcal{N}^{-1}(\eta_{s_1}, Q_{s_1})$ and $p(\xi_{s_2}|X_{s_2}^*) = \mathcal{N}^{-1}(\eta_{s_2}, Q_{s_2})$ are predicted by the learned GMRF model. Then, each prior submap is updated with the observations $\mathbf{z}_{s_1} = \mathbf{z}_a$ and $\mathbf{z}_{s_2} = [\mathbf{z}_b, \mathbf{z}_c]$ using the maximum *a posteriori* estimator. Note that to avoid double counting information, only the non-common part of each set of measurements is used to update each submap, except for last submap where both are incorporated. We denote the posterior submaps $p(\xi_{s_1}|\mathbf{z}_{s_1}, X_{s_1}^*)$ and $p(\xi_{s_2}|\mathbf{z}_{s_2}, X_{s_2}^*)$ as⁵

$$p(\xi_A, \xi_B|\mathbf{z}_{s_1}) = \mathcal{N}^{-1} \left(\begin{bmatrix} \hat{\eta}_A^a \\ \hat{\eta}_B^a \end{bmatrix}, \begin{bmatrix} Q_{AA}^a & Q_{AB}^a \\ Q_{BA}^a & Q_{BB}^a \end{bmatrix} \right) \quad (4)$$

$$p(\xi_B, \xi_C|\mathbf{z}_{s_2}) = \mathcal{N}^{-1} \left(\begin{bmatrix} \hat{\eta}_B^{bc} \\ \hat{\eta}_C^{bc} \end{bmatrix}, \begin{bmatrix} Q_{BB}^{bc} & Q_{BC}^{bc} \\ Q_{CB}^{bc} & Q_{CC}^{bc} \end{bmatrix} \right), \quad (5)$$

⁴This is precisely the same independence as the pairwise Markov independence [22].

⁵ $p(\xi_A, \xi_B|\mathbf{z}_{s_1})$ denotes $p(\xi_A, \xi_B|X_{s_1}^*, \mathbf{z}_{s_1})$ for short; and so on.

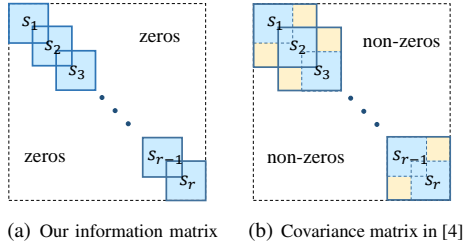


Fig. 2: Schematic representation of the entries of the (a) information matrix and (b) covariance matrix during CI submapping [4]. The blue areas show the non-zeros entries in our information matrix. The yellow areas are the extra non-zero entries that need to be computed during [4].

where the superscript indicates the set of observations that have been incorporated after Bayesian fusion, and the subscript represents the state components in submaps.

Now our goal is to recover the optimal joint distribution⁶:

$$p(\xi_A, \xi_B, \xi_C | z_{s_1}, z_{s_2}) = \mathcal{N}^{-1} \left(\begin{bmatrix} \hat{\eta}_A^{abc} \\ \hat{\eta}_B^{abc} \\ \hat{\eta}_C^{abc} \end{bmatrix}, \begin{bmatrix} Q_{AA}^{abc} & Q_{AB}^{abc} & 0 \\ Q_{BA}^{abc} & Q_{BB}^{abc} & Q_{BC}^{abc} \\ 0 & Q_{CB}^{abc} & Q_{CC}^{abc} \end{bmatrix} \right). \quad (6)$$

Q_{AC}^{abc} and Q_{CA}^{abc} are zero based on local CI property:

$$\xi_A, z_A \perp \xi_C, z_C | \xi_B. \quad (7)$$

This is because the shared part ξ_B is a separating subset for s_1 and s_2 . Given ξ_B , s_1 and s_2 do not carry any additional information about each other.

To couple s_1 and s_2 after fusion, the new information z_{s_2} need to propagate to s_1 . The correlation propagation algorithm is obtained following the deduction in the Appendix. Basically, this algorithm is obtained using the marginalisation and conditioning operations for Gaussian distributions and the CI property $p(\xi_A | \xi_B, \xi_C, z_a, z_b, z_c) = p(\xi_A | \xi_B, z_a)$ and $p(\xi_C | \xi_A, \xi_B, z_a, z_b, z_c) = p(\xi_C | \xi_B, z_b, z_c)$. Here we only show the conclusion:

$$\begin{aligned} 1: \hat{\eta}_B^{abc} &= \hat{\eta}_B^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} \hat{\eta}_A^a; & Q_{BB}^{abc} &= Q_{BB}^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} Q_{AB}^a; \\ 2: \hat{\eta}_A^{abc} &= \hat{\eta}_A^a; & Q_{AA}^{abc} &= Q_{AA}^a; & Q_{AB}^{abc} &= Q_{AB}^a; & Q_{BC}^{abc} &= Q_{BC}^{bc}; \\ \hat{\eta}_C^{abc} &= \hat{\eta}_C^{bc}; & Q_{CC}^{abc} &= Q_{CC}^{bc}. \end{aligned}$$

Note that only the information vector and matrix of the common elements between consecutive submaps are required to be re-computed. On the contrary, both the non-overlapping and the cross-correlated parts need to be updated in its dual form [4], which uses the estimation difference in common components to correct the state and thus enforcing the cross-correlation backwards. Generally, each submap could be corrected sequentially. A schematic view is shown in Fig.2.

C. Map recovery

After correcting all submaps, one can recover the global map in covariance form. Note that because of the global CI dependence (see Section V-A), there is no need to solve

⁶Note the information matrix of local map $p(\xi_B, \xi_C | z_{s_1}, z_{s_2})$ is not identical to the corresponding block in the global map $p(\xi_A, \xi_B, \xi_C | z_{s_1}, z_{s_2})$. Although it is true in covariance form.

for the full map, and the recovery takes place locally. This allows opportunities for parallelisation. For each updated submap $\{\hat{\eta}, Q\}$, the mean and variance are recovered using the method proposed in [23], which requires Cholesky factorisation. To reduce fill-ins in the Cholesky factor, we use the bandwidth reduction method [24] to re-order Q since we have long and thin graphs for each submap. Otherwise nested dissection re-ordering [25] could be employed. In 2.5D mapping, band re-ordering requires $O(N^2)$ flops and nested dissection takes $O(N^{3/2})$, N is the size of η . After reordering, each local information matrix is factorised $Q = LL^T$, which typically consumes the most memory and time. Nonetheless, at the submap level, we have a small and re-ordered Q which can be factorised in $O(N^{3/2})$. After that, we obtain L which inherits the non-zero pattern from Q . Then, we solve for μ using the forwards and backwards substitutions. For bandwidth L with a few non-zero entries far-off the diagonal, recovery is done in $O(N)$ using the method presented in [26].

VI. EXPERIMENT RESULTS

We evaluate the ability of three former methods and the proposed approach to generate accurate high-resolution elevation maps, via Bayesian fusion of synthetic terrain data from different resolutions.

A. Benchmarking methods

CGM (optimal) - Covariance-form Global Mapping [1]: we model Ψ_1 via a Matérn GP, and then use this GP to predict the high-resolution map, which is then fused with Ψ_2 . The spatial correlation is incorporated via the covariance matrix. We regard this method as optimal in the sense that it considers all the cross-correlations.

IGM - Information-form Global Mapping [2]: we model Ψ_1 via the continuously indexed GMRF, and then use GMRF to predict a high-resolution map, which is later fused with Ψ_2 . The spatial correlation is incorporated in the sparse information matrix.

CCIS - Covariance-form CI Submapping [4]: we model Ψ_1 via a Matérn GP, and then use this GP to predict each submap. The predicted submaps are sequentially fused with disjoint observations from Ψ_2 . Finally, we will propagate information to other submaps to obtain the global map.

B. Dataset

A challenging scenario is simulated with data incompleteness, high uncertainty and inconsistency. Synthetic elevation data is generated from Canadian Digital Elevation Data (CDED) [6] in a similar manner as in [27]. CDED consists of the ground elevations at regularly spaced intervals. We randomly choose one map which covers the area of -116 to -114° longitude and 56 to 57° latitude. There are 9608×4804 grid cells in this map, and cell coverage is $23\text{m} \times 23\text{m}$. The elevation values range between $[450\text{m}, 830\text{m}]$.

We first down-sample the original map into a 100×200 grid map, which is regarded as groundtruth (GT). To get Ψ_1 , we randomly sample 1476 points from the original map and

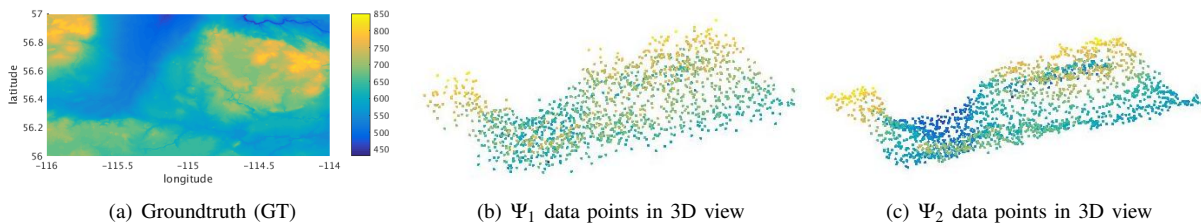


Fig. 3: Groundtruth and two synthetic datasets.

set noise level to be $\sigma_\epsilon = 45.6\text{m}$ (12% of the elevation range). To obtain Ψ_2 , we randomly sample 1960 points from the original map and set the noise level to be $\sigma_z = 11.4\text{m}$ (3% of the elevation range). Therefore we obtain two heterogeneous, noisy and sometimes inconsistent datasets.

2.5D and 3D plots of GT, Ψ_1 and Ψ_2 are shown in Fig.3. Elevation values are shown in colour. In 2.5D plots, the vertical axis corresponds to the latitude and horizontal axis is the longitude. In 3D plots, we enlarge the axis ratio of the 3rd dimension to make the elevation change visible.

C. Evaluation

Experimental results are shown in Figures 4, 5, 6 and 7. For simplicity, we ignore the axis labels. The figures show the correlation propagation, either in covariance or information form, correct the maps which closely approximates the optimal global solution. Both GP and GMRF, which are trained using only 0.0032% of the original map, could correctly model the global trend. GP includes all correlations resulting in a smooth map. GMRF is conservative in the prediction, therefore the uncertainty is bigger than the GP as shown in the figure. CGM posterior map Fig.4(c) varies smoothly, while IGM depicts local variations. Fig.5(c) clearly shows the bottom-right peak and mid-left valley. As expected the proposed method approximates quite well IGM, but with substantial gain in speed as shown in Table I.

Table I compares the computational complexity of all algorithms. The proposed method shows significant advantages over all others. It speeds up IGM by exploiting the GMRF properties and CI submaps. It is usually faster than CCIS and CGM which involve factorising the dense covariance matrix.

TABLE I: Computational Complexity of Each Step

	Learning	Prediction	Bayesian Fusion	Reordering + Recovery
CGM	$O(N^3)$	$O(N^2M)$	$O(M^3)$	/
IGM	$O(N^{3/2})$	$O(M)$	$O(M)$	$O(M^2 + M)$
CCIS	$O(N^3)$	$O(N^2m)$	$O(Mm^2)$	/
Our method	$O(N^{3/2})$	$O(m)$	$O(M)$	$O(m^{3/2} + m)$

N , M and m are the size of training data, test data, and the submap, respectively. $m \ll M$. For high-resolution maps, $N \lesssim M$. Our approach allows for parallel computation for submaps prediction, re-ordering and recovery. Bandwidth reduction permutation method is used in IGM.

Table II quantifies the results in terms of Root Mean Squared Error (RMSE). The average and standard deviation of the absolute RMSE of a 5-runs Monte Carlo Simulation for all methods is presented in this table. The accuracy obtained by our approach is comparable with IGM. Moreover,

as expected, the GP based-methods are more accurate than both GMRFs. The table also shows clearly that the correlation propagation step greatly improves the final results.

TABLE II: RMSE \pm std

	CGM	IGM	CCIS	Our method
Prior map	20.61 \pm 0.93	26.67 \pm 0.96	15.49 \pm 0.94	19.31 \pm 0.98
Final map	9.36 \pm 0.14	12.48 \pm 0.32	10.65 \pm 0.20	13.92 \pm 0.43

The unit is metre (m). RMSE for dataset Ψ_1 is 46.2m; for Ψ_2 is 7.7m. For CCIS and our method, "prior map" refers to submap-level fusion results.

VII. CONCLUSION AND FUTURE WORK

This paper proposes an efficient information-form CI submapping method via fusing multi-modal uncertain data in a spatially correlated and statistically sound way. GMRF, when combined with CI property, significantly accelerates the submapping albeit closely approximates the optimal global solution. The new information can be propagated to all submaps to correct the whole map without loss of information. Almost all computations are done at a local level (submaps) and thus efficiently.

A key finding is that in covariance form, the information propagated from submap to submap uses the estimation difference in the overlapping areas between submaps to correct the non-overlapping and the cross-correlated areas [4]. While in information form, only the overlapping areas need to be corrected thanks to the CI property.

Results on a public elevation map datasets show that, when compared with three benchmark approaches, the presented approach exhibits considerable computational gain while maintaining competitive accuracy, making it appealing for fast large-scale 2.5D mapping.

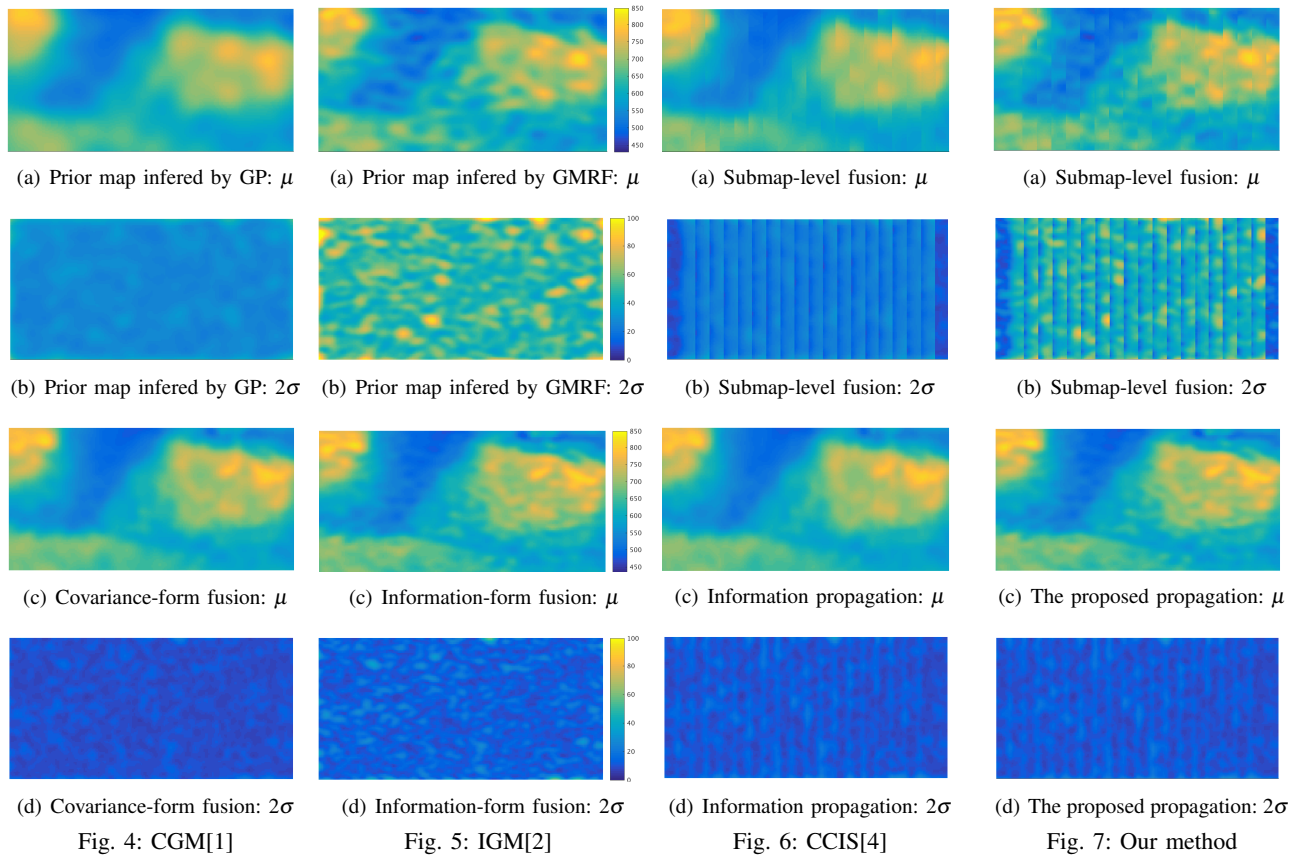
Further plans include extending this work for online incremental submapping, and adapting it to situations when non-consecutive submaps overlap.

ACKNOWLEDGMENT

This work is an outcome from the Critical Pipes Project funded by Sydney Water, Water Research Foundation USA, Melbourne Water, Water Corporation(WA), UK Water Industry Research Ltd, South Australia Water, South East Water, Hunter Water, City West Water, Monash University, University of Technology Sydney and University of Newcastle.

REFERENCES

- [1] T. Vidal-Calleja, D. Su, F. De Bruijn, and J. V. Miro, "Learning spatial correlations for Bayesian fusion in pipe thickness mapping," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014, pp. 683–690.



[2] L. Sun, T. Vidal-Calleja, and J. Valls Miro, "Gaussian markov random fields for fusion in information form," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2016, pp. 1840–1845.

[3] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, Mass.: MIT Press, 2006.

[4] L. Sun, T. Vidal-Calleja, and J. Valls Miro, "Bayesian fusion using conditionally independent submaps for high resolution 2.5d mapping," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2015, pp. 3394–3400.

[5] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*. CRC Press, 2005.

[6] "Natural Resources Canada. Laval, Quebec." <http://www.geobase.ca/>, Ottawa: Natural Resources Canada, Geomatics Canada, 1995.

[7] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[8] A. Kelly and A. Stentz, "Rough terrain autonomous mobility—part 2: An active vision, predictive control approach," *Auton. Robots*, vol. 5, no. 2, pp. 163–198, 1998.

[9] P. Pfaff, R. Triebel, and W. Burgard, "An efficient extension to elevation maps for outdoor terrain mapping and loop closing," *I. J. Robotic Res.*, vol. 26, no. 2, pp. 217–230, 2007.

[10] S. Vasudevan, F. Ramos, E. Nettleton, and H. Durrant-Whyte, "Gaussian process modeling of large-scale terrain," *Journal of Field Robotics*, vol. 26, no. 10, pp. 812–840, 2009.

[11] —, "Heteroscedastic gaussian processes for data fusion in large scale terrain modeling," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, 2010, pp. 3452–3459.

[12] J. Quiñero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate gaussian process regression," *J. Machine Learning Res.*, vol. 6, no. Dec, pp. 1939–1959, 2005.

[13] V. Tresp, "A bayesian committee machine," *Neural Computation*, vol. 12, no. 11, pp. 2719–2741, 2000.

[14] S. Kim and J. Kim, "Continuous occupancy maps using overlapping local gaussian processes," in *IEEE/RSJ Int. Conf. Intel. Robot. Syst.*, 2013, pp. 4709–4714.

[15] D. Moore and S. J. Russell, "Gaussian process random fields," in *Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 3357–3365.

[16] F. Lindgren, H. Rue, and J. Lindström, "An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach," *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, vol. 73, no. 4, pp. 423–498, 2011.

[17] P. Piniés and J. D. Tardós, "Large-Scale SLAM building conditionally independent local maps: Application to monocular vision," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1094–1106, 2008.

[18] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[19] M. R. Walter, R. M. Eustice, and J. J. Leonard, "Exactly sparse extended information filters for feature-based slam," *Int. J. Robot. Res.*, vol. 26, no. 4, pp. 335–359, 2007.

[20] H. Durrant-Whyte and T. Henderson, *Multisensor Data Fusion*, 2008, ch. 26, pp. 585–610.

[21] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural computation*, vol. 13, no. 10, pp. 2173–2200, 2001.

[22] D. Koller and N. Friedman, "Gaussian network models," in *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: MIT Press, 2009.

[23] G. H. Golub and R. J. Plemmons, "Large-scale geodetic least-squares adjustment by dissection and orthogonal decomposition," *Linear Algebra Appl.*, vol. 34, pp. 3–28, 1980.

[24] E. Cuthill and J. McKee, "Reducing the bandwidth of sparse symmetric matrices," in *Proceedings of the 1969 24th National Conference*, 1969, pp. 157–172.

[25] A. George, "Nested dissection of a regular finite element mesh," *SIAM J. Numerical Anal.*, vol. 10, no. 2, pp. 345–363, 1973.

[26] G. H. Golub and C. F. Van Loan, "Matrix computations. 1996," *Johns Hopkins University, Press, Baltimore, MD, USA*, pp. 374–426, 1996.

[27] M. P. Gerardo-Castro, T. Peynot, F. Ramos, and R. Fitch, "Non-parametric consistency test for multiple-sensing-modality data fusion," in *IEEE Int. Conf. on Inform. Fusion*, 2015, pp. 443–451.

APPENDIX
CORRELATION PROPAGATION

This appendix deduces the proposed information-form correlation propagation algorithm.

Given the first submap (4) and the second submap (5), the goal is to obtain the optimal global map (6).

Based on CI property, as in (7), we get

$$p(\xi_A | \xi_B, \xi_C, z_a, z_b, z_c) = p(\xi_A | \xi_B, z_a) \quad (8)$$

$$p(\xi_C | \xi_A, \xi_B, z_a, z_b, z_c) = p(\xi_C | \xi_B, z_b, z_c) \quad (9)$$

In (8), the left term can be obtained from the global map (6) by first marginalising out ξ_C and then conditioning on ξ_B . First, we marginalise out ξ_C and we get

$$p(\xi_A, \xi_B | z_{s_1}, z_{s_2}) \triangleq \mathcal{N}^{-1}(\hat{\eta}_{AB}^{*,abc}, Q_{AB}^{*,abc}), \quad (10)$$

where

$$\begin{aligned} \hat{\eta}_{AB}^{*,abc} &= \begin{bmatrix} \hat{\eta}_A^{abc} \\ \hat{\eta}_B^{abc} \end{bmatrix} - \begin{bmatrix} 0 \\ Q_{BC}^{abc} \end{bmatrix} (Q_{CC}^{abc})^{-1} \hat{\eta}_C^{abc} \\ &= \begin{bmatrix} \hat{\eta}_A^{abc} \\ \hat{\eta}_B^{abc} - Q_{BC}^{abc} (Q_{CC}^{abc})^{-1} \hat{\eta}_C^{abc} \end{bmatrix}, \end{aligned} \quad (11a)$$

$$\begin{aligned} Q_{AB}^{*,abc} &= \begin{bmatrix} Q_{AA}^{abc} & Q_{AB}^{abc} \\ Q_{BA}^{abc} & Q_{BB}^{abc} \end{bmatrix} - \begin{bmatrix} 0 \\ Q_{BC}^{abc} \end{bmatrix} (Q_{CC}^{abc})^{-1} \begin{bmatrix} 0 & Q_{CB}^{abc} \end{bmatrix} \\ &= \begin{bmatrix} Q_{AA}^{abc} & Q_{AB}^{abc} \\ Q_{BA}^{abc} & Q_{BB}^{abc} - Q_{BC}^{abc} (Q_{CC}^{abc})^{-1} Q_{CB}^{abc} \end{bmatrix}. \end{aligned} \quad (11b)$$

Then, when conditioning (10) on ξ_B , we get

$$p(\xi_A | \xi_B, z_{s_1}, z_{s_2}) = \mathcal{N}^{-1}(\hat{\eta}_A^{abc} - Q_{AB}^{abc} \xi_B, Q_{AA}^{abc}). \quad (12)$$

Meanwhile, the right term $p(\xi_A | \xi_B, z_{s_1})$ in (8) can be obtained from the first submap (4) by conditioning on ξ_B

$$p(\xi_A | \xi_B, z_{s_1}) = \mathcal{N}^{-1}(\hat{\eta}_A^a - Q_{AB}^a \xi_B, Q_{AA}^a). \quad (13)$$

According to CI property (8), (12) equals to (13), thus

$$\hat{\eta}_A^{abc} - Q_{AB}^{abc} \xi_B = \hat{\eta}_A^a - Q_{AB}^a \xi_B \quad (14a)$$

$$Q_{AA}^{abc} = Q_{AA}^a. \quad (14b)$$

Besides, since ξ_A and z_{s_2} are conditionally independent

$$\hat{\eta}_A^{abc} = \hat{\eta}_A^a. \quad (15)$$

By substituting (15) into (14a), we get:

$$Q_{AB}^{abc} = Q_{AB}^a. \quad (16)$$

In (9), the left term can be obtained from the global map by first marginalising out ξ_A and then conditioning on ξ_B . We first marginalise out ξ_A and get

$$p(\xi_B, \xi_C | z_{s_1}, z_{s_2}) \triangleq \mathcal{N}^{-1}(\hat{\eta}_{BC}^{*,abc}, Q_{BC}^{*,abc}), \quad (17)$$

where

$$\begin{aligned} \hat{\eta}_{BC}^{*,abc} &= \begin{bmatrix} \hat{\eta}_B^{abc} \\ \hat{\eta}_C^{abc} \end{bmatrix} - \begin{bmatrix} Q_{BA}^{abc} \\ 0 \end{bmatrix} (Q_{AA}^{abc})^{-1} \hat{\eta}_A^{abc} \\ &= \begin{bmatrix} \hat{\eta}_B^{abc} - Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} \hat{\eta}_A^{abc} \\ \hat{\eta}_C^{abc} \end{bmatrix} \end{aligned} \quad (18a)$$

$$\begin{aligned} Q_{BC}^{*,abc} &= \begin{bmatrix} Q_{BB}^{abc} & Q_{BC}^{abc} \\ Q_{CB}^{abc} & Q_{CC}^{abc} \end{bmatrix} - \begin{bmatrix} Q_{BA}^{abc} \\ 0 \end{bmatrix} (Q_{AA}^{abc})^{-1} \begin{bmatrix} Q_{AB}^{abc} & 0 \end{bmatrix} \\ &= \begin{bmatrix} Q_{BB}^{abc} - Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} Q_{AB}^{abc} & Q_{BC}^{abc} \\ Q_{CB}^{abc} & Q_{CC}^{abc} \end{bmatrix}. \end{aligned} \quad (18b)$$

When (17) is conditioned on ξ_B ,

$$p(\xi_C | \xi_B, z_{s_1}, z_{s_2}) = \mathcal{N}^{-1}(\hat{\eta}_C^{abc} - Q_{CB}^{abc} \xi_B, Q_{CC}^{abc}). \quad (19)$$

Meanwhile, the right term in (9) could be computed from (5) by conditioning on ξ_B that

$$p(\xi_C | \xi_B, z_{s_2}) = \mathcal{N}^{-1}(\hat{\eta}_C^{bc} - Q_{CB}^{bc} \xi_B, Q_{CC}^{bc}). \quad (20)$$

According to CI property (9), (19) is equal to (20), thus

$$\hat{\eta}_C^{abc} - Q_{CB}^{abc} \xi_B = \hat{\eta}_C^{bc} - Q_{CB}^{bc} \xi_B \quad (21a)$$

$$Q_{CC}^{abc} = Q_{CC}^{bc}. \quad (21b)$$

In SLAM literature, (17) and (5) are equal if s_2 is initialised with the common elements $\xi_B | z_a$ from s_1 . However, we could not initialise s_2 in this way because we do not have a transition model. Therefore we make the only assumption in this approach that they two are equal, and this assumption makes the global map sub-optimal. Then we obtain

$$\begin{bmatrix} \hat{\eta}_B^{bc} \\ \hat{\eta}_C^{bc} \end{bmatrix} = \begin{bmatrix} \hat{\eta}_B^{abc} - Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} \hat{\eta}_A^{abc} \\ \hat{\eta}_C^{abc} \end{bmatrix} \quad (22a)$$

$$\begin{bmatrix} Q_{BB}^{bc} & Q_{BC}^{bc} \\ Q_{CB}^{bc} & Q_{CC}^{bc} \end{bmatrix} = \begin{bmatrix} Q_{BB}^{abc} - Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} Q_{AB}^{abc} & Q_{BC}^{abc} \\ Q_{CB}^{abc} & Q_{CC}^{abc} \end{bmatrix}. \quad (22b)$$

Then by comparing each entries in (22a) and (22b), we get

$$\hat{\eta}_B^{abc} = \hat{\eta}_B^{bc} + Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} \hat{\eta}_A^{abc} \quad (23a)$$

$$Q_{BB}^{abc} = Q_{BB}^{bc} + Q_{BA}^{abc} (Q_{AA}^{abc})^{-1} Q_{AB}^{abc}. \quad (23b)$$

When substituting (15) and (16) into (23a) and (23b), we get

$$\hat{\eta}_B^{abc} = \hat{\eta}_B^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} \hat{\eta}_A^a \quad (24a)$$

$$Q_{BB}^{abc} = Q_{BB}^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} Q_{AB}^a. \quad (24b)$$

In conclusion, given the former submap, and the current submap that is optimal in the sense that it contains all the currently available observations, we could obtain the optimal global map using below algorithm

$$\hat{\eta}_A^{abc} = \hat{\eta}_A^a \quad (25a)$$

$$\hat{\eta}_B^{abc} = \hat{\eta}_B^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} \hat{\eta}_A^a \quad (25b)$$

$$\hat{\eta}_C^{abc} = \hat{\eta}_C^{bc} \quad (25c)$$

$$Q_{AA}^{abc} = Q_{AA}^a \quad (25d)$$

$$Q_{AB}^{abc} = Q_{AB}^a \quad (25e)$$

$$Q_{BB}^{abc} = Q_{BB}^{bc} + Q_{BA}^a (Q_{AA}^a)^{-1} Q_{AB}^a \quad (25f)$$

$$Q_{BC}^{abc} = Q_{BC}^{bc} \quad (25g)$$

$$Q_{CC}^{abc} = Q_{CC}^{bc}. \quad (25h)$$