A Trust-based Mixture of Gaussian Processes Model for Reliable Regression in Participatory Sensing

Qikun Xiang
Nanyang Technological University
qxiang003@e.ntu.edu.sg

Jie Zhang
Nanyang Technological University
zhangj@ntu.edu.sg

Ido Nevat
TUMCREATE
ido.nevat@tum-create.edu.sg

Pengfei Zhang
University of Oxford
pengfei@robots.ox.ac.uk

Abstract

Data trustworthiness is a crucial issue in real-world participatory sensing applications. Without considering this issue, different types of worker misbehavior, especially the challenging collusion attacks, can result in biased and inaccurate estimation and decision making. We propose a novel trust-based mixture of Gaussian processes (GP) model for spatial regression to jointly detect such misbehavior and accurately estimate the spatial field. We develop a Markov chain Monte Carlo (MCMC)-based algorithm to efficiently perform Bayesian inference of the model. Experiments using two real-world datasets show the superior robustness of our model compared with existing approaches.

1 Introduction

Recently, crowdsourcing has become a viable alternative to outsourcing. Crowdsourcing platforms, e.g. Amazon Mechanical Turk (https://www.mturk.com) and CrowdFlower (https://www.crowdflower.com), have demonstrated the advantages of crowdsourcing, such as being fast and inexpensive. One notable crowdsourcing application is participatory sensing, in which workers collect sensory information of spatial phenomena (e.g. temperature, noise, air pollution, etc.) via mobile devices [Zenonos *et al.*, 2015]. Through collected data, the field of spatial phenomenon can be estimated at any given point in space via regression. However, trustworthiness of collected data is a crucial issue [Mousa *et al.*, 2015]. Faulty sensors, inappropriate measurements, and worker misbehaviors (especially the challenging collusion attacks), all result in unreliable (erroneous or malicious) data. Without considering this issue, estimations can be negatively affected, becoming biased and inaccurate.

Related Work. A number of approaches have been proposed to address the reliability issue in crowdsourcing and participatory sensing. They are related to different types of tasks including object labelling, rating-based crowdsourcing, parameter estimation, and spatial field regression.

In object labelling, workers are given images of objects, such as text that contains mistakes [Tran-Thanh *et al.*, 2015] and celestial objects [Kamar *et al.*, 2012], and are asked to classify them. There is usually a unique true answer for each question, and each worker submits his/her guess of the answer. In rating-based crowdsourcing, workers are asked to provide subjective opinion about text, videos, or events [Tarasov *et al.*, 2014]. These questions have no unique answers. In parameter estimation, the objective is to estimate continuous-valued quantities, such as spatial locations of target objects [Salek *et al.*, 2013] and Wi-Fi hotspots [Venanzi *et al.*, 2015]. Workers report noisy (and possibly erroneous) observations of these quantities, which are aggregated to produce estimations. The most challenging task type in crowdsourcing and participatory sensing is spatial field regression, in which the objective is to estimate a continuous spatial field at every spatial location. In such tasks, workers report noisy and potentially erroneous observations of spatial fields (e.g. urban noise and air quality [Zenonos *et al.*, 2016]). Subsequently, regression models are used to estimate these spatial fields. This type of tasks is challenging because the intensity of a spatial field varies across space, and worker observations are made at different spatial locations. Hence, there is no unique ground truth, and methods in object labelling that rely on consensus [Kamar and Horvitz, 2012], or predefined gold standard questions with known answers [Shah and Zhou, 2015] are not applicable. On the other hand, observations are not independent but spatially correlated, rendering methods that detect colluders via inter-worker similarities [KhudaBukhsh *et al.*, 2014] in rating-based crowdsourcing inapplicable. Trust-based robust estimation methods for estimating continuous-valued quantities such as [Venanzi *et al.*, 2013b] and [Venanzi *et al.*, 2015] are also inapplicable because a spatial field is a continuous function that is different from a finite number of quantities.

Many robust models in regression have been developed in the past. M-estimation [Huber, 2011] was developed to improve the robustness of models to outliers. M-estimators are obtained by minimising general cost functions, and are generalizations to the least-squares estimator (LSE). Other robust models replace the normal distribution of noises in ordinary LSE by heavy-tailed distributions such as Student's t-distribution [Jylänki *et al.*, 2011] and contaminated normal distribution [Huber, 1964]. These methods can be applied to

spatial field regression in participatory sensing to minimise effects of outliers, but they fail to model complex human-like behaviors, such as collusion. In [Venanzi *et al.*, 2013a], a trust-based regression model was proposed for spatial field regression tasks in participatory sensing. It constructs a heteroscedastic Gaussian process (HGP) model, where the accuracies of worker observations are scaled by worker trustworthiness. Attackers are assumed to have low trustworthiness, and report observations with higher variances. A maximum marginal likelihood method is used to jointly estimate model parameters and worker trustworthiness. Due to overly simplified assumptions, this model lacks the ability to mitigate complex real-world malicious attacks, such as injecting biased observations. In addition, the method is unable to incorporate past trustworthiness of workers in future tasks.

Our Contributions. In this paper, we propose a novel trust-based mixture of Gaussian processes (GP) model[1] to yield accurate estimations of spatial fields in the presence of various types of misbehaving workers. The mixture model is the natural choice for this problem. By introducing mixture components that are also spatial fields, the model is made robust against various kinds of complex attacks, including collusion attacks. Our contributions are as follows: (i) We define attacks in spatial regression via a mixture of GP model. (ii) We develop a Bayesian trust framework to maintain and update trustworthiness of participatory sensing workers. Updated trustworthiness improves reliability of future tasks. (iii) We design a novel and efficient Markov chain Monte Carlo (MCMC) sampling-based algorithm for Bayesian inference of the proposed model. (iv) We compare our model with state-of-the-art models using two real-world datasets and demonstrate its robust predictive performance.

2 Problem Definition

We first define the spatial field regression problem in participatory sensing. Suppose there are W potentially dishonest workers in the system. A sequence of tasks are assigned to these workers. We focus on one particular task, where we collected N_i data points $\{(\mathbf{x}_{i,j}, y_{i,j})\}_{j=1}^{N_i}$ from the i-th worker ($N_i \geq 1$). $\mathbf{x}_{i,j} \in \mathcal{X} \subset \mathbb{R}^d$ represents the d-dimensional vector-valued covariates (e.g. geographical coordinates), and $y_{i,j} \in \mathbb{R}$ represents an observation at $\mathbf{x}_{i,j}$. Each observation could either be truthful or untruthful. Truthful observations are made from the target spatial field $f : \mathcal{X} \to \mathbb{R}$, while untruthful observations are unrelated to f (to be defined later). The objective is to reliably estimate $f(\mathbf{x}_*)$ for any $\mathbf{x}_* \in \mathcal{X}$ with its probability distribution.

To motivate the model, we show an example of the stated problem. A participatory sensing task aims to estimate the concentration of a certain air pollutant in a region. For a data point $(\mathbf{x}_{i,j}, y_{i,j})$ from the i-th worker, $\mathbf{x}_{i,j}$ represents the geographical coordinates (longitude and latitude), and $y_{i,j}$ represents the concentration reading at $\mathbf{x}_{i,j}$. The target field f corresponds to the spatial field of the concentration. Suppose that the i-th worker reports observations that are higher than the actual concentrations. These observations would be deemed as untruthful, and the i-th worker would be regarded as dishonest.

[1] A preliminary version of the model has been published in [Xiang *et al.*, 2017]

3 Threat Model

We here discuss various kinds of threats that result in unreliable data. These include sensor faults, inappropriate measurements, and dishonest workers [Mousa *et al.*, 2015].

According to [Sharma *et al.*, 2010], we identify three types of sensor faults. *Noisy*: sensors produce readings with abnormally high variance. *Outlier*: sensors produce abnormally high or low readings. *Uncalibrated*: sensors produce systematically biased readings. Some workers may perform measurements inappropriately. For example, a sensing device may be placed inside a pocket, which would systematically bias the readings, similar to uncalibrated sensors. Dishonest workers deliberately alters measurements at specific locations by injecting data points that changes the estimation of the target field in their favor [Mousa *et al.*, 2015]. They may work in either collusive or non-collusive manner.

Since the task objective is to minimise estimation errors, which could be due to bias and variance, we summarize the effective threats by categorizing them into three types of strategies, denoted by *S1, S2* and *S3*. *S1*: introducing bias into observations. *S2*: increasing variance in observations. *S3*: introducing bias and increasing variance in observations. Besides different strategies, the spatial distribution of untruthful data points may be clustered or dispersed.

A collusion attack involves coalitions in which dishonest workers behave cooperatively to achieve a common objective. For example, a coalition of dishonest workers may adopt a common strategy to alter observations, or focus on a common sub-region. Among all types of threats, the collusion attack that introduces bias is the most damaging type, as it can cause estimations of spatial fields to be highly biased, and is hard to detect using outlier detection-based methods. In existing works related to collusion, the size of coalition is usually upper bounded (e.g. $O(\log n)$ [Celis *et al.*, 2016]), or assumed to be the minority [Wang *et al.*, 2013], which is unrealistic. We do not make this assumption in our model.

4 Trust-based Mixture of GP Model

In this section, we introduce our trust-based mixture of GP model and the Bayesian inference procedures. The model maintains worker trustworthiness by beta prior distribution. It models the probability of a worker reporting truthfully. We differentiate truthful and untruthful observations by different assumptions. It is assumed that there are K coalitions of dishonest workers, each contributing untruthful observations with distinct features. Truthful observations are made from the underlying spatial field f, while untruthful observations from a coalition follow a common strategy function that is independent to f. The value of K is set sufficiently large initially, and the actual value can be determined after the inference.

Bayesian inference of the model is done via an efficient Markov chain Monte Carlo (MCMC) sampling algorithm with two phases. In the first phase, samples of latent variables

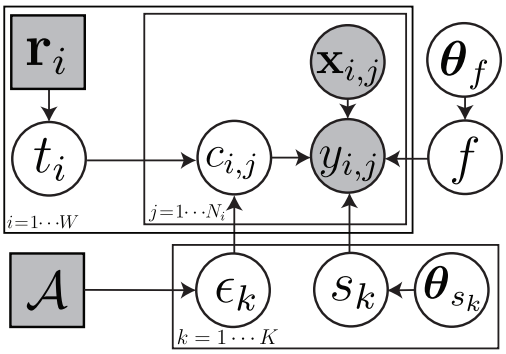

Figure 1: Bayesian model DAG (white circles represent latent variables, shaded circles represent observed data, shaded squares represent known variables).

are simulated from the posterior distribution of the model. In the second phase, predictions of the spatial field f is computed by Monte Carlo integration using samples generated in the first phase. After the inference, trustworthiness of workers is updated. The updated trustworthiness is used in future tasks as prior, resulting in more accurate estimations.

Trustworthiness of Workers. The directed acyclic graph (DAG) of the model is shown in Figure 1. As observations are either truthful or untruthful, we use $c_{i,j} \in \{0, 1, \cdots, K\}$ as the indicator of the truthfulness of an observation $y_{i,j}$, where $c_{i,j} = 0$ indicates $y_{i,j}$ is truthful and $c_{i,j} = k \in \{1, \cdots, K\}$ indicates that $y_{i,j}$ is untruthful and is generated from the k-th coalition. Let $\mathbf{c} := \{c_{i,j} : i = 1, \cdots, W, j = 1, \cdots, N_i\}$. The truthfulness of $y_{i,j}$ depends on the honesty of the i-th worker t_i, defined to be the probability that the i-th worker reports a truthful observation ($t_i \in [0, 1]$). Hence, $t_i := \Pr(c_{i,j} = 0)$. When $c_{i,j} \neq 0$, we define $\epsilon_k := \Pr(c_{i,j} = k | c_{i,j} \neq 0)$, for $k \in \{1, \cdots, K\}$. All indicators in $\mathbf{c}$ are assumed to be pairwise independent condition on $\{t_i\}_{i=1}^W$ and $(\epsilon_1, \cdots, \epsilon_K)$.

We place a beta prior distribution on t_i, with parameters $\mathbf{r}_i := (\alpha_i, \beta_i)$, defined as the trustworthiness of the i-th worker. When a worker first enters the system, he is assigned an initial trustworthiness (e.g. based on his skill level and the type of sensing device). After completion of each task, the trustworthiness of workers will be updated, as elaborated later in Section 4.2, and used in future tasks as priors. Workers who have behaved dishonestly in the past are more likely to be distrusted in future tasks, or are banned from entering the system. A symmetric Dirichlet prior distribution with concentration parameter $\mathcal{A}$ is placed over $(\epsilon_1, \cdots, \epsilon_K)$. We use the beta trust model as it is intuitive for modelling the probability of binary events. It also results in a simple update rule as shown in Section 4.2.

Observation Model. Truthful observations are generated by the target function f. Untruthful observations from the k-th coalition are generated by a strategy function $s_k : \mathcal{X} \to \mathbb{R}$. The strategies can be any of the aforementioned threats. We assume that observations from both f and $\{s_k\}_{k=1}^K$ contain normally distributed noises,

$$(y_{i,j} | c_{i,j} = 0) \sim N(f(\mathbf{x}_{i,j}), \sigma_{n0}^2), \tag{1}$$

$$(y_{i,j} | c_{i,j} = k) \sim N(s_k(\mathbf{x}_{i,j}), \varsigma_k^2),\ k \in \{1, \cdots, K\}. \tag{2}$$

We model f and $\{s_k\}_{k=1}^K$ as realizations from Gaussian processes (GP) [Rasmussen, 2006]. Their priors are specified by the following Gaussian processes,

$$f \sim \mathcal{GP}(\mu_f, \mathcal{C}_f), \tag{3}$$

$$s_k \sim \mathcal{GP}(\mu_{s_k}, \mathcal{C}_{s_k}),\ \text{for } k \in \{1, \cdots, K\}, \tag{4}$$

where $\mu_f, \{\mu_{s_k}\}_{k=1}^K$ are mean functions, and $\mathcal{C}_f, \{\mathcal{C}_{s_k}\}_{k=1}^K$ are covariance functions ($\mathcal{C}_f$ could be stationary or non-stationary depending on assumptions about the f process). Let $\boldsymbol{\theta}_f$ denote hyperparameters of the f process and let $\boldsymbol{\theta}_{s_k}$ denote the hyperparameters of the s_k process. Let π_f and π_s denote the prior distributions of $\boldsymbol{\theta}_f$ and $\{\boldsymbol{\theta}_{s_k}\}_{k=1}^K$ respectively. This creates a mixture of $K + 1$ Gaussian processes.

This versatile model caters to all types of aforementioned threats, and captures untruthful observations using flexible strategy functions $\{s_k\}_{k=1}^K$. In reality, the number of coalitions K is unknown. We can set K sufficiently large initially, and set the prior on $(\epsilon_1, \cdots, \epsilon_K)$ to be weakly informative. This will cause the redundant mixture components to have close to zero posterior probabilities [Rousseau and Mengersen, 2011], which allows us to estimate the actual number of coalitions without performing Bayesian model determination or implementing a trans-dimensional sampling method, since both methods are computationally costly.

4.1 Bayesian Inference of Spatial Field

We can integrate out $\{t_i\}_{i=1}^W$ and $(\epsilon_1, \cdots, \epsilon_K)$ from the posterior distribution of the model. Let $\mathbf{L} := \left(\mathbf{c}, \boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_k}\}_{k=1}^K\right)$ denote remaining latent variables, and $\mathbf{y} := \{y_{i,j} : i = 1, \cdots, W, j = 1, \cdots, N_i\}$ denote the collection of observations. The posterior distribution of $\mathbf{L}$ factorizes as follows,

$$\begin{aligned} p(\mathbf{L}|\mathbf{y}) &= p\left(\mathbf{c}, \boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_k}\}_{k=1}^K | \mathbf{y}\right) \\ &\propto \Pr(\mathbf{c}) p(\mathbf{y}_0 | \boldsymbol{\theta}_f) \pi_f(\boldsymbol{\theta}_f) \textstyle\prod_{k=1}^K p(\mathbf{y}_k | \boldsymbol{\theta}_{s_k}) \pi_s(\boldsymbol{\theta}_{s_k}), \end{aligned} \tag{5}$$

where $\mathbf{y}_0$ represents truthful observations and $\mathbf{y}_k$ represents untruthful observations from the k-th coalition.

To obtain posterior predictive distribution of $f_* := f(\mathbf{x}_*)$, we need to marginalize over latent variables $\mathbf{L}$, i.e.

$$p(f_* | \mathbf{y}) = \int p(f_* | \mathbf{y}, \mathbf{L}) \mathrm{d}p(\mathbf{L} | \mathbf{y}). \tag{6}$$

The first term in the integral is a Gaussian density, as derived by standard GP regression [Rasmussen, 2006], and $p(\mathbf{L}|\mathbf{y})$ is the same as (5). Since it is intractable to perform the integration in (6) analytically, we apply Markov chain Monte Carlo (MCMC) to simulate samples from (5) to approximate this integral by Monte Carlo integration. The Bayesian inference of the spatial field is done in two phases.

In the first phase, samples of latent variables $\mathbf{L}$ are generated from its posterior distribution (5). Each component of $\mathbf{L}$ is sampled from its corresponding conditional posterior distribution via Gibbs sampling. Truthfulness indicators $\mathbf{c}$ is sampled component-wise. Let $\mathbf{c}_- := \mathbf{c} \setminus \{c_{i,j}\}$, $\mathbf{y}_- := \mathbf{y} \setminus \{y_{i,j}\}$. For $l \in \{0, \cdots, K\}$,

$$\begin{aligned} &\Pr(c_{i,j} = l | \mathbf{c}_-, \boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_k}\}_{k=1}^K, \mathbf{y}) \\ &\propto p(y_{i,j} | c_{i,j} = l, \mathbf{c}_-, \boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_k}\}_{k=1}^K, \mathbf{y}_-) \Pr(c_{i,j} = l | \mathbf{c}_-). \end{aligned} \tag{7}$$

In (7), the first term is a Gaussian density derived via standard GP regression and can be computed by removing terms related to $\mathbf{x}_{i,j}$ from the covariance matrix. Note that rank one update can be performed to compute the inverse of the covariance matrix excluding $\mathbf{x}_{i,j}$, which makes sampling of $\mathbf{c}$ computationally efficient. The second term in (7) can be obtained by integrating out t_i and $(\epsilon_1, \cdots, \epsilon_K)$,

$$\Pr(c_{i,j} = 0|\mathbf{c}_-) = \frac{\alpha_i + N_{i,0-}}{\alpha_i + \beta_i + N_i - 1}, \tag{8}$$

$$\Pr(c_{i,j} = k|\mathbf{c}_-) = \frac{n_{k-} + \mathcal{A}}{\sum_k n_{k-} + K\mathcal{A}}[1 - \Pr(c_{i,j} = 0|\mathbf{c}_-)], \quad \text{for } k \in \{1, \cdots, K\}, \tag{9}$$

where $N_{i,0-}$ is the total number of truthful observations from the i-th worker excluding $y_{i,j}$, n_{k-} is the total number of observations from the k-th coalition excluding $y_{i,j}$. Hyperparameters of GP $(\boldsymbol{\theta}_f, \{\boldsymbol{\theta}_{s_k}\}_{k=1}^K)$ can be efficiently sampled via Hybrid Monte Carlo (HMC) [Duane *et al.*, 1987].

The nature of the posterior distribution in (5) implies that it may be multimodal. We introduce parallel tempering [Swendsen and Wang, 1986] to facilitate convergence of the Markov chain by including an auxiliary variable τ as the temperature ($\tau \geq 1$). The joint posterior distribution is,

$$p(\mathbf{L}, \tau|\mathbf{y}) \propto p(\tau)p(\mathbf{L}|\mathbf{y})^{\frac{1}{\tau}}. \tag{10}$$

We set $p(\tau)$ to be uniform. When $\tau = 1$, the marginal posterior is the desired posterior as in (5). Chains with $\tau > 1$ simulate samples from flattened posterior distributions that are easier to sample. We run multiple Markov chains with different τ in parallel, and periodically propose swapping of states $\mathbf{L}$ between chains. A swap of states between two chains $(\mathbf{L}, \tau)$ and $(\mathbf{L}', \tau')$ is accepted with probability as follows,

$$\Pr(accept) = \frac{p(\mathbf{L}', \tau|\mathbf{y})p(\mathbf{L}, \tau'|\mathbf{y})}{p(\mathbf{L}, \tau|\mathbf{y})p(\mathbf{L}', \tau'|\mathbf{y})}, \tag{11}$$

to preserve the detailed balance. When the Markov chain has mixed, we only take samples from the chain with $\tau = 1$ and proceed to the next phase (initial samples are discarded).

In the second phase of the inference algorithm, the posterior predictive distribution of f_* is computed by approximating the integral in (6) via Monte Carlo integration. Suppose we have generated Q samples in the first phase $\left\{\mathbf{L}^{(q)}\right\}_{q=1}^Q$. The integral in (6) can be approximated by

$$p(f_*|\mathbf{y}) \approx \frac{1}{Q}\sum_{q=1}^{Q} p\left(f_*|\mathbf{y}, \mathbf{L}^{(q)}\right). \tag{12}$$

The mean and variance of the distribution in (12) can be computed analytically. Other statistics such as the median and percentiles can be estimated from the histogram.

The time complexities of the two phases of the algorithm are $O(n^3)$ and $O(n^2)$ respectively, which coincide with other GP-based models. Therefore, our model is able to handle datasets with thousands of data points. For larger datasets, reduced-rank approximation [Rasmussen, 2006] may be used.

4.2 Update of Trustworthiness

Throughout different tasks, we maintain the trustworthiness of workers via $\{\mathbf{r}_i\}_{i=1}^W$. For the i-th worker, his trustworthiness before performing a task is $\mathbf{r}_i = (\alpha_i, \beta_i)$. After the task, the updated trustworthiness will become $\mathbf{r}_i' = (\alpha_i', \beta_i')$. We now derive the update rule. Since t_i has been integrated out from the posterior distribution, we need to approximate $p(t_i|\mathbf{y})$ from samples generated through MCMC,

$$p(t_i|\mathbf{y}) = \sum_{\mathbf{c}} p(t_i|\mathbf{c})\Pr(\mathbf{c}|\mathbf{y}). \tag{13}$$

The distribution in (13) is a mixture of beta distributions. We then compute the updated trustworthiness by finding (α_i', β_i') that minimise the Kullback-Leibler (KL) divergence between $\mathrm{Beta}(\alpha_i', \beta_i')$ and (13). Let P denote the distribution in (13) and $\widehat{P}$ denote the approximate distribution $\mathrm{Beta}(\alpha_i', \beta_i')$,

$$(\alpha_i', \beta_i') = \arg\max_{\alpha_i', \beta_i'} \int_0^1 P(t_i) \ln \widehat{P}(t_i) \mathrm{d}t_i. \tag{14}$$

This is a convex optimization problem and can be computed efficiently via Newton's method.

Sometimes it is desirable to control the rate at which trustworthiness is updated. For example, we may associate a weight to each task to indicate its relative importance, and update trustworthiness according to the importance. We may also set the dropping rate of trustworthiness to be faster than the growing rate to increase the penalty for misbehaviors. This control could be achieved by modifying the term $p(t_i|\mathbf{c})$ in (13) to take other factors into account. In this way, temporal strategies (e.g. building up trustworthiness in unimportant tasks and subsequently misbehave in important ones) can be effectively mitigated.

5 Experimentation

In this section, we demonstrate the robustness of our model by applying it to two real-world datasets and comparing its predictive performance with several state-of-the-art models.

Datasets and Experimental Settings. The first dataset is the air quality index (AQI) data retrieved from the World Air Quality Index project (http://aqicn.org). It consists of 435 AQI readings from East Asia collected by Environmental Protection Agencies on January 10, 2017. In this dataset, covariates are longitudes and latitudes where data were observed, and observations are measured PM2.5 AQIs. The second dataset is the US monthly precipitation dataset collected by the Institute for Mathematics Applied to Geosciences, US National Center for Atmospheric Research during 1895-1997 (http://www2.image.ucar.edu). In this dataset, covariates are longitudes and latitudes of weather stations and observations are the total monthly precipitation in millimeters. We use a subset of data from October of 1997 in a region, consisting of 1618 data points. Since these datasets are official, we assume all readings are truthful and refer to these uncontaminated datasets as original datasets. From each dataset, we randomly select a subset of data points (100 from the AQI dataset and 200 from the US precipitation dataset) and use them as predictive points for evaluating different models.

We compare our regression model with three baseline models, including standard GP regression, robust GP regression with Student's t-distributed noises [Jylänki *et al.*, 2011], and TrustHGP [Venanzi *et al.*, 2013a]. We adapted all three models into fully-Bayesian models and applied MCMC to sample latent variables and approximate the posterior predictive distribution. The same weakly informative priors on GP hyperparameters are used for all models. By using MCMC as the sole inference method, we focus only on the differences in performance due to different model specifications, and exclude the effects of inference techniques (e.g. Laplace approximate, expectation propagation, etc.)

In the experiments, we use anisotropic Matérn covariance functions with $\nu = \frac{3}{2}$ for all stationary GP priors, same as in [Zenonos *et al.*, 2015]. Additionally, non-stationary Matérn covariance function introduced in [Paciorek and Schervish, 2004] is used in conjunction with the proposed model in the first part of the experiment to demonstrate its applicability in non-stationary spatial fields. Although most physical phenomena are non-stationary due to their spatially varying smoothness properties, we use stationary covariance functions for comparing models due to the much higher computational cost of non-stationary regression. In practice, non-stationary covariance functions work better with our model as discussed later in this section, and is preferred when the computational efficiency is not a major concern and the underlying phenomenon is likely to be non-stationary.

To compare different models, we contaminate a subset of observations in original datasets with three types of strategies introduced in Section 3, changing them into untruthful observations. Bias in observations is introduced by a linear transformation of actual observations, and extra variance is introduced by adding a zero-mean Gaussian noise term to the reading. Performance is evaluated by both root-mean-square error (RMSE) and mean of log predictive density (MLPD). RMSE evaluates the point prediction (posterior mean) accuracies, while MLPD provides a more comprehensive measurement of predictive performance by taking into account posterior uncertainties. A higher MLPD indicates better performance.

Experiments on AQI Dataset. We use the AQI dataset to evaluate model performance when a single strategy is present. We assign data points in the dataset to 20 imaginary workers and then alter readings from 5 of the workers by one of three strategies (*S1, S2* or *S3*). For each strategy, we build one dataset with clustered untruthful data points and another with dispersed untruthful data points.

We first demonstrate the Bayesian inference results and the updated trustworthiness of our model by an experiment on a dataset with dispersed untruthful data points and strategy *S1*. In Figure 2, the top-right, bottom-left and bottom-right subfigures show the posterior mean resulted from standard GP and our model with non-stationary and stationary covariance functions. The top-left subfigure shows the posterior mean of standard GP on the original dataset as a reference. Our model has reconstructed a smooth spatial field that explains the spatial phenomenon well, while the standard GP produced a rapidly varying spatial field due to the presence of untruthful data points, resulting in poor predictive

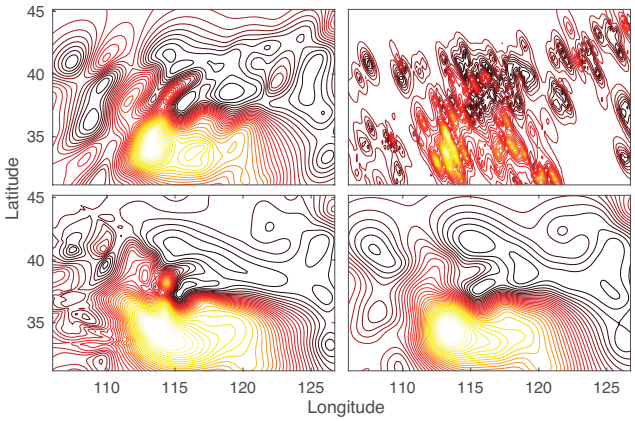

Figure 2: Contour plots of posterior mean from standard GP in original dataset (top-left), and contaminated dataset (top-right), our model with non-stationary (bottom-left) and stationary (bottom-right) covariance in contaminated dataset. [**Best Viewed in Color**]

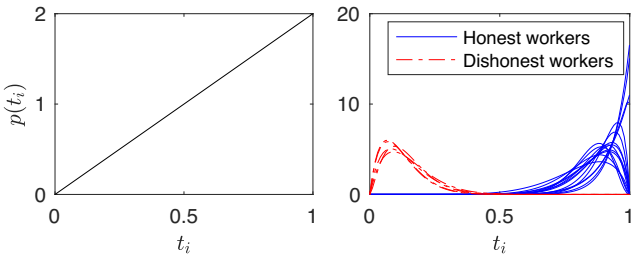

Figure 3: Probability density functions of initial trustworthiness of workers (left) and updated trustworthiness of honest and dishonest workers (right). [**Best Viewed in Color**]

performance. In this experiment, the initial trustworthiness of all workers was set as $\mathbf{r}_{init} = (2, 1)$. After performing Bayesian inference, our model updated the trustworthiness of each worker. Figure 3 shows the probability density function $p(t_i)$ of the initial trustworthiness and updated trustworthiness of all 20 imaginary workers. Honest workers were more trusted, while dishonest workers became distrusted after the inference. The updated trustworthiness could be used as prior in future tasks, which would further reduce the impact of untruthful data points provided by dishonest workers. The number of coalitions K was set as 8. The mean posterior probability of $(\epsilon_1, \cdots, \epsilon_K)$ was (0.78, 0.09, 0.05, 0.03, 0.02, 0.01, 0.01, 0.01). Therefore, one major coalition could be identified from this dataset.

Figure 4 shows the predictive performance (RMSE and MLPD) of five models on different strategies and spatial distribution of untruthful data points in a bar chart, with the performance on the original dataset on the leftmost column for reference, and error bars indicating the standard deviation of RMSEs and MLPDs. TrustMix(NS) and TrustMix(S) refer to our trust-based model with non-stationary and stationary covariance functions, respectively. Both TrustMix(NS) and TrustMix(S) consistently achieved good predictive performances with low RMSEs and high MLPDs. Three baseline models were shown to be susceptible to strategies that involve biases, especially when untruthful data points are clustered. Extra variances had insignificant effects on pre-

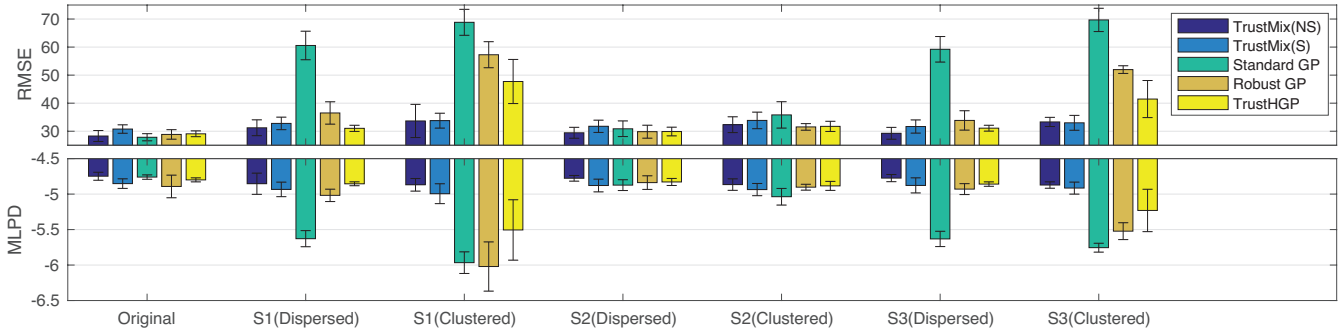

Figure 4: RMSE and MLPD of predictions resulted from five models in contaminated AQI datasets. [**Best Viewed in Color**]

Table 1: MLPD of predictions on US precipitation datasets.

Data	*S1*	*S2*	*S3*	TrustMix	Std. GP	Robust GP	TrustHGP
A	0	0	0	-2.0439	-2.0588	-2.0217	-2.0821
B1	10	10	10	-2.1110	-2.9723	-4.3511	-2.3060
B2	18	6	6	-2.1259	-2.7952	-3.4426	-5.4699
B3	6	18	6	-2.1646	-2.4841	-3.8650	-2.1348
B4	6	6	18	-2.1310	-2.8929	-3.5105	-2.2868
B5	24	3	3	-2.1550	-2.9895	-3.1315	-7.3516
B6	3	24	3	-2.1593	-2.3940	-2.2644	-2.0986
B7	3	3	24	-2.1487	-7.8555	-3.5302	-2.3003
C	20	20	20	-2.1153	-	-	-

dictive performances. One notable aspect of the results is that our model with non-stationary covariance function outperformed the stationary version. We speculate that this is due to model mis-specification, as the AQI dataset is likely to be non-stationary. Since a mixture of stationary GP may fit a non-stationary function well, mixture components in our model may be used to fit the non-stationary function instead of real strategies. Thus, in practice, non-stationary covariance functions are preferred when using this model.

Experiments on US Precipitation Dataset. We use this dataset to demonstrate the capability of our model to handle multiple strategies. We create nine datasets for the experiment, A, B1, B2, B3, B4, B5, B6, B7 and C in a similar way as in the previous experiments. Dataset A is the original uncontaminated dataset. Dataset B1-B7 are datasets contaminated by 30 dishonest workers forming three coalitions with varied proportions of dishonest workers and strategies *S1*, *S2* and *S3*. Dataset C demonstrates a special case where the majority of workers are dishonest, but honest workers have higher trustworthiness resulted from past behaviors. These datasets are used to evaluate the predictive performance of our model and three baseline models (all four models use stationary covariance functions).

Table 1 shows the performance in MLPD of four models on nine datasets (Dataset C is inapplicable to baseline models). RMSE results give the same conclusions, and thus are omitted for saving space. Columns 2 to 4 show the number of dishonest workers in the coalition following strategies *S1*, *S2* and *S3*. Our model has achieved good predictive performance in all datasets. While TrustHGP achieved slightly better performance when the majority of dishonest workers follow *S2*, its performance was significantly affected when the majority of dishonest workers follow *S1*, which could be the case when there is collusion attack. Apart from being more robust than baseline models when dishonest workers are present, the performance of our model on dataset C demonstrates its ability to incorporate past trustworthiness maintained in the model to improve its identification of misbehaviors, even in the extreme case where dishonest workers become the majority. This could not be achieved in baseline models. Same as the experiments on the AQI dataset, in the experiment with dataset C, the number of coalitions K was initially set as 8, and the mean posterior probability of $(\epsilon_1, \cdots, \epsilon_K)$ was (0.35, 0.29, 0.23, 0.07, 0.03, 0.02, 0.01, 0.00), and roughly three coalitions could be identified from this dataset.

Although the spatial field of precipitation is likely to be non-stationary considering the fact that the data comes from a large geographical area with a variety of terrain features, we used only stationary models for efficiency considerations. Nevertheless, we could expect a better predictive performance if non-stationary models were used, as we have demonstrated in the experiments on the AQI dataset.

6 Conclusion and Future Work

This paper introduced a novel trust-based mixture of Gaussian processes model to resolve the problem of spatial field regression in the presence of untruthful data in participatory sensing. Bayesian inference of the model is done via an efficient MCMC-based algorithm. We demonstrated the predictive performance of the model using two real-world datasets, comparing with three baseline models. Our model was shown to be significantly more effective than baseline models against various attacks, which can cause detrimental effects if not mitigated. The trustworthiness of workers is efficiently updated after each task, which improves the accuracy of future tasks. For future work, we will investigate the possibility of using reduced-rank approximation to improve the scalability of the model, especially in the non-stationary case.

Acknowledgements

We wish to acknowledge the funding for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme.

References

[Celis *et al.*, 2016] L Elisa Celis, Sai Praneeth Reddy, Ishaan Preet Singh, and Shailesh Vaya. Assignment techniques for crowdsourcing sensitive tasks. In *Proceedings of ACM CSCW*, 2016.

[Duane *et al.*, 1987] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[Huber, 1964] Peter J Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[Huber, 2011] Peter J Huber. *Robust statistics*. Springer, 2011.

[Jylänki *et al.*, 2011] Pasi Jylänki, Jarno Vanhatalo, and Aki Vehtari. Robust gaussian process regression with a student-t likelihood. *Journal of Machine Learning Research*, 12:3227–3257, 2011.

[Kamar and Horvitz, 2012] Ece Kamar and Eric Horvitz. Incentives for truthful reporting in crowdsourcing. In *Proceedings of AAMAS*, 2012.

[Kamar *et al.*, 2012] Ece Kamar, Severin Hacker, and Eric Horvitz. Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of AAMAS*, 2012.

[KhudaBukhsh *et al.*, 2014] Ashiqur R KhudaBukhsh, Jaime G Carbonell, and Peter J Jansen. Detecting non-adversarial collusion in crowdsourcing. In *Proceedings of AAAI HCOMP*, 2014.

[Mousa *et al.*, 2015] Hayam Mousa, Sonia Ben Mokhtar, Omar Hasan, Osama Younes, Mohiy Hadhoud, and Lionel Brunie. Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, 90:49–73, 2015.

[Paciorek and Schervish, 2004] Christopher J. Paciorek and Mark J. Schervish. Nonstationary covariance functions for gaussian process regression. In *Proceedings of NIPS*, 2004.

[Rasmussen, 2006] Carl Edward Rasmussen. *Gaussian processes for machine learning*. MIT Press, 2006.

[Rousseau and Mengersen, 2011] Judith Rousseau and Kerrie Mengersen. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710, 2011.

[Salek *et al.*, 2013] Mahyar Salek, Yoram Bachrach, and Peter Key. Hotspotting-a probabilistic graphical model for image object localization through crowdsourcing. In *Proceedings of AAAI*, 2013.

[Shah and Zhou, 2015] Nihar Bhadresh Shah and Denny Zhou. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. In *Proceedings of NIPS*, 2015.

[Sharma *et al.*, 2010] Abhishek B Sharma, Leana Golubchik, and Ramesh Govindan. Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Transactions on Sensor Networks (TOSN)*, 6(3):23, 2010.

[Swendsen and Wang, 1986] Robert H Swendsen and Jian-Sheng Wang. Replica monte carlo simulation of spin-glasses. *Physical Review Letters*, 57(21):2607, 1986.

[Tarasov *et al.*, 2014] Alexey Tarasov, Sarah Jane Delany, and Brian Mac Namee. Dynamic estimation of worker reliability in crowdsourcing for regression tasks: Making it work. *Expert Systems with Applications*, 41(14):6190–6210, 2014.

[Tran-Thanh *et al.*, 2015] Long Tran-Thanh, Trung Dong Huynh, Avi Rosenfeld, Sarvapali D Ramchurn, and Nicholas R Jennings. Crowdsourcing complex workflows under budget constraints. In *Proceedings of AAAI*, 2015.

[Venanzi *et al.*, 2013a] Matteo Venanzi, Alex Rogers, and Nicholas R Jennings. Crowdsourcing spatial phenomena using trust-based heteroskedastic gaussian processes. In *Proceedings of AAAI HCOMP*, 2013.

[Venanzi *et al.*, 2013b] Matteo Venanzi, Alex Rogers, and Nicholas R Jennings. Trust-based fusion of untrustworthy information in crowdsourcing applications. In *Proceedings of AAMAS*, 2013.

[Venanzi *et al.*, 2015] Matteo Venanzi, W. T. Luke Teacy, Alex Rogers, and Nicholas R. Jennings. Bayesian modelling of community-based multidimensional trust in participatory sensing under data sparsity. In *Proceedings of IJCAI*, 2015.

[Wang *et al.*, 2013] Xinlei Oscar Wang, Wei Cheng, Prasant Mohapatra, and Tarek Abdelzaher. Artsense: Anonymous reputation and trust in participatory sensing. In *Proceedings of INFOCOM*, 2013.

[Xiang *et al.*, 2017] Qikun Xiang, Jie Zhang, Ido Nevat, and Pengfei Zhang. A trust-based mixture of gaussian processes model for robust participatory sensing. In *Proceedings of AAMAS*, 2017.

[Zenonos *et al.*, 2015] Alexandros Zenonos, Sebastian Stein, and Nicholas R Jennings. Coordinating measurements for air pollution monitoring in participatory sensing settings. In *Proceedings of AAMAS*, 2015.

[Zenonos *et al.*, 2016] Alexandros Zenonos, Sebastian Stein, and Nicholas R Jennings. An algorithm to coordinate measurements using stochastic human mobility patterns in large-scale participatory sensing settings. In *Proceedings of AAAI*, 2016.