Spatio-Temporal Bayesian Modeling for Mobile Edge Computing

Laha Ale
Department of Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, USA
Email: lale@islander.tamucc.edu

Instructor:Dr. Jose Guardiola
Department of Mathematics & Statistics
Texas A&M University-Corpus Christi
Corpus Christi, USA
Email: jose.guardiola@tamucc.edu

***Abstract*—With the emergence of the Internet of Things(IoT), the number of mobile devices connects to wireless are growing exponentially, posing a severe strain on the core network and backhaul bandwidth. Also, the mobile network has its unique properties and challenges. First, changing over space-time in one of the main characteristics of the mobile network. Second, mobile devices such as mobile phone attached more personal information and preferences. Third, devices are connected to the mobile network have less computational resources and limited power supply. Mobile Edging Computing (MEC) is proposed to tackle abovementioned challenges. In this work, we try to model the distribution connection over space-time, which is a fundamental and considerably difficult problem in wireless network communication. Concretely, we adopt the Spatio-Temporal Bayesian method to model and predict the distribution of the connection over space and time. Furthermore, we train and test our model on real-world data, and the results show that our method can achieve considerably high accuracy. The code of this work has been posted on GitHub[1] and free available to the public.**

***Index Terms*—Mobile Edging Computing, Bayesian Modeling, Spatio-Temporal**

I. Introduction

With the wave of Internet of Things (IoTs) [1], [2], we have witnessed a rapid proliferation of wireless devices, including mobile phones, wearable devices, connected cars, and sensors. As a result, wireless traffic has skyrocketed dramatically [3], [4], causing potential congestion in core networks and posing severe strain on backhaul bandwidth. Moreover, the emergence of various services and applications, such as augmented reality, virtual reality, and intelligent transportation, require efficient and real-time data collection and data processing. To address these challenges, storage and computation resources expects to be provisioned to end devices in proximity, giving birth to mobile edge computing (MEC) [5], [6]. Consequently, with the aid of MEC, the demands from end devices can be served locally and quickly, rather than being served by the remote cloud through backhaul. In other words, through decentralizing cloud service and spreading the burden of cloud servers in core networks to base stations (BS), the requests for both contents and computations can be satisfied locally, which significantly mitigates the pressure on core network and reduces the service latency for end devices.

[1]https://github.com/ainilaha/Spatio-Temporal-Bayesian-MEC

MEC poses many challenges in terms of deployment and resource management [7], [8]. By incorporation of storage and computation resources in networks, it is of importance yet very challenging to utilize various forms of resources efficiently to meet different users' needs. For the different space-time requests from end devices, it is found that the requests may exhibit certain similarity among different users over time. The same popular contents may be requested at similar locations at different time instants. Therefore, as presented in [9], and [10], proactively caching reusable contents such as video streams and high-resolution images during off-peak periods and reusing cached content in peak hours, can reduce traffic loads in backhaul and service latency of end users so that the QoE [11] can be improved [9], [12]. For computation requests, some computation applications and data can also be stored such that the response can be made promptly for tasks from users [13].

The performance of proactive caching heavily depends on the accuracy of the prediction of content popularity, which is typically unknown and change over space-time. Content popularity prediction is very challenging due to the high dynamics of users' requests and mobility. By harnessing the storage and computing resources brought from MEC, machine learning has great potential to learn from the requested data for caching management. Specifically, caching control units in local BS can harness machine learning techniques [14] to predict content popularity, adapt to the time-varying popularity of requests from users, and update the cache intelligently. In addition, there are some research efforts along this direction, including cache prediction machine learning models [15][16], transfer learning based caching management methods [17] [18] [19], and reinforcement learning based models [20][21]. However, the prediction models in [15][16] assume the popularity of files is time-invariant, which might not be true in the real world. Though transfer learning models can reduce the training effort, they can cause high complexity and require high computational cost due to additional features in the pre-trained models. Moreover, methods, as mentioned earlier, assume that the locations are static or not take the spatial dimension as consideration. Resources are hardly exploited or offloading contents without knowing the space-time distribution.

Additionally, space-time based models can offer personal-

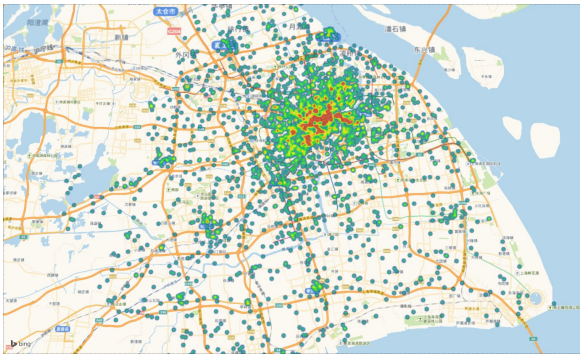

Fig. 1. data

ized services such as contents recommendation [22], space-time based reminder, and navigation. However, both mobile devices and edge devices such as location base station are quipped limited resource that cannot run highly computational cost models. In MEC application, it is an online learning process that the models expected to keep learning over time, which indicate the model has frequently been training, rising a significant amount of computational resources if we adopt classical statistics models our or deep learning methods as mentioned earlier.

In this work, our goal is to try to represent and predict mobile users connection over space-time and providing the foundation of many other applications such as contents offloading, contents recommendation, and computing recommendation. First, we performed spatial analysis over the data to find the spatial relation of the datasets. Second, we formulate the problem as spatiotemporal Bayesian models [23] and build an online learning Bayesian model. Third, we evaluate on real-world dataset [24] and visualize the results.

The remainder of the paper is organized as follow. Section II Methodology; III Experiment setting and Results Analysis Section; IV discuss the future work; Section V concludes this work.

II. Methodology

A. Classical Geospatial Statistics Methods

In this work, we adopted a published data provided by [24] to set up our experiment. Before we build our Bayesian model, we will explore and analyze the data and explore the data is correlated spatially. First, we use Variogram (Eq.1), where $N(d)$ is the set of points such that $||s_i - s_j|| = d$ and $|N(d)|$ is the number of pairs. And the covariance (Eq.3) and Variogram (Eq.4) functions, we adopted the *exponential* functions in this work; where τ is nugget indicated non-spatial term.

$$\widehat{\gamma}(d) = \frac{1}{2N(d)} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(d)} \left[Y(\mathbf{s}_i) - Y(\mathbf{s}_j) \right]^2 \tag{1}$$

$$N(d_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}, k = 1, \ldots, K \tag{2}$$

$$C(d) = \begin{cases} \sigma^2 \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases} \tag{3}$$

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi d)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases} \tag{4}$$

A classical spatial prediction method is $Kriging$, specially spatial prediction in the point-referenced data setting. The $Kriging$ method can be shown in Eq.5, where $\widehat{\gamma} = \left(\hat{\sigma}^2 \rho\left(\hat{\phi}; d_{01}\right), \ldots, \hat{\sigma}^2 \rho\left(\hat{\phi}; d_{0n}\right)\right)^T$ and $\widehat{\boldsymbol{\beta}} = \left(X^T \widehat{\Sigma}^{-1} X\right)^{-1} X^T \widehat{\Sigma}^{-1} \mathbf{y}$. Besides, since our data observations are well distributed over the domain, the spatial pattern can be detected by estimating continuous surfaces using interpolation functions.

$$\widehat{h(\mathbf{y})} = \mathbf{x}_0^T \widehat{\boldsymbol{\beta}} + \widehat{\gamma}^{T\widehat{\Sigma}^{-1}} (\mathbf{y} - \overline{X}\widehat{\boldsymbol{\beta}}) \tag{5}$$

The observed samples of our data can be considered as an online generation; the connections from mobile devices would move around and disconnect one base station and connect to another. Therefore, we can exploit Bayes modeling that can continuously update the model by newly generated data. A basic model has shown Eq.6, where the mean structure $\mu(s) = x^T(s)\theta$. the residual is partitioned into two pieces, one is spatial $w(s)$ and non-spatial $\epsilon(s)$.

$$Y(s) = \mu(s) + w(s) + \epsilon(s) \tag{6}$$

Typically, independent priors are chosen for the different parameters shown Eq.7, and natural candidates are the multivariate for β and inverse gamma for σ^2 and τ. Specification for ϕ depends upon the choice of ρ function; in the simple exponential case $\rho(\mathbf{s}_i - \mathbf{s}_j; \phi) = \exp(-\phi \|\mathbf{s}_i - \mathbf{s}_j\|)$, a gamma might seem sensible.

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}) p\left(\sigma^2\right) p\left(\tau^2\right) p(\phi) \tag{7}$$

Since we will often want to make inferential statements about the parameters separately, we will need to obtain marginal posterior distributions. For instance, a point estimate or credible interval for β raise from Eq.8.We will resort Markov chain Monte Carlo (MCMC) to perform Bayesian computations. There are three most popular MCMC algorithms, the Gibbs sampler (Alg.1), the Metropolis-Hasting (Alg.2), and the slice sampler, and we applied the first two in this work.

Algorithm 1: Gibbs sampler

```
Result: Updated parameters
for t ∈ 1 : T do
    for k ∈ Steps do
        Step k: Draw θ_k^(t) from
          p(θ_k | θ_1^(t), θ_2^(t), ..., θ_{k-1}^(t), y)
    end
end
```

Algorithm 2: Metropolis-Hasting

Result: Updated parameters

for $t \in 1:T$ **do**

1. Draw θ^* from $q\left(\cdot|\boldsymbol{\theta}^{(t-1)}\right)$

2. compute the ratio:

$r = h\left(\boldsymbol{\theta}^*\right)/h\left(\boldsymbol{\theta}^{(t-1)}\right)$

=exp $\left[\log h\left(\boldsymbol{\theta}^*\right) - \log h\left(\boldsymbol{\theta}^{(t-1)}\right)\right]$;

3.If $r \geq 1$, set $\theta^{(t)} = \theta^*$;

If $r < 1$, set $\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^* \text{ with probability } r \\ \boldsymbol{\theta}^{(t-1)} \text{ with probability } 1-r \end{cases}$

end

$$
\begin{aligned}
p(\boldsymbol{\beta}|\mathbf{y}) &= \iiint p\left(\boldsymbol{\beta}, \sigma^2, \tau^2, \phi|\boldsymbol{y}\right) d\sigma^2 d\tau^2 d\phi \\
&\propto p(\boldsymbol{\beta}) \iiint f(\mathbf{y}|\boldsymbol{\theta}) p\left(\sigma^2\right) p\left(\tau^2\right) p(\phi) d\sigma^2 d\tau^2 d\phi
\end{aligned} \tag{8}
$$

B. Spatio-Temporal Bayesian Modeling

The above methods do not consider time dimension whereas the mobile connections space-time dynamic problem; therefore, we adopted spatio-temporal Bayesian models to learn historical data and predict future connections. Let l and t denote the two units of time where l denotes the more extended unit, e.g., year,$l = 1, ..., r$, and t denotes the shorter unit, e.g., minutes, $t = 1, ..., T_l$ where r and T_l denote the total number of two time units, respectively. Let $Z_l(\mathbf{s}_i, t)$ denote the observed point-referenced data and $O_l(\mathbf{s}_i, t)$ be the true value corresponding to $Z_l(\mathbf{s}_i, t)$ at site $\mathbf{s}_i, i = 1, \ldots, n$ at time denoted by two indices l and t. Let $\mathbf{Z}_{lt} = \left(Z_l(\mathbf{s}_1, t), \ldots, Z_l(\mathbf{s}_n, t)\right)^\top$ and $\mathbf{O}_{lt} = \left(O_l(\mathbf{s}_1, t), \ldots, O_l(\mathbf{s}_n, t)\right)^\top$. We shall denote all the observed data by z and z^* Let $N = n\sum_{l=1}^{r} T_l$ will denote all the missing data. be the total number of observations to be modeled.

Also, $\epsilon_{lt} = \left(\epsilon_l(\mathbf{s}_1, t), \ldots, \epsilon_l(\mathbf{s}_n, t)\right)^\top$ will be used to denote the nugget eeffect or the pure error term assumed to be independently normally distributed $N\left(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n\right)$. The spatio-temporal random effects will be denoted by $\boldsymbol{\eta}_{lt} = \left(\eta_l(\mathbf{s}_1, t), \ldots, \eta_l(\mathbf{s}_n, t)\right)$ and these will be assumed to follow $N(\mathbf{0}, \Sigma_\eta)$ independently in time, where $\Sigma_\eta = \sigma_\eta^2 S_\eta, \sigma_\eta^2$ is the site invariant spatial variance and S_η is the spatial correlation matrix obtained from the, often used, general Matern correlation function (Eq.9).

$$
\begin{aligned}
\kappa(\mathbf{s}_i, \mathbf{s}_j; \phi, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu}\left\|\mathbf{s}_i - \mathbf{s}_j\right\|\phi\right)^\nu \\
K_\nu\left(2\sqrt{\nu}\left\|\mathbf{s}_i - \mathbf{s}_j\right\|\phi\right), \quad \phi > 0, \nu > 0
\end{aligned} \tag{9}
$$

The independent Gaussian process (GP) model is specified hierarchically by Eq.10 and Eq.11.

$$
\mathbf{Z}_{lt} = \mathbf{O}_{lt} + \boldsymbol{\epsilon}_{lt} \tag{10}
$$

$$
\mathbf{O}_{lt} = \mathbf{X}_{lt}\boldsymbol{\beta} + \boldsymbol{\eta}_{lt} \tag{11}
$$

for each $l = 1, \ldots, r$ and $t = 1, \ldots, T_l$ where we assume that ϵ_{lt} and $\boldsymbol{\eta}_{lt}$ are independent and each are normally distributed with their respective parameters. Let $\mathbf{O}$ denote all the random effects, $\mathbf{O}_{lt}, l = 1, \ldots, r$ and $t = 1, \ldots, T_l$. Let $\boldsymbol{\theta} = \left(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\eta^2, \phi, \nu\right)$ denote all the parameters of this model and let $\pi(\boldsymbol{\theta})$ denote the prior distribution that we shall specify later. The logarithm of the joint posterior distribution of the parameters and the missing data for this GP model is given by Eq.12.

$$
\begin{aligned}
\log \pi\left(\boldsymbol{\theta}, \mathbf{O}, \mathbf{z}^*|\mathbf{z}\right) \propto -\frac{N}{2}\log\sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2}\sum_{l=1}^{r}\sum_{t=1}^{T_l}\left(\mathbf{Z}_{lt} - \mathbf{O}_{lt}\right)^\top \\
\left(\mathbf{Z}_{lt} - \mathbf{O}_{lt}\right) - \frac{\sum_{l=1}^{r} T_l}{2}\log\left|\sigma_\eta^2 S_\eta\right| \\
-\frac{1}{2\sigma_\eta^2}\sum_{l=1}^{r}\sum_{t=1}^{T_l}\left(\mathbf{O}_{lt} - \rho\mathbf{O}_{lt-1} - \mathbf{X}_{lt}\boldsymbol{\beta}\right)^\top S_\eta^{-1} \\
\left(\mathbf{O}_{lt} - \rho\mathbf{O}_{lt-1} - \mathbf{X}_{lt}\boldsymbol{\beta}\right) - \frac{1}{2}\sum_{l=1}^{r}\log\left|\sigma_l^2 S_0\right| \\
-\frac{1}{2}\sum_{l=1}^{r}\frac{1}{\sigma_l^2}\left(\mathbf{O}_{l0} - \boldsymbol{\mu}_l\right)^\top S_0^{-1} \\
\left(\mathbf{O}_{l0} - \boldsymbol{\mu}_l\right) + \log\pi(\boldsymbol{\theta})
\end{aligned} \tag{12}
$$

III. Experiments and Results Analysis

We will demonstrate the experiments and interpreted the results.

A. Data Clean

Most of the data from the real-world need preprocessing and clean in order to further analysis. We have tidy and assigned the columns name properly we can load the data (Fig.2) into our code and visualize on the map as shown in Fig.3; however, we find out there are few data far away from the most of data in spatially distance. We are more interested in the major of data as shown in Fig.4; therefore, we need to find extra data to help us to filter out the outlier so that we can focus on the majority of the data. We adopted the polygon map to filter out the the outlier and derived the clean data as shown in Fig.5. In addition, we can plot the connect sum of time spatially distribution and colored with connected time as Fig.6.

B. Interpolation

We can plot the data distribution with spatially with interpolation and produces. As we can see the sum of connection by location in Fig.7 and Fig.8, the values are variant from a location from another, and its spatial distribution. Mobile users stay connection longer in some of the areas other some areas because of some of the areas like work zone, they stay on place longer whereas places like shopping mall would be different. Another more straightforward measurement is the connection number of users grouped by location which is shown in Fig.9 and Fig.10; there is a place close to the right corner has a larger number connection than other places.

```
> head(data,n=15)
    X  month date          start_time            end_time                          user_id      lat     long
1   1 201406   16 2014-06-16 09:52:25 2014-06-16 10:04:24 88ac431480911d3c7a41a2bde5c74b85 31.23787 121.4703
2   2 201406   16 2014-06-16 01:51:22 2014-06-16 02:24:55 56987a61b95afb9d7cb04e26e19ad769 31.23787 121.4703
3   3 201406   16 2014-06-16 22:55:40 2014-06-16 22:58:43 f8206ab58b9bdb070673f7050242e9ee 31.23787 121.4703
4   4 201406   16 2014-06-16 23:00:32 2014-06-16 23:05:24 f8206ab58b9bdb070673f7050242e9ee 31.23787 121.4703
5   5 201406   16 2014-06-16 22:40:32 2014-06-16 22:40:46 f8206ab58b9bdb070673f7050242e9ee 31.23787 121.4703
6   6 201406   16 2014-06-16 08:32:58 2014-06-16 08:42:25 d5d113ac402205dee8ff6866a5c01b78 31.23787 121.4703
7   7 201406   16 2014-06-16 08:31:54 2014-06-16 08:45:20 d5d113ac402205dee8ff6866a5c01b78 31.23787 121.4703
8   8 201406   16 2014-06-16 16:38:39 2014-06-16 16:39:03 1709f1d6ad0d111ecedc9e64ab11437d 31.23787 121.4703
9   9 201406   16 2014-06-16 15:46:36 2014-06-16 16:28:04 1709f1d6ad0d111ecedc9e64ab11437d 31.23787 121.4703
10 10 201406   16 2014-06-16 17:19:32 2014-06-16 17:25:28 1709f1d6ad0d111ecedc9e64ab11437d 31.23787 121.4703
11 11 201406   16 2014-06-16 17:51:30 2014-06-16 18:19:37 1709f1d6ad0d111ecedc9e64ab11437d 31.23787 121.4703
12 12 201406   16 2014-06-16 16:54:25 2014-06-16 17:04:20 1709f1d6ad0d111ecedc9e64ab11437d 31.23787 121.4703
13 13 201406   16 2014-06-16 10:42:08 2014-06-16 10:45:21 22e385838db90a9ad0cc862e022605f5 31.23787 121.4703
14 14 201406   16 2014-06-16 12:50:40 2014-06-16 13:27:50 f3f7108fee650c001c53478da1e13fb2 31.24695 121.5139
15 15 201406   16 2014-06-16 11:07:40 2014-06-16 11:14:34 b6f11a26170aabb49062f79b03818bf2 31.24695 121.5139
```

Fig. 2. Partial Data

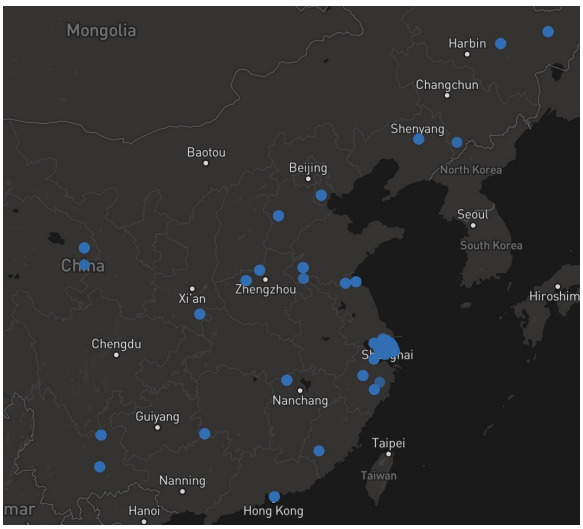

Fig. 3. Sparse Data

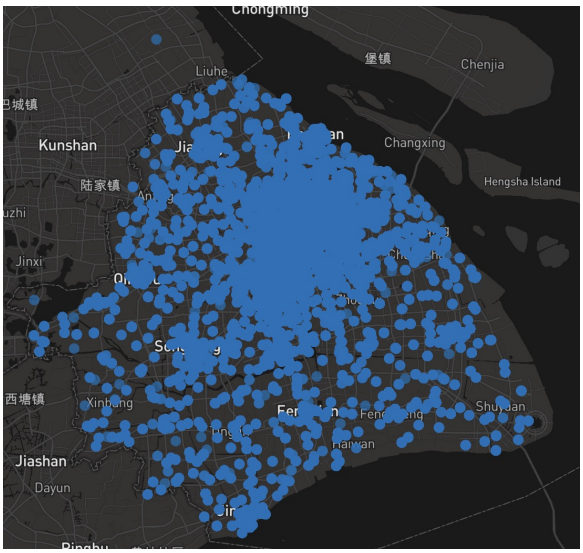

Fig. 4. Main data set

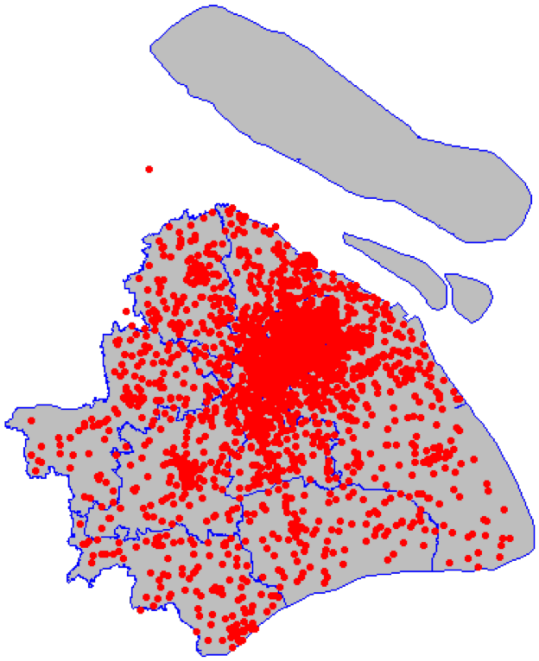

Fig. 5. Clean Dataset

C. Virogram

As we can see from the Variogram Fig.11, it connection number is spatially correlated. In this experiment, we set sill $\sigma^2 = 8205$, the range $\phi = 0.02$, and $\tau = 2495$. The non-spatial range is considerably large given the dataset locations.Besides, the Variogram is fit the dataset well with $exponential$ term.

D. Kriging

We have also run the Kriging prediction model for our point-data. The parameter setting for Kriging model is produced by Variogram fitting, and $\sigma^2 = 8205$,$\tau = 2495$, and $\phi = 0.0811$. The 95% Prediction Interval (PI) is between -176.20 and

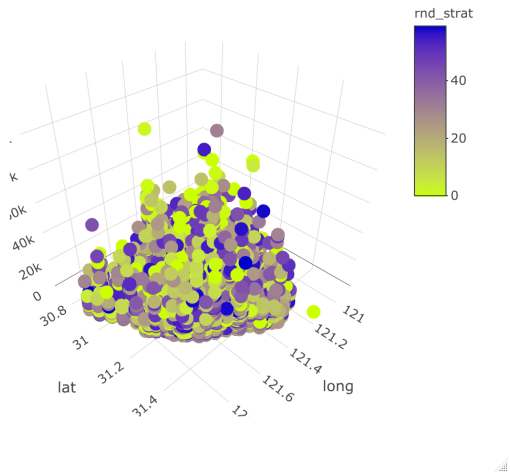

Fig. 6. connection sum time and start to connection time

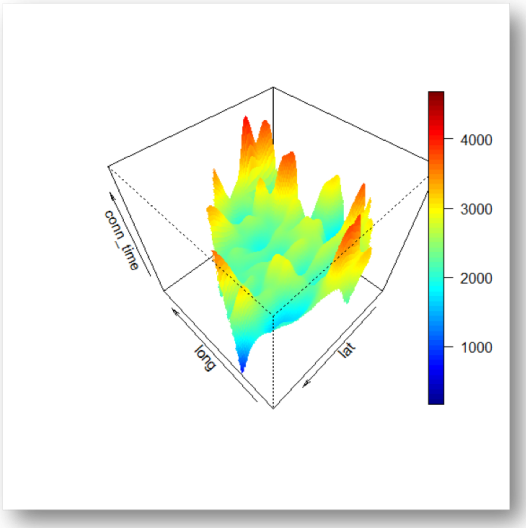

Fig. 8. Sum connection time 3d plot

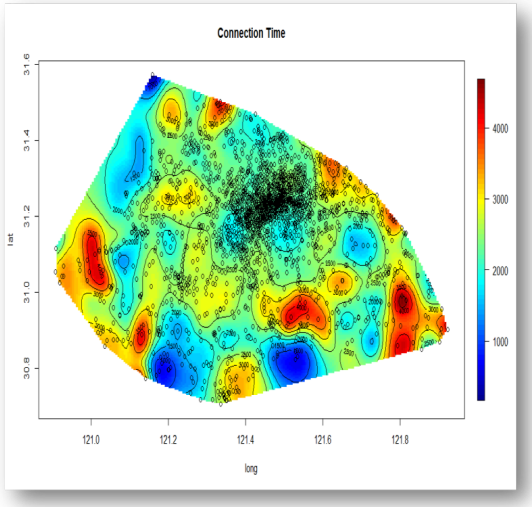

Fig. 7. Sum connection time contour plot

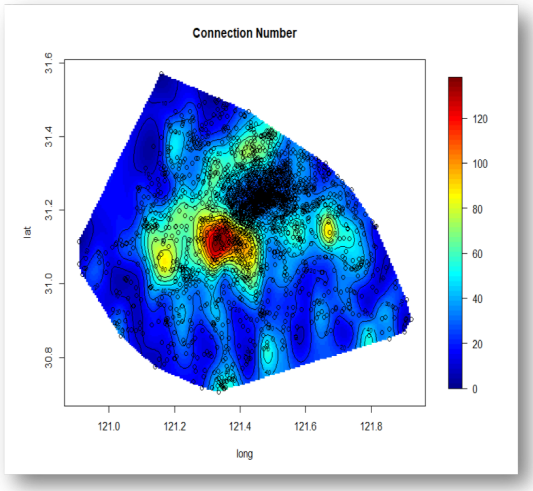

Fig. 9. connection number contour plot

240.68. This result is not far from good because the connection number cannot possibly be a negative number and the range is too big given the observed samples.

E. Geospatial Bayesian

We have also run a Bayesian model to estimate the number of connection. In this work, we have set the priors of σ and τ as the inverse of gamma with parameters $(0.1, 0.1)$. Its prediction of the mean can be shown in Fig.12 and Fig.13, and residual estimates can be shown in Fig.14 and Fig.15. As we can see from the results, although the estimation is spatially correlated, it different from the population statistics because there only 500 samples inputted to the model and 1000 times sample. The mean estimator is smoother than the residual estimator even though the mean estimator seems not a reasonable estimate of the population distribution.

Additionally, we have run with Bayesian models with $exponential$ got DIC 4351 and $spherical$ 4358, which indicated that $exponential$ slightly better than $spherical$.

F. Spatio-Temporal Bayesian

Finally, we have run a Spatio-Temporal Bayesian with GP with a small portion of dataset considered limited computation resources we have. The model is trained with 200 records and prediction 100 records, and the prediction results can be shown Fig.(16) and Fig.17. It seems a reasonable prediction given such limited data. Admittedly, the results have to be compare the actual true value from the data and compute the overall error to verify the performance of the model. Currently, the prediction results variants and far different from the population distribution.

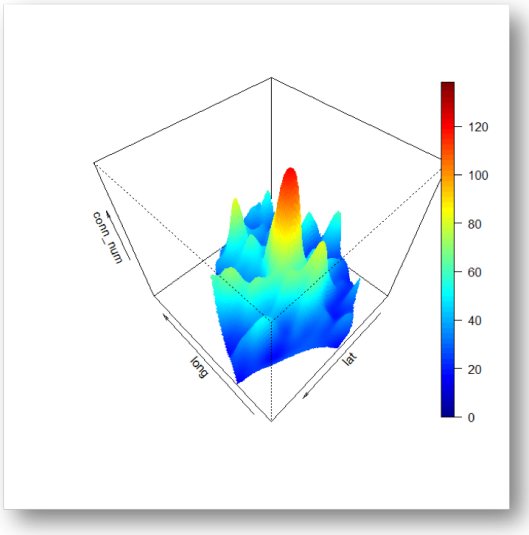

Fig. 10. Connection number 3d plot

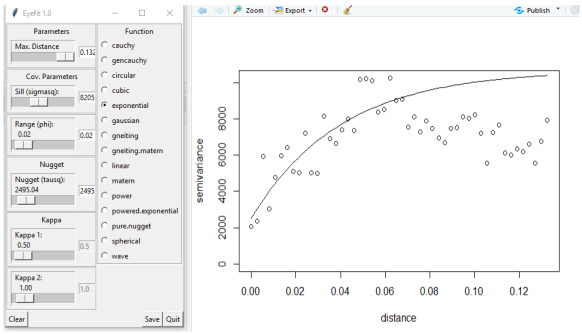

Fig. 11. Variogram fit for Number of connections

IV. Future work

As we can see from the above results, the analysis and prediction results are sufficient yet. First, we can perform geospatial data mining to get more features to further analysis. Second, Spatio-Temporal Bayesian modeling in the above work is just nutshell; therefore, we need further models that can explore and exploit the dataset. Third, online learning manner would be suitable for our application scenarios since the application produces the data over time and the model can learn timely. Lastly, Recurrent Neural Network could be considered because it has chain-like time slot architecture that can address the time-series problem.

V. Conclusion

In this work, we have performed data preprocessing and remove the outliers from the dataset; additionally, we have visualized the dataset and in varies ways to explore the data. Furthermore, we have run many experiment with geospatial statistics methods, visualize and analysis results. Lastly, we have a tested Spatio-Temporal Bayesian model and produced the results. The goal of this work predicts the connection

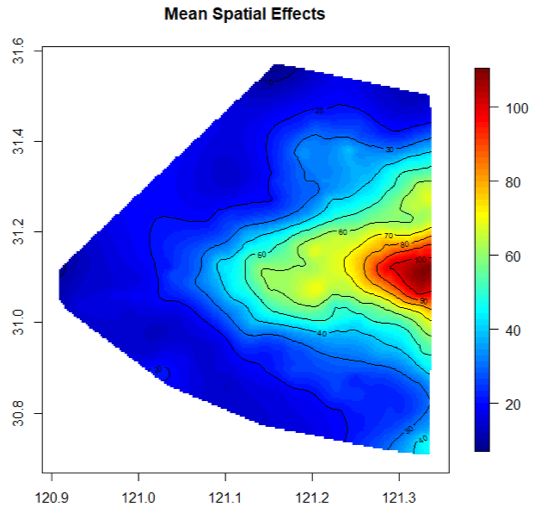

Fig. 12. Mean of connection number contour plot estimated by Bayesian

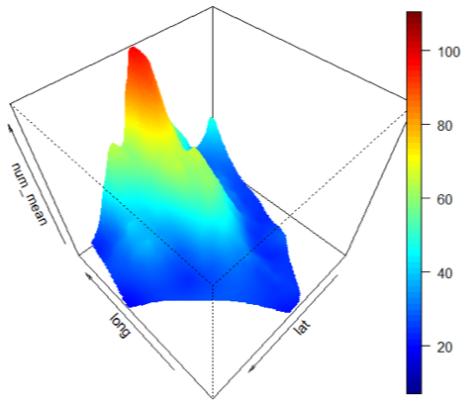

Fig. 13. mean connection number estimated by Bayesian

of users space-time distribution so that many other works such as content offloading, computing offloading and content recommendation. However, this is an early work, and it still demands much work to complete this work.

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] D. Chen, N. Zhang, Z. Qin, X. Mao, Z. Qin, X. Shen, and X.-Y. Li, "S2m: A lightweight acoustic fingerprints-based wireless device authentication protocol," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 88–100, 2017.

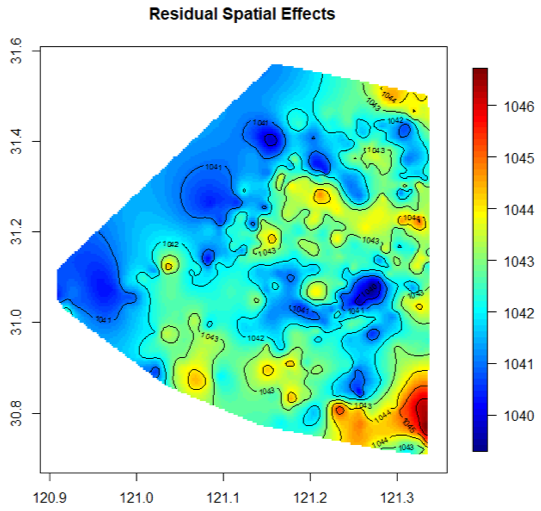

Fig. 14. Residual of connection number contour plot estimated by Bayesian

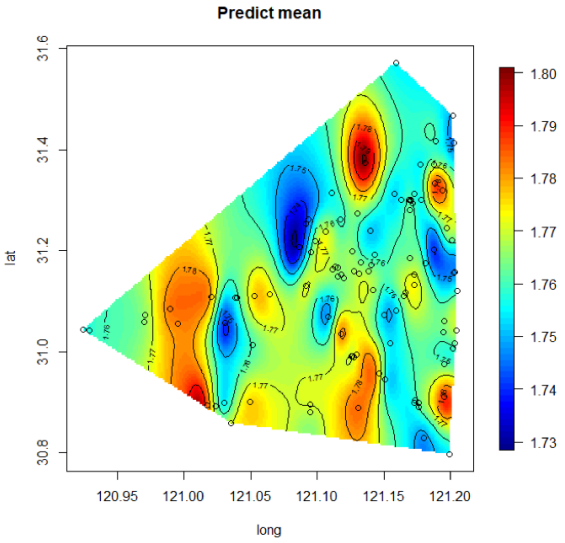

Fig. 16. Mean of connection number contour plot prediction by Spatio-Temporal Bayesian

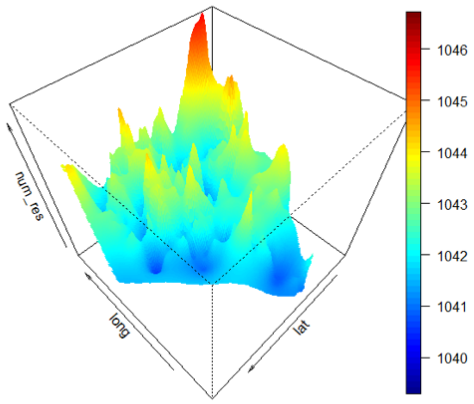

Fig. 15. Residual of connection number estimated by Bayesian

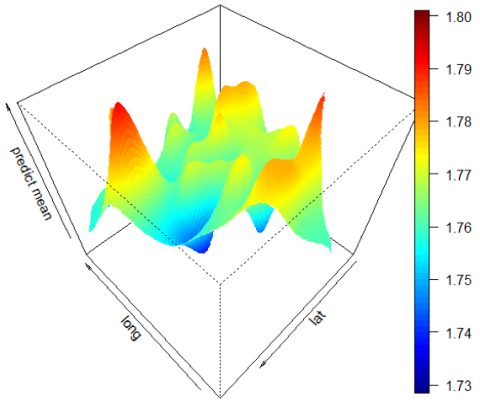

Fig. 17. Mean of connection number prediction by Spatio-Temporal Bayesian

[3] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update," *Cisco 20162021 White Paper*, vol. Document ID:1454457600805266, 2017.

[4] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, "Synergy of big data and 5g wireless networks: opportunities, approaches, and challenges," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 12–18, 2018.

[5] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, Feb 2018.

[6] S. Mu, Z. Zhong, D. Zhao, and M. Ni, "Joint job partitioning and collaborative computation offloading for internet of things," *IEEE Internet of Things Journal*, 2018, DOI: 10.1109/JIOT.2018.2866945.

[7] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, Dec 2013.

[8] G. Jia, G. Han, J. Jiang, S. Chan, and Y. Liu, "Dynamic cloud resource management for efficient media applications in mobile computing environments," *Personal and Ubiquitous Computing*, vol. 22, no. 3, pp. 561–573, 2018.

[9] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5g systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131–139, February 2014.

[10] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[11] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "Qoe driven decentralized spectrum sharing in 5g networks: potential game approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7797–7808, 2017.

[12] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K.

Soong, and J. C. Zhang, "What will 5g be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[13] Y. Hao, M. Chen, L. Hu, M. S. Hossain, and A. Ghoneim, "Energy efficient task caching and offloading for mobile edge computing," *IEEE Access*, vol. 6, pp. 11 365–11 373, 2018.

[14] W. Yuan, C. Li, D. Guan, G. Han, and A. M. Khattak, "Socialized healthcare service recommendation using deep learning," *Neural Computing and Applications*, vol. 30, no. 7, pp. 2071–2082, 2018.

[15] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: modeling and methodology," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 77–83, August 2016.

[16] J. Xu and S. Ren, "Online learning for offloading and autoscaling in renewable-powered mobile edge computing," in *Proceedings of IEEE GLOBECOM*, 2016.

[17] E. Batu, M. Bennis, and M. Debbah, "A transfer learning approach for cache-enabled wireless networks," in *Proceedings of WiOpt*, 2015.

[18] B. N. Bharath, K. G. Nagananda, and H. V. Poor, "A learning-based approach to caching in heterogenous small cell networks," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1674–1686, April 2016.

[19] T. Hou, G. Feng, S. Qin, and W. Jiang, "Proactive content caching by exploiting transfer learning for mobile edge computing," in *Proceedings of IEEE GLOBECOM*, 2017.

[20] M. S. ElBamby, M. Bennis, W. Saad, and M. Latva-aho, "Content-aware user clustering and caching in wireless small cell networks," in *Proceedings of IEEE ISWCS*, 2014.

[21] A. Sadeghi, F. Sheikholeslami, and G. B. Giannakis, "Optimal and scalable caching for 5g using reinforcement learning of space-time popularities," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 180–190, Feb 2018.

[22] F. Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong, "Mobile multimedia recommendation in smart communities: A survey," *IEEE Access*, vol. 1, pp. 606–624, 2013.

[23] S. K. S. Khandoker Shuvo Bakar, "sptimer: Spatio-temporal bayesian modeling using r," *Journal of Statistical Software*, vol. 64, p. 4135, 2 2015.

[24] S. Wang, Y. Zhao, L. Huang, J. Xu, and C.-H. Hsu, "Qos prediction for service recommendations in mobile edge computing," *Journal of Parallel and Distributed Computing*, vol. 127, pp. 134 – 144, 2019.