SPATIO-TEMPORAL PRRS EPIDEMIC FORECASTING VIA FACTORIZED DEEP GENERATIVE MODELING

Mohammadsadegh Shamsabardeh[†]   *Bahar Azari*[⋆]   *Beatriz Martínez-López*[†] [*]

[†]Center for Animal Disease Modeling and Surveillance, University of California, Davis, CA, USA
[⋆]Northeastern University, Boston, MA, USA

ABSTRACT

We propose an epidemic analysis framework for the outbreak prediction in the livestock industry, focusing on the study of the most costly and viral infectious disease in the swine industry – the PRRS virus. Using this framework, we can predict the PRRS outbreak in all farms of a swine production system by capturing the spatio-temporal dynamics of infection transmission based on the intra-farm pig-level virus transmission dynamics, and inter-farm pig shipment network. We simulate a PRRS infection epidemic based on the shipment network and the SEIR epidemic model using the statistics extracted from real data provided by the swine industry. We develop a hierarchical factorized deep generative model that approximates high dimensional data by a product between time-dependent weights and spatially dependent low dimensional factors to perform per farm time series prediction. The prediction results demonstrate the ability of the model in forecasting the virus spread progression with average error of NRMSE = 2.5%.

Index Terms— Spatio-temporal Analysis, Infectious Disease, Deep Generative Model, PRRS Outbreak Modeling, Epidemic Network Analysis.

1. INTRODUCTION

The Porcine Reproductive and Respiratory Syndrome (PRRS) is arguably the most challenging and costly viral infectious disease in the US swine industry [1]. It is the high cost of disease control practices such as testing and vaccination as well as the financial damage caused by an outbreak that highlights the need to develop predictive models that can help to identify farms at high risk of infection to support risk-based, more cost-effective, target interventions. Such a framework will allow for more efficient testing, vaccination and outbreak prevention. Due to the high level of specialization in the swine industry, vast amount of data has been collected by the swine production systems. However, they have not yet
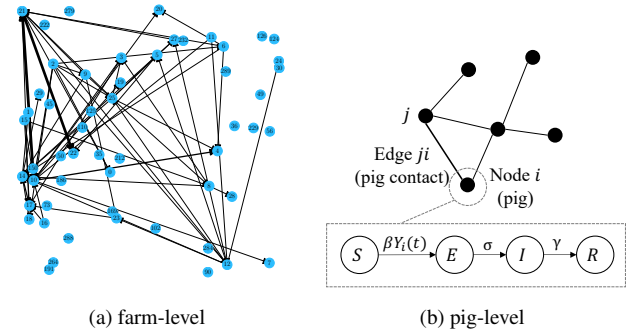
(a) farm-level   (b) pig-level

Fig. 1: Contact Network. (a) The swine shipment network (directed graph). The premises are displayed by a number-labeled node and edge weights corresponds to the shipment rate. The between-premises shipment rate network is showcased for 10% of nodes randomly selected among over 300 existing nodes. (b) **Top:** Pig level network graph. **Bottom:** State-transition diagram for a single node.

been exploited enough due to difficulty of data access, integration and analyses (i.e., data are not consistently gathered, are non-standardized which makes their integration difficult, and are usually scattered across stakeholders). Examples of these data include diagnostic information, including the number of infected or dead animals, animal movements between farms or production data, which can give insights regarding farm health status and its contact network. Usually, these data do not satisfy the granularity required for learning an advanced predictive model (e.g. diagnostic samples are only taken once or twice per month per farm). However, using the real-world data, we can simulate epidemics to produce fine-grained time series data to analyze it further with an advanced novel prediction method based on a generative and variational inference model.

The direct contact with an infected pig is the main pathway for PRRS virus transmission. The network of between-farm movements, shown in Fig. 1a together with the intra-farm (local) pig-level contact pattern based on Susceptible-Exposed-Infectious-Recovered (SEIR) epidemic model (see Fig. 1b) allows for constructing a system-level (global) pig-level disease transmission contact network, [2]. Network-

based SIR or SIR-extended epidemic models have been extensively studied in the literature [3]. In [4], Newman studied a network-based SIR epidemic model where infection is transmitted through a random network of contacts between individuals. The disease transmission contact network is a probabilistic graph that can be sampled to generate virtual contact. Using this, we generate fine-grained spatio-temporal time-series data based on statistics of real-world data.

The spatio-temporal data is often considered to have a high level of correlation between spatial dimensions, and, therefore, they can be assumed to be governed by a smaller number of underlying components. For modeling the temporal dynamic of the time series including the number of susceptible, infected, or recovered pigs, we employ a non-linear vector auto-regressive latent model inspired by the work of [5]. Our spatio-temporal time-series data is first *factorized* into temporal weights and spatial factors. The temporal weights are modeled using a non-linear auto-regressive model parameterized by neural networks governed with a Markovian chain of discrete switches to capture higher-order multimodal latent dependencies, [6, 7, 8, 9, 10].

2. TIME SERIES DATA SIMULATION

Based on the rich database of an extensive anonymous swine production system located in the Midwest of the United States, we have access to farm-level pig shipment data, and PRRSV testing results [11]. From 2006 to 2021, there have been over 260,000 movement records to or from farms within this production system. For each movement entry, the data includes the source and destination information, the number of transported pigs, and the date of the movement. Based on the farm-level shipment data we generate a *farm-level movement network* for the entire production system. Furthermore, the frequent PRRSV testing in each farm gives insight into how the virus is transmitted, e.g., what is the virus's transmission rate, incubation time, etc. Using the SEIR model we can produce a *pig-level contact network*, [2]. The combination of this two network built on the statistics extracted from our real data results in an intricate contact network by which we can simulate complex time series data showing the number of susceptible, infected or recovered pigs in each farm in the production system.

To create the farm-level movement network we build a probabilistic graph, i.e., a graph in which the existence of edges is uncertain with some probability. A node in the graph represents a farm, while the weight of an edge is proportional to the shipment rate (probability of shipment) between farms. Fig. 1a shows a part of the shipment network of over 300 farms for our real data. The edge thickness is representative of the shipment rate. For local (intra-farm) pig-level contact network in each farm, we consider a basic random graph model based on Erdös—Rényi model [12, 2], that produces pig contact graphs with an edge probability of 0.5 between any pair

of pigs. The global (inter-farm) pig-level contact network is constructed when we sample a random generalization of a between-farm shipment over time (each day) and as a result, we create pig contact between farms (see Fig. 1b, top).

The simulation is formed on the network-based SEIR epidemic model for PRRS. In a network-based model, we consider a graph in which nodes represent individual pigs, and edges indicate direct or indirect contacts between pigs, which are considered infection pathways of PRRS. Each animal can be in one of the four states, Susceptible (S), Exposed (E), Infected (I), or Recovered (R) as the result of the epidemic progression. The state-transition diagram between these states is shown in figure. 1b, Bottom. In the generated swine pig-level network, a PRRS outbreak is introduced by randomly selecting a pig farm, and infecting an arbitrary random number of pigs. We collect several time snapshots representing the progression dynamics of the disease spread, such as the number of infected pigs in each farm over time. The healthy pigs which are free from PRRS virus infection are classified as Susceptibles. If such a healthy pig comes into contact with infected pigs containing the virus, it may get infected at the rate $\beta Y_i(t)$, where $Y_i(t)$ is the number of infected neighbors of recoverednode i at time t. If the transmission of pathogen occurs, a healthy pig enters into the Exposed group where it stays for the duration of the incubation period. On average, this period is denoted by $1/\sigma$. Once it shows symptoms, it moves into the Infected group. It stays there for an average time of $1/\gamma$ before it is recovered. We choose the parameters values $\beta = 0.087, \sigma = 7, \gamma = 6.5$ based on [13, 14]. This dynamic produces a spatio-temporal time series from all farms over time that can be used for predication modeling in the next section (Sec. 3).

3. FARM DISEASE PROPAGATION PREDICATION

Our spatio-temporal data indicates the number of pigs categorized within a particular stage, e.g., infected, recovered, etc., in every time instance in each farm. We denote this data as the matrix $X \in \mathbb{R}^{T \times D}$, where T is the number of time points and D the number of spatial locations, e.g., the number of farms. Building on previous work by [5], our assumption is that X can be decomposed into a weighted summation of $K \ll D$ factors over time as:

$$X \approx [w_1, \cdots, w_T]^\top [f_1; \cdots; f_K] = W^\top F, \quad (1)$$

where $f_k \in \mathbb{R}^D$ is the kth spatial factor and $w_t \in \mathbb{R}^K$ is the weight vector at time t. Our intuition for adopting this model for some pig-specific collected measurements in D farm over T time points is that there are $K \ll D$ underlying factors using which we can approximate the overall dynamics of the disease propagation in the data. Our main goal is to predict the future outbreak behavior given the past time samples in each farm.

We assume that the weights, $W = \{w_t\}_{t=1}^T$, are generated according to a set of temporal lags, ℓ, through a deep proba-
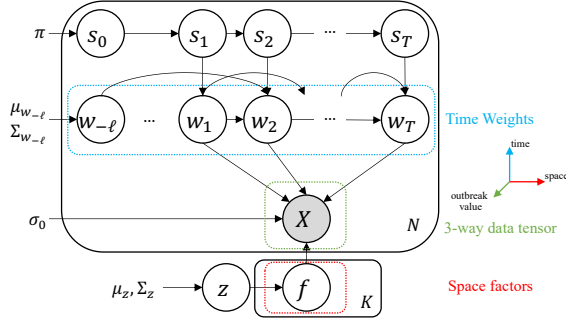
Fig. 2: Probabilistic graphical model.

bilistic switching auto-regressive model. These weights are furthermore governed by a Markovian chain of discrete latent states, $\mathcal{S} = \{s_t\}_{t=1}^{T}$ as follows: $w_t \sim p(w_t|w_{t-\ell}, s_t)$, $s_t \sim p(s_t|s_{t-1})$. In addition, we assume that spatial factors, $F = \{f_k\}_{k=1}^{K}$, are controlled by a shared low dimensional latent variable, z, as follows: $f_{1:K} \sim p(F|z)$, $z \sim p(z)$. Fig. 2 shows the relation among above-mentioned random variables in a probabilistic graphical model diagram form.

We train the model using stochastic variational methods [15, 16, 17, 18] by approximating the posterior $p_\theta(\mathcal{S}, W, z, F|X)$ using a variational distribution $q_\phi(\mathcal{S}, W, z, F)$, and by maximizing a lower bound (known as ELBO) $\mathcal{L}(\theta, \phi) \leq \log p_\theta(X)$:

$$
\begin{aligned}
\mathcal{L}(\theta, \phi) &= \mathbb{E}_{q_\phi(\mathcal{S}, W, z, F)} \left[\log \frac{p_\theta(X, \mathcal{S}, W, z, F)}{q_\phi(\mathcal{S}, W, z, F)} \right] \quad (2) \\
&= \log p_\theta(X) - \mathrm{KL}(q_\phi(\mathcal{S}, W, z, F) \,||\, p_\theta(\mathcal{S}, W, z, F|X)).
\end{aligned}
$$

By maximizing the bound with respect to the parameters θ, we learn the generative distribution over datasets $p_\theta(X)$, and by maximizing the bound over the parameters ϕ, we do Bayesian inference by approximating the distribution $q_\phi(\mathcal{S}, W, z, F) \simeq p_\theta(\mathcal{S}, W, z, F|X)$ over latent variables for each data point. According to the graphical model in Fig. 2, the joint distribution of observations and latents will be:

$$
\begin{aligned}
p_\theta(X, \mathcal{S}, \mathcal{Z}) = p(F|z)p(z) \prod_{n=1}^{N} p(X_n|W_n, F)p(w_{n,-\ell})p(s_{n,0}) \\
\prod_{t=1}^{T} p(s_{n,t}|s_{n,t-1})p(w_{n,t}|w_{n,t-\ell}, s_{n,t}), \quad (3)
\end{aligned}
$$

where $\mathcal{Z} = \{W, z, F\}$). Furthermore, we assume a fully factorized variational distribution for the latent variables posterior as:

$$
q(\mathcal{S}, \mathcal{Z}) = q(F)q(z) \prod_{n=1}^{N} q(w_{n,-\ell})q(s_{n,0}) \prod_{t=1}^{T} q(s_{n,t})q(w_{n,t}). \quad (4)
$$

3.1. Generative Parameters

Here we describe the generative distribution parameters and model assumptions. We assume the variable $s_{n,t}$ to represents

a categorical variable of dimensionality S, i.e., the number of modes/switches that a system can be in at a specific time t. The sequence of the discrete latents, $s_{n,1:T}$, are in form of a Markov chain and govern the state transitions over time with distributions:

$$
\begin{aligned}
p_\theta(s_t|s_{t-1}) &= \mathrm{Cat}(\boldsymbol{\Phi}_\theta \, \boldsymbol{\pi}_{s_{t-1}}) \\
q_\phi(s_{t-1}) &= \mathrm{Cat}(\boldsymbol{\pi}_{s_{t-1}}), \quad (5)
\end{aligned}
$$

where $\boldsymbol{\pi}_{s_{t-1}} = [\pi_1, \cdots, \pi_S]$ represents the probabilities of the categorical distribution for s_{t-1}, and $\boldsymbol{\Phi}_\theta \in \mathbb{R}^{S \times S}$ is a valid probability transition matrix.

For the temporal weights, w_t, we assume a switching Gaussian dynamic for the temporal latent transitions governed by the discrete latent states, s_t. In other words, we assume that the marginal distribution of temporal weights follows a Gaussian mixture distribution in the latent space, as:

$$
p_\theta(w_t|w_{t-\ell}, s_t = s) = \mathcal{N}\Big(\mu_{\theta_s}^{w}(w_{t-\ell}), \Sigma_{\theta_s}^{w}(w_{t-\ell})\Big),
$$

where $s \in \{1, \cdots, S\}$, and state-specific $\mu_{\theta_s}^{w}(\cdot)$ and diagonal $\Sigma_{\theta_s}^{w}(\cdot)$ are parameterized by multilayer perceptrons (MLPs), hence, follow a *nonlinear* vector auto-regressive model given $w_{t-\ell}$. Namely, we feed $w_{t-\ell}$ to a multi-head MLP for estimating the Gaussian parameters, e.g.,

$$
\mu_{\theta_s}^{w} = \mathrm{FC}_s(h_s), \quad h_s = \sum_{l \in \ell} \sigma(\mathrm{FC}_{s,l}(w_{t-l})),
$$

where FC denotes a fully connected layer, and σ is a nonlinear activation function.

For the spatial factors, F, we assume a diagonal Gaussian distribution for spatial factors parameterized with an MLP as

$$
p_\theta(F|z) = \mathcal{N}\big(\mu_\theta^{F}(z), \Sigma_\theta^{F}(z)\big) \quad (6)
$$

where z is sampled from a normal distribution: $z \sim \mathcal{N}(0, I)$. The latent z is introduced as a low dimensional spatial embedding that encourages the estimation of a multimodal distribution among spatial factors. Given the temporal weights and spatial factors, we reconstruct the data by consolidating the two factorized part as:

$$
X_n \sim p_\theta(X_n|W_n, F) = \mathcal{N}\Big(\big[w_{n,1}, \cdots, w_{n,T}\big]^\top F, \sigma_0^2 \Big), \quad (7)
$$

where σ_0 is a hyperparameter for observation noise.

3.2. Variational Parameters

The trainable variational parameters, ϕ, are assumed to have fully factorized distributions. The variational distribution for the continuous variables, $q(z; \phi^z)$, $q(F; \phi^F)$, and $\{q(w_{n,t}; \phi_{n,t}^w)\}_{n=1, t=-\ell}^{N, T}$, are considered to be Gaussian distributions with diagonal covariances. In addition, the variational parameters for the distribution of discrete latents, $\{q(s_{n,t}; \phi_{n,t}^s)\}_{n=1, t=1}^{N, T}$, are considered based on the mean-field approximation assumption to compensate information loss (see [5] for more detail.)
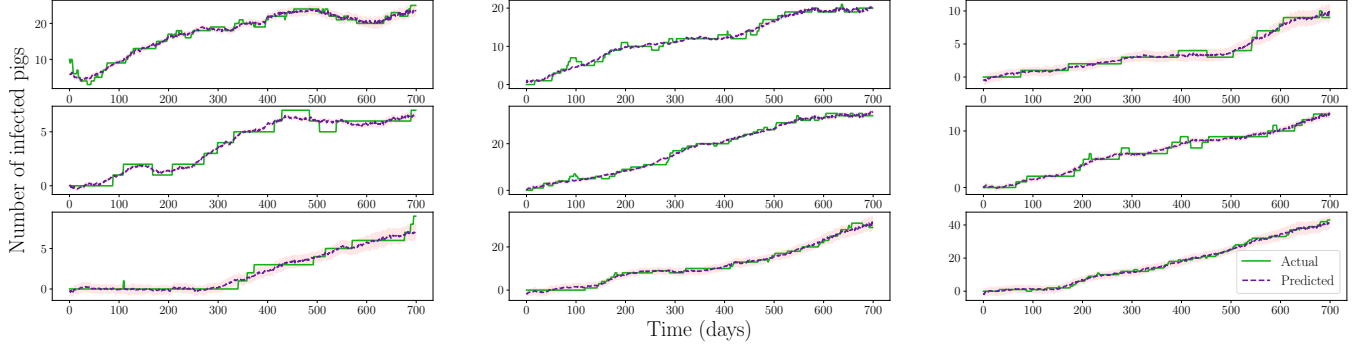
Fig. 3: Short-term (one-day) predication. Each plot demonstrates the actual number of infected pigs in the simulated data (solid green), the mean estimate of the predictive model (dashed purple), and the standard deviation of the prediction estimate (shaded red error bar). Each row represents neighbouring farms that are connected in terms of pigs movement.

3.3. Training Procedure

The Monte-Carlo estimate of the gradient of ELBO is computed with respect to generative, θ, and variational, ϕ, parameters using a re-parameterized sample, [17], from the posterior of continuous latents, $\{W, z, F\}$. For the discrete latent, $\mathcal{S}$, however, we compute the expectations by summing over the $\mathcal{S}$, without the need for explicit sampling. This regularizes the $\mathcal{S}$ nonlinear auto-regressive priors based on their corresponding weighting. We can analytically calculate the Kullback-Leibler (KL) divergence terms of ELBO for both multivariate Gaussian and categorical distributions, which leads to lower variance gradient estimates and faster training as compared to e.g., noisy Monte Carlo estimates often used in literature. We use the Adam optimizer, [19], with learning rate of 0.01 for training. We initialized all the parameters randomly, and adopted a linear KL annealing [20] schedule to increase from 0.01 to 1 over the course of 100 epochs.

4. EXPERIMENTAL RESULTS

We used time series of epidemic progression from over 300 farms simulated for 700 time points. We kept last 20% of the time series as the test set. We then performed a short-term prediction tasks by adopting a rolling prediction scheme reported in [21]. For short-term prediction, the next time point is predicted on the test set using the generative model and spatial factors learned on the train set. We reported the test set normalized root-mean-square error (NRMSE%), which is related to the expected negative test-set log-likelihood for the case of Gaussian distributions, and it is used for evaluating the predictive generative models. We obtained NRMSE of 2.5% averaged over all the farms. Fig. 3 shows the number of infected pigs over time for nine selected farms. In this figure, we illustrate the actual number of infected pigs in the simulated data with a solid green curve, the mean estimate of the predictive model with a dashed purple curve, and the standard deviation of the data with a shaded red error bar. Each row in

the figure represents a group of relatively highly connected farms in terms of the frequency of the pig shipments. Note that each group show a relatively strong correlation regarding the outbreak progression and the predictive model was able to capture this.

One observation regarding the performance of the model in situations when we do not have a curve with smooth behavior is the fact that when the increase and decrease of the number of infected pigs is abrupt, the model tends to converge to a middle point between the current value and the future one. This issue can be addressed using a model selection choice. Specifically, we can control the number of factors K in order to increase the flexibility of the model in capturing different levels of curve smoothness.

5. DISCUSSION & CONCLUSION

The PRRS outbreak cause an economic loss of over \$664 million annually [1], which can be significantly mitigated by early detection and risk-based intervention practices. Direct contact is the main disease transmission pathway. Therefore, the pig contact network provides a substantial basis to develop an outbreak prediction framework. We create a system-wide pig contact network by combining the SEIR epidemic model based on intra-farm infection transmission parameters, and inter-farm pig shipment network. We presented a hierarchical factorized deep generative model of our spatio-temporal data that can capture the underlying dynamics of the disease spread with the aim to predict the number of infected pigs in all farms. Our result demonstrates the ability of the model in forecasting the virus spread progression with an average one-day prediction error of NRMSE = 2.5%. A potential future direction is to incorporate the per farm disease transmission parameters for the SEIR model to represent variations in the disease control implementation. Additionally, indirect disease transmission pathways, such as airborne, can be included in the framework for a more comprehensive analysis.

6. REFERENCES

[1] Derald J Holtkamp, James B Kliebenstein, Eric J Neumann, Jeffrey J Zimmerman, Hans F Rotto, Tiffany K Yoder, Chong Wang, Paul E Yeske, Christine L Mowrer, and Charles A Haley, "Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on united states pork producers," *Journal of Swine Health and Production*, vol. 21, no. 2, pp. 72–84, 2013.

[2] Tanvir Ferdousi, Sifat Afroj Moon, Adrian Self, and Caterina Scoglio, "Generation of swine movement network and analysis of efficient mitigation strategies for african swine fever virus," *PloS one*, vol. 14, no. 12, pp. e0225785, 2019.

[3] Kyuyoung Lee, Dale Polson, Erin Lowe, Rodger Main, Derald Holtkamp, and Beatriz Martínez-López, "Unraveling the contact patterns and network structure of pig shipments in the united states and its association with porcine reproductive and respiratory syndrome virus (prrsv) outbreaks," *Preventive veterinary medicine*, vol. 138, pp. 113–123, 2017.

[4] Mark EJ Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, pp. 016128, 2002.

[5] Amirreza Farnoosh, Bahar Azari, and Sarah Ostadabbas, "Deep switching auto-regressive factorization: Application to time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 7394–7403.

[6] Chaw-Bing Chang and Michael Athans, "State estimation for discrete systems with switching parameters," *IEEE Transactions on Aerospace and Electronic Systems*, , no. 3, pp. 418–425, 1978.

[7] Zoubin Ghahramani and Geoffrey E Hinton, "Switching state-space models," Tech. Rep., Citeseer, 1996.

[8] Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski, "Bayesian learning and inference in recurrent switching linear dynamical systems," in *Artificial Intelligence and Statistics*, 2017, pp. 914–922.

[9] J Nassar, SW Linderman, M Bugallo, and IM Park, "Tree-structured recurrent switching linear dynamical systems for multi-scale modeling," in *International Conference on Learning Representations (ICLR)*, 2019.

[10] Philip Becker-Ehmck, Jan Peters, and Patrick van der Smagt, "Switching linear dynamics for variational bayes filtering," 2019.

[11] M Shamsabardeh, Shabaz Rezaei, Jose Pablo Gomez, Beatriz Martínez-López, and Xin Liu, "A novel way to predict prrs outbreaks in the swine industry using multiple spatio-temporal features and machine learning approaches," *Frontiers in Veterinary Science*, vol. 6, 2019.

[12] P. Erdos and A. Rényi, "On random graphs i," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[13] Phithakdet Phoo-ngurn, Chanakarn Kiataramkul, and Farida Chamchod, "Modeling the spread of porcine reproductive and respiratory syndrome virus (prrsv) in a swine population: transmission dynamics, immunity information, and optimal control strategies," *Advances in Difference Equations*, vol. 2019, no. 1, pp. 1–12, 2019.

[14] Céline Charpin, Sophie Mahé, André Keranflec'h, Catherine Belloc, Roland Cariolet, Marie-Frédérique Le Potier, and Nicolas Rose, "Infectiousness of pigs infected by the porcine reproductive and respiratory syndrome virus (prrsv) is time-dependent," *Veterinary Research*, vol. 43, no. 1, pp. 1–11, 2012.

[15] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley, "Stochastic variational inference," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1303–1347, 2013.

[16] Rajesh Ranganath, Chong Wang, Blei David, and Eric Xing, "An adaptive learning rate for stochastic variational inference," in *International Conference on Machine Learning*, 2013, pp. 298–306.

[17] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *stat*, vol. 1050, pp. 1, 2014.

[18] Danilo Jimenez Rezende and Shakir Mohamed, "Variational inference with normalizing flows," in *International Conference on Machine Learning*, 2015, pp. 1530–1538.

[19] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[20] Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio, "Generating sentences from a continuous space," *CoNLL 2016*, p. 10, 2016.

[21] Xinyu Chen, Zhaocheng He, Yixian Chen, Yuhuan Lu, and Jiawei Wang, "Missing traffic data imputation and pattern discovery with a bayesian augmented tensor factorization model," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 66–77, 2019.