2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)

Building Explainable Predictive Analytics for Location-Dependent Time-Series Data

Yao-Yi Chiang*, Yijun Lin†, Meredith Franklin‡, Sandrah P. Eckel§, José Luis Ambite¶
Spatial Sciences Institute, Department of Computer Science†, Department of Preventive Medicine‡§, Information Sciences Institute¶*
University of Southern California
Los Angeles, CA USA
*yaoyic@usc.edu, †yijunlin@usc.edu, ‡meredith.franklin@usc.edu,§eckel@usc.edu,¶ambite@isi.edu

Wei-Shinn Ku
Department of Computer Science and Software Engineering
Auburn University
Auburn, AL USA
weishinn@auburn.edu

***Abstract*—There are increasing numbers of online sources of real-time and historical location-dependent time-series data describing various types of environmental phenomena, e.g., traffic conditions and air quality levels. When coupled with the information that characterizes the natural and built environments, these location-dependent time-series data can help better understand interactions between and within human social systems and the ecosystem. Nevertheless, these data are still limited by their spatial and temporal resolution for downstream use (e.g., generating residential-level environmental exposures for human health studies). In this paper, we present a vision of a general machine learning framework for explainable predictive analytics for location-dependent time-series data. The framework will effectively deal with data- and model-related challenges for general scientific predictive analytics on spatiotemporal environmental phenomena. The challenges include how to identify the main features driving the phenomena, how to handle complex spatiotemporal variations in the phenomena, and how to utilize sparse ground truth measurements for training and validation. The resulting framework will enable fine spatial and temporal scale environmental exposure assessment and allow researchers to carry out unprecedented inquiries, such as understanding relationships between health outcomes and long-term air pollution exposures.**

***Index Terms*—spatial data science, spatiotemporal data, predictive analytics, machine learning**

I. Vision

With the widespread use of sensors that measure a variety of environmental phenomena and the fast-growing application of the Internet of Things (IoT), there are increasing amounts of real-time and historical spatiotemporal data available to the scientific community through large-scale and publicly available databases and application programming interfaces (API). For example, traffic sensors provide real-time assessments of the traffic volume along the freeways and arterial roads (e.g., [1]). Air quality monitoring stations offer both real-time and long-term measurements of air pollutants at multiple locations.[1] Effective uses of these location-dependent time-series data can have wide-reaching impacts across multiple domains such as smart cities, urban planning, policymaking, and public health. For instance, bus GPS trajectory data can help generate performance metrics for investigating public transportation systems towards reducing the operating costs and increasing ridership [2]. Regardless, these data are limited by the number of available sensors, which are often not sufficient in terms of their spatial and temporal resolutions for downstream use (e.g., the human health studies that require residential level environmental exposures [3]). This data-related challenge makes it difficult to use sensor data alone to accurately track the phenomena with complex spatiotemporal variations. Therefore, to fully exploit the potential of available sensors for close modeling and tracking of human-environment interactions, there is an urgent need for a general method that can effectively leverage the location-dependent time-series data, together with the information of the natural and built environment, for the accurate prediction of spatiotemporal phenomena over fine spatial and temporal resolutions.

Traditionally, spatiotemporal predictions require labor- and expert-intensive efforts to process heterogeneous data to then make meaningful interpretations or predictions from the data (e.g., [4]). Machine learning technologies could alleviate these requirements, but they typically require large amounts of training data, which are costly to produce. Moreover, careful considerations of the spatial relationships in the data are essential but often overlooked, causing inaccurate results (e.g., overfitting from spatial non-stationarity and statistical bias from spatial autocorrelations and the modifiable areal unit problem, see [5, 6, 7]). These considerations are challenging to realize for complex machine learning models (e.g., deep neural networks) without a deep understanding of spatial

[1]https://www.epa.gov/air-quality-management-process/managing-air-quality-ambient-air-monitoring

978-1-7281-6737-4/19/$31.00 ©2019 IEEE
DOI 10.1109/CogMI48466.2019.00037

sciences. Hence, very often, machine learning models would only incorporate simple assumptions (e.g., nearby objects are similar) [8] or directly treat the spatial dimension as yet another independent variable (e.g., adding longitude and latitude as location indicators in the model). This model-related challenge hinders the studies that require an effective investigation and interpretation of the interactions between these location-dependent time-series data and the contributors (e.g., natural and built environment) to fully understand and explain the phenomena.

We envision a general machine learning architecture for explainable fine-scale predictions of universal location-dependent time-series from limited observations. The resulting architecture will contribute to fine spatial and temporal scale assessments of various environmental phenomena (e.g., urban air quality and noise levels) . Linking such results to health studies (e.g., cancer and asthma [9, 10, 11]) or wearable sensor measurements (e.g., [12]) will enable researchers to carry out unprecedented inquiries of the relationships between health outcomes and exposure impacts.

II. Current Limitations & Challenges

Without loss of generality, here we define a scientific prediction process for spatiotemporal data as follows. The process aims to predict a value (e.g., $PM_{2.5}$ concentrations) on a spatial grid with a high resolution (e.g., a grid map with a uniform cell size of 500 meters by 500 meters, 500m×500m). The input data is a multi-dimensional matrix $\boldsymbol{X} = (O, F, H, W)$ representing observation values, O, and sets of environmental features, F, over the space where H and W represent the height and width of the grid, respectively. Note that O can be a null value representing no observation at a specific cell. The output matrix is $\boldsymbol{Y} = (P, H, W)$, where P represents the cell predictions (dimension=1).

Let $\boldsymbol{X}^{(t)}$ represent the input signal at time t; T' is the number of previous hours (i.e., from $t - T' + 1$ to t). The prediction technology aims to learn a function h that maps T' historical input signals to the output at t:

$$[\boldsymbol{X}^{(t-T'+1)}, \cdots, \boldsymbol{X}^{(t)}] \xrightarrow{h} [\boldsymbol{Y}^{(t)}]$$

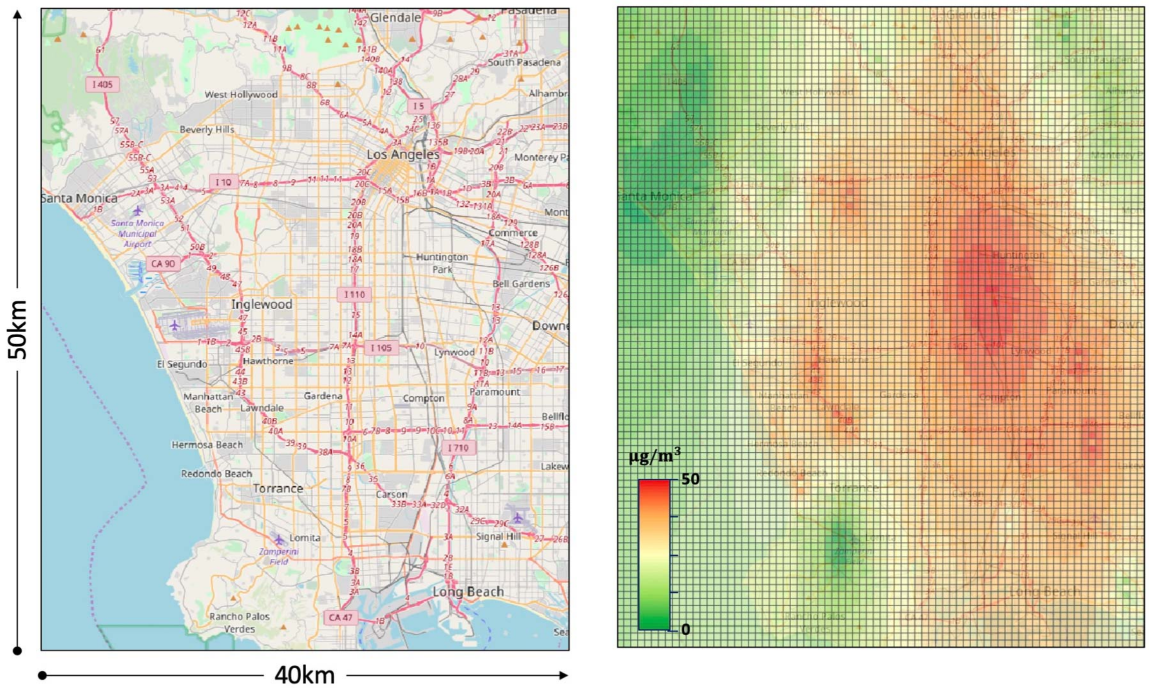
where the function h can jointly model the spatial and temporal dependencies between features in F and their interactions with the observation values O.

In general, scientific prediction processes dealing with location-dependent time-series data and their underlying spatiotemporal phenomena are inherently difficult due to several data- and modeling-related challenges. These challenges include identifying the main contributors driving the phenomena, handling of complex spatiotemporal dependencies in the data, and sparse ground truth measurements for training and validation. Specifically, current prediction technologies have the following limitations. **First**, they do not generate explainable prediction results automatically (Section II-A). Some of the existing techniques do not explicitly include explanatory variables about surrounding environmental characteristics (e.g., inverse distance weighting). Other techniques require careful considerations of manually selecting related factors using domain knowledge (e.g., land use regression [4]) or only provide "black box" results that are not interpretable (e.g., deep learning models). **Second**, existing technologies cannot effectively capture relationships between the target phenomenon and explanatory variables (e.g., built environment) in both space and time simultaneously (e.g., [13]) (Section II-B). **Third**, they do not perform well with sparse sensor locations (e.g., [5]) (Section II-C).

A. Generating Explainable Prediction Results

Generating explainable prediction results allows the explicit understanding of how the environmental characteristics contribute to spatiotemporal phenomena, such as air pollution, and enables informed policymaking or interventions (e.g., actions for air pollution prevention and control). Explainable (or interpretable) prediction models have the ability to present in terms understandable to a human [14]. Traditional spatial interpolation methods for spatial prediction tasks, such as the Inverse Distance Weighting (IDW) and ordinary Kriging [15] do not explicitly include explanatory variables about environmental characteristics such as meteorology and topography. Although these methods are generally computational efficient compared to machine learning or data-driven methods, the results are not explainable, and their ability to produce reliable estimates is limited [5]. As an example, Figure 1 shows the $PM_{2.5}$ predictions from IDW over a target area in Los Angeles County. IDW generates a smooth prediction surface over the region, which only offers a general idea about the variation of the $PM_{2.5}$ concentrations at a coarse spatial scale (e.g., the Los Angeles downtown area has poorer air quality level than other areas).

More sophisticated prediction technologies statistically investigate the direct correlations between the time-series observations and the environmental characteristics to infer spatiotemporal variations of the underlying phenomena. These technologies can use the identified correlations to explain the prediction results; however, they usually require expert-domain knowledge to decide useful predictors from a variety of environmental characteristics (i.e., independent variables). For example, for air quality prediction, one popular approach to predict long-term spatial variations in air pollution levels is land-use regression (LUR) [4, 16, 17, 18]. LUR leverages expert-selected environmental factors, including various types of geographical features (e.g., commercial and industrial), traffic conditions, population density, and meteorological data, for modeling each combination of study areas (e.g., Los Angeles County), pollutant types (e.g., $PM_{2.5}$), and spatiotemporal scales (e.g., daily average concentrations). For the LUR-type of method, the expert-selected environmental factors can change significantly across study areas. Table I provides three example cases of LUR using varying independent variables and spatial buffer sizes for separate study areas and pollutants. This makes LUR-type approaches highly localized and vulnerable to inaccuracies when applied to heterogeneous study areas. Furthermore, models for different geographic regions



(a) Target area

(b) IDW prediction results

Fig. 1. (a) The target prediction region covering an area of 50km×40km in Los Angeles (b) The aggregated prediction results of $PM_{2.5}$ concentrations in November with 500m×500m resolution using IDW

require expert-selected location characteristics before model fitting to achieve the best regression results (e.g., distance to the ocean has a high correlation to air quality in San Diego but not in every coastal city).

Another example of using expert-selected sets of environmental characteristics for prediction is the U.S. Federal Highway Administration (FHWA) Traffic Noise Model (TNM2.5) for the prediction of ambient environmental noise. Ambient environmental noise is an understudied risk factor that has been associated deleterious human mental and physical health [19]. While noise itself can be considered as an independent contributor to health risk [20], examined in a multi-pollutant context it has been shown to amplify the physical health effects associated with exposure to air pollution, including respiratory health [21], stress [22], and obesity [23]. The current practice for ambient noise modeling relies heavily on the use of deterministic acoustic models such as TNM2.5 (Figure 2). Such models require pre-defined inputs on roadway locations, type, fleet composition and traffic volumes, and tailored for assessing traffic noise for policy compliance related to highway policy.

In contrast to the methods that require expert domain knowledge, data-driven approaches using machine learning approaches, such as linear regression and random forests, enable us to identify the critical predictors by learning the feature weights (e.g., [5]). However, these algorithms usually fail to achieve a reliable performance due to their limitation

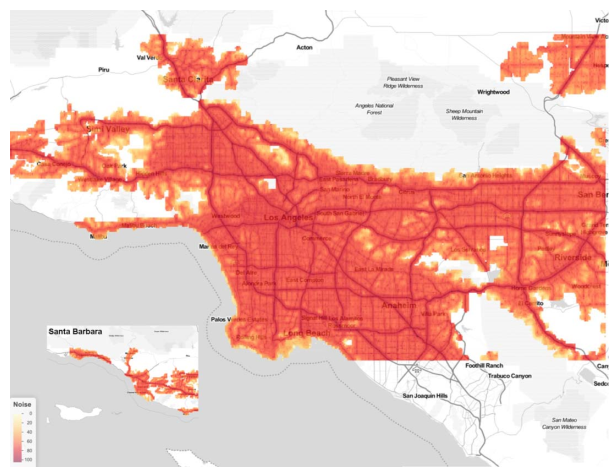


Fig. 2. TNM2.5 traffic-related noise estimates over Southern California

on learning the complex interactions between environmental features and the spatiotemporal phenomena. For example, for traffic volume prediction, the traffic conditions rely on not only spatial features (e.g., if there is a commercial area nearby) but also temporal features (e.g., workday or weekend).

Advanced machine learning approaches, e.g., deep learning models, coupled with the increasing availability of big data describing the environment, become popular in dealing with

TABLE I
EXAMPLE LAND-USE REGRESSION STUDIES

Reference	Study area	Pollutant	Selected variables	Buffer sizes
Moore et al. [16]	Los Angeles, USA	$PM_{2.5}$	Land uses, traffic volume, population, distance to the ocean, elevation	50-5,000m
Franklin et al. [17]	Southern California, USA	NO, NO_2	Distance to roads, traffic volume, population, elevation, land uses	300m
Wu et al. [18]	Beijing, China	$PM_{2.5}$	Road length, land uses, population, bus stops, intersections	100-3,000m

the prediction problem of location-dependent time-series data from a data-driven perspective. These methods are generally not explainable since they often directly adopt raw data features, such as weather, road networks, and points of interest, and expect the model to select and use the best features for the prediction task automatically without generating interpretable results (e.g., [13, 24, 25]).

There are some existing feature selection techniques, such as Lasso [26] and Group Lasso [27], that can provide model interpretation for deep learning models. Qi et al. [8] propose an air quality prediction model that effectively removes the redundant or irrelevant features by introducing an extra sparse layer to minimize the Kullback–Leibler (KL) divergence between the weights and a vector of tiny values. However, such regularization-type of feature selection methods only examine the importance of individual features. Interpreting and quantifying the interacting process between features, such as the joint effect of wind speed and factory emission on $PM_{2.5}$ concentrations, is still challenging.

To learn the interactions between the environmental characteristics and their impacts on prediction values in an interpretable way, one promising direction to provide model interpretability is using variational autoencoders (VAE) (e.g., the beta VAE [28]) to generate factorized latent patterns in a lower dimension. One advantage of variational autoencoders is that it learns the distribution of the latent representations with a vector of means, μ, and a vector of standard deviations, σ, instead of directly outputting a condensed vector in traditional autoencoders. Such distribution in the latent space can help understand the compression process and relate the latent features to the prediction results.

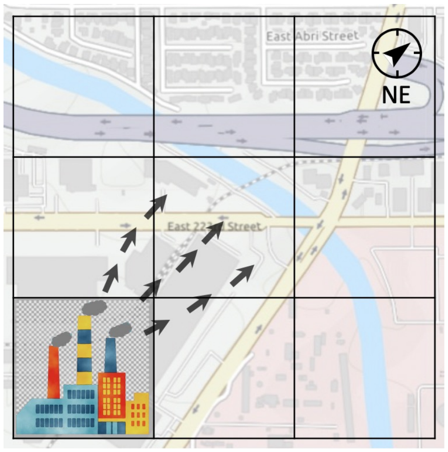
B. Modeling Spatiotemporal Dependencies

Spatiotemporal phenomena have strong spatial and temporal dependencies that should be considered jointly for accurate prediction. For example, current air quality is highly correlated with current and past surrounding environmental characteristics and is also influenced by the conditions from neighboring locations. Traffic volumes are related to the road features in the current cell and also affected by the information from neighboring cells such as the upstreaming traffic flow. We define these spatiotemporal impacts as the joint effects of features referenced in both space and time. For example, if a South-West power plant emits pollution at time T (Figure 3(a)), the North-East cells can be significantly polluted at $T+1$ with a North-East wind direction.

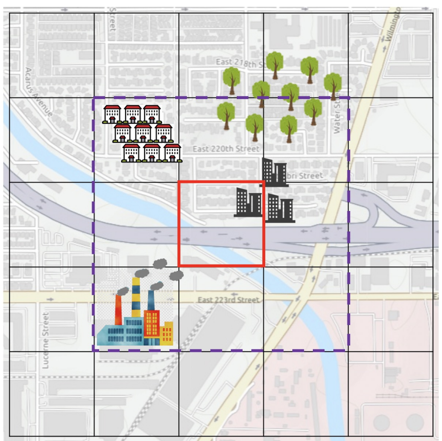
Existing prediction methods often overlook complex spatial and temporal variations in the interactions between the observations and external contextual data about the environment. For example, ordinary Kriging and LUR build prediction models for individual time points without considering temporal dependencies, i.e., either linear patterns (e.g., Figure 4(a)) or non-linear patterns (e.g., 4(b)) in the time-series data. Besides, Zheng et al. [13] propose a method that trains separate classifiers for spatial and temporal features and augments predictions via a co-training mechanism. However, separate models cannot effectively capture complex spatiotemporal interactions, such as emission patterns that diffuse over space with a high wind speed in short time frames.

With enough training data, deep learning technologies provide useful tools for modeling both spatial and temporal dependencies in the input data and their relationships to the prediction values. For modeling spatial dependencies, computer vision models commonly use convolution operations in convolutional neural networks (CNN) to extract important salient information from neighboring pixels [29]. One advantage of using convolution operations for location-dependent data is that convolution retains the positional relationships between data cells and is less computationally expensive than graph-based approaches (e.g., graph convolutional neural networks [30, 31]). For example, the influence of the left cell on the center cell should be heavier than the influence of the right cell when the wind direction is to the left. For modeling the temporal dependency, recurrent neural network (RNN) performs better than traditional models by automatically computing the information passing to the next time step in the sequence and using the final stored memory to make decisions (e.g., [6]).

To jointly model the spatiotemporal dependencies in the location-dependent time-series data, Shi et al. [32] propose the Conv-LSTM that adds the convolution operation directly in the recurrent neural network. For example, in Figure 3(b), when predicting the air quality value for the center cell (red box) by considering one-step neighbors (in the purple dotted box), the model should learn the interactive effects from the North-East green areas, the North-West residential areas, the South-East industrial areas, and the East commercial areas. Using the

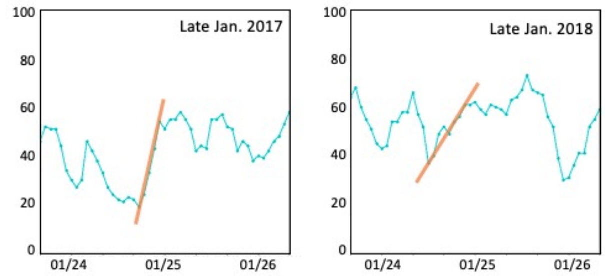


(a) A factory emission at time T

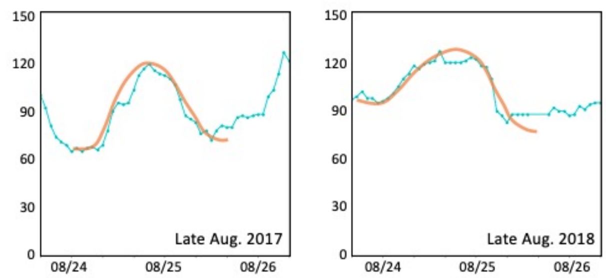


(b) Considering one-step neighbors

Fig. 3. An example of spatiotemporal effects. The cell size is 500m×500m. (a) A pollution emission releases from South-West power plant at time T, and its North-East cells can be polluted at time $T+1$ with the North-East wind, and (b) The red box is the target cell and the purple dotted box contains environment characteristics of the one-step neighbors.



(a) Linear patterns



(b) Non-linear patterns

Fig. 4. Examples of temporal variations of $PM_{2.5}$ concentrations. (a) linear patterns in late January 2017 and 2018, and (b) non-linear patterns in late August 2017 and 2018.

Conv-LSTM operation, the model learns useful information from the combination of the current latent embedding (i.e., the interactive effects) of the selected features from neighbors and the previous hidden memory.

One challenge here is that the spatial extent is not well defined and cannot be learned in a plain Conv-LSTM operation. Models using multiple Conv-LSTM layers with various convolution kernel sizes can learn the impacts from the neighbors within various distances. For example, if the kernel size is three and the cell size is 500m×500m, the model looks at one-step neighbors within approximately 500m. Similarly, for a kernel size of five, the model learns the information from two-step neighbors within approximately 1,000m. The outputs of these multiple Conv-LSTM layers can then be concatenated and fed to the fully connected layers to generate final predictions. However, how to enable the network to learn and focus on an appropriate kernel size or multiple kernel sizes is still an open research problem for spatiotemporal predictions. Computer vision approaches, such as the inception module [33], could help address this problem. The idea of the inception module is that instead of using a convolutional filter of fixed size, the network can learn to select the most effective kernels from filters of varying sizes and combine the image features learned from these kernels.

Another challenge is that the prediction model needs to take into account the fact that some cells should not contain prediction results (e.g., traffic volumes only occur in the cells with road features). Such data constraint (e.g., the traffic-road dependency) can be learned from the "No Data" (or null) cells in the grid and their nearby cells with some attention-based convolution layers. In this way, the model will be able to handle spatiotemporal phenomena that only occur in specific areas of the grid (e.g., traffic occurs only on roads).

C. Handling Sparse Observations

In real world prediction problems, the input data usually contain abundant explanatory variables (i.e., large feature vectors of contextual data) but limited observations. As an example, for air quality monitoring, in the Greater Los Angeles area, there are more than 13 million people within an area of approximately 30,000 square miles but only around 25 regulatory air quality monitoring stations (measuring multiple air pollutants) established by the U.S. Environmental Protection Agency (EPA). The sparse measurement locations largely limit the diversity of learnable environmental effects. More recently, low-cost sensors, such as PurpleAir, provide a more widespread network that captures finer spatial and temporal variability than Federal- and State-operated monitoring networks [34]. However, there remain limitations in terms of proper sensor installation, calibration, and maintenance which can leave some areas without reliable data. For example, Figure 5(a) shows the PurpleAir sensor locations covering an area of 2,000 square meters in Los Angeles, where sensors in the western coastal region are clustered together while sensors in the West-North region are scattered. As a result, the spatial coverage cannot satisfy the needs of the studies that require close tracking of exposure-response relationships (e.g., [3]).

Another example is the fixed traffic sensors that usually provide minute-level information about real-time and historical traffic volumes and flows in urban areas. These sensor measurements are critical inputs to air and noise pollution models [35, 36, 37]. While temporally resolved, traffic sensors are generally only on freeways and major arterials, with scant coverage of most primary, secondary, and residential roads. Using sensor-based traffic data in environmental exposure estimation has been an ongoing challenge due to these limitations. Figure 5(b) shows the locations of more than 9,000 traffic sensors covering highways and major roads in Los Angeles, many minor roads in populous areas such as Glendale, Burbank, and Huntington Park, but overall poor sensor coverage in Long Beach [1].

Semi-supervised learning methods can address the challenge of sparse sensors using deep neural networks by utilizing labeled samples (locations with an observation) together with the information from unlabeled samples (locations without an observation) to improve the model performance [7]. There are many types of semi-supervised learning methods, e.g., smoothness regularization [8], co-training [13], and the generative mixture models with EM algorithm (see more details in [38]).

Smoothness regularization assumes that samples close in space and time have a high chance to share similar values. Thus, the loss function of the deep neural networks usually consists of the prediction accuracy calculated using the labeled data and the smoothness of predictions on neighboring samples:

$$L = \sum_i \bar{L}(p_i, q_i) + \lambda \sum_i \sum_{j \in N_i} W_{ij} \bar{L}(p_i, p_j) \qquad (1)$$

where p_i and q_i are the predicted value and true label for sample i. N_i contains the neighbors of the sample i over space and time, and W_{ij} defines the spatial proximity or temporal similarity between sample i and j. $\bar{L}$ is the risk function, e.g., mean square error (MSE). Here, λ is a hyper-parameter. During the training process, the model minimizes the loss function (Equation 1) to avoid focusing on only labeled data. The smoothness regularization is simple and intuitive but might not be adaptable when the assumption on neighboring smoothness is invalid. Another challenge here is to adjust the hyper-parameters that 1) specify how close in space and time the prediction values should be similar and 2) incorporate the semi-supervised loss function with other loss functions in the entire network. For the first challenge, empirical methods typically work the best (e.g., setting fixed weights in space and time using k-fold cross-validation). For the second challenge, dynamic weight adjustments (e.g., [39, 40]) might overcome this problem by taking the sub-model uncertainty into account and improve model generalization for heterogeneous prediction domains.

Co-training is another semi-supervised learning approach that continuously augments the training set by selecting samples with the highest confidence from the prediction results. Zheng et al. [13] leverage the co-training technique for air quality prediction task by iteratively adding the most confident predictions from a spatial classifier and a temporal classifier into the training set. However, the disadvantage is that the method relies heavily on accurate model predictions from previous steps. Adding a highly confident but inaccurate prediction into the training set might propagate the errors, which further results in poor performance. Additionally, the computational cost is much higher than smoothness regularization since the model needs to be re-trained with "new" training samples at each iteration.

III. Summary

The location of things in space and how they change over time is the key to understand complex environmental phenomena as well as human-environmental interactions in the past, present, and future (e.g., [41]). This paper presented a vision of a general machine learning framework for explainable, fine-scale prediction of environmental phenomena, represented in location-dependent time-series data. The resulting framework will enable fine spatial and temporal scale environmental exposure assessment and allow researchers to carry out unprecedented inquiries, such as understanding relationships between health outcomes and long-term exposure histories.

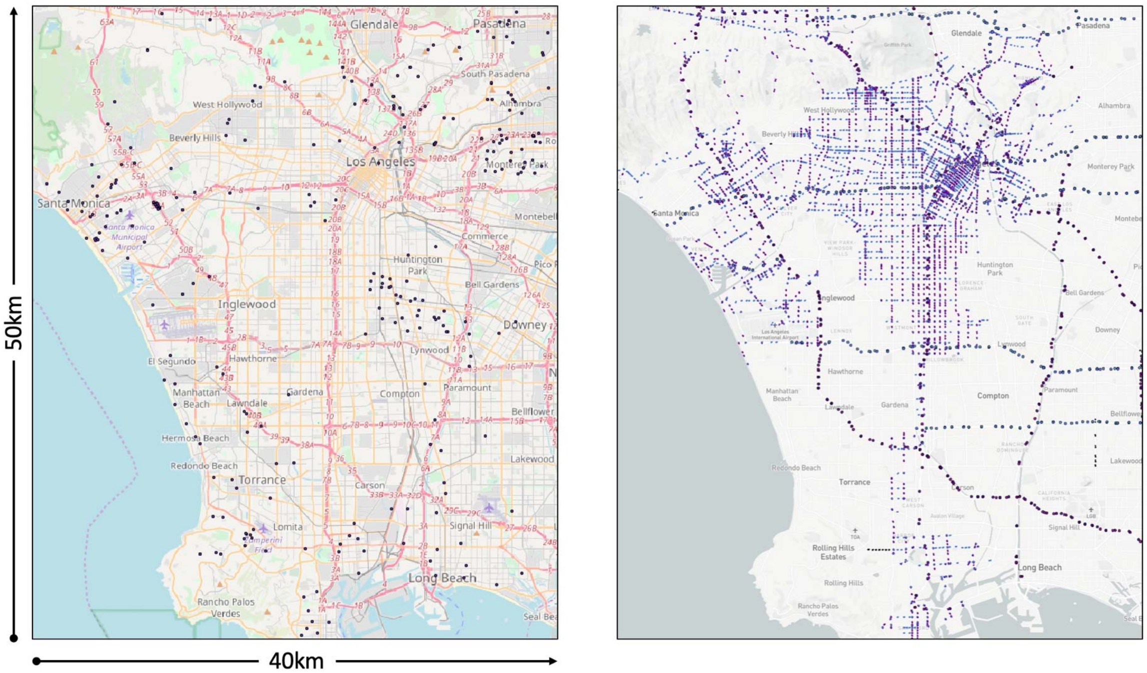
References

[1] C. Anastasiou, J. Lin, C. He, Y.-Y. Chiang, and C. Shahabi, "Admsv2: A modern architecture for transportation data management and analysis," in *ACM SIGSPATIAL International Workshop on Advances on Resilient and Intelligent Cities*, 2019.

[2] K. Nguyen, J. Yang, Y. Lin, J. Lin, Y.-Y. Chiang, and C. Shahabi, "Los angeles metro bus data analysis using gps trajectory and schedule data (demo paper)," in *ACM SIGSPATIAL*, 2018, pp. 560–563.

[3] C. S. Murray, G. Poletti, T. Kebadze, J. Morris, A. Woodcock, S. L. Johnston, and A. Custovic, "Study of modifiable risk factors for asthma exacerbations: virus infection and allergen exposure increase the risk of asthma hospital admissions in children," *Thorax*, vol. 61, no. 5, pp. 376–382, 2006.

[4] D. J. Briggs, S. Collins, P. Elliott, P. Fischer, S. Kingham, E. Lebret, K. Pryl, H. Van Reeuwijk, K. Smallbone, and A. Van Der Veen, "Mapping urban air pollution using



(a) Low-cost air quality sensors

(b) Traffic sensors

Fig. 5. Sparse and uneven distribution of sensors in the city of Los Angeles

gis: a regression-based approach," *International Journal of Geographical Information Science*, vol. 11, no. 7, pp. 699–718, 1997.

[5] Y. Lin, Y.-Y. Chiang, F. Pan, D. Stripelis, J. L. Ambite, S. P. Eckel, and R. Habre, "Mining public datasets for modeling intra-city pm2.5 concentrations at a fine spatial resolution," in *ACM SIGSPATIAL*, 2017, p. 25.

[6] Y. Lin, N. Mago, Y. Gao, Y. Li, Y.-Y. Chiang, C. Shahabi, and J. L. Ambite, "Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning," in *ACM SIGSPATIAL*, 2018, pp. 359–368.

[7] Z. Jiang, "A survey on spatial prediction methods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 9, pp. 1645–1664, 2019.

[8] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2285–2297, 2018.

[9] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian, "The impact of pm2.5 on the human respiratory system," *Journal of Thoracic Disease*, vol. 8, no. 1, p. E69, 2016.

[10] M. Guarnieri and J. R. Balmes, "Outdoor air pollution and asthma," *The Lancet*, vol. 383, no. 9928, pp. 1581–1592, 2014.

[11] R. D. Brook, S. Rajagopalan, C. A. Pope, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, A. Peters, D. Siscovick, S. C. Smith, L. Whitsel, and J. D. Kaufman, "Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association," *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.

[12] K. Li, R. Habre, H. Deng, R. Urman, J. Morrison, F. D. Gilliland, J. L. Ambite, D. Stripelis, Y.-Y. Chiang, Y. Lin *et al.*, "Applying multivariate segmentation methods to human activity recognition from wearable sensors' data," *JMIR mHealth and uHealth*, vol. 7, no. 2, p. e11201, 2019.

[13] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *ACM SIGKDD*, 2013, pp. 1436–1444.

[14] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[15] H. Bayraktar and F. S. Turalioglu, "A kriging-based approach for locating a sampling site—in the assessment of air quality," *Stochastic Environmental Research and Risk Assessment*, vol. 19, no. 4, pp. 301–305, 2005.

[16] D. Moore, M. Jerrett, W. Mack, and N. Künzli, "A land use regression model for predicting ambient fine particulate matter across los angeles, ca," *Journal of Environmental Monitoring*, vol. 9, no. 3, pp. 246–252, 2007.

[17] M. Franklin, H. Vora, E. Avol, R. McConnell, F. Lurmann, F. Liu, B. Penfold, K. Berhane, F. Gilliland, and W. J. Gauderman, "Predictors of intra-community variation in air quality." *Journal of Exposure Science & Environmental Epidemiology*, vol. 22, no. 2, pp. 135–47, 2012.

[18] J. Wu, J. Li, J. Peng, W. Li, G. Xu, and C. Dong, "Applying land use regression model to estimate spatial variation of pm 2.5 in beijing, china," *Environmental Science and Pollution Research*, vol. 22, no. 9, pp. 7045–7061, 2015.

[19] O. Hänninen, A. B. Knol, M. Jantunen, T.-A. Lim, A. Conrad, M. Rappolder, P. Carrer, A.-C. Fanetti, R. Kim, J. Buekers, R. Torfs, I. Iavarone, T. Classen, C. Hornberg, and O. C. L. Mekel, "Environmental Burden of Disease in Europe: Assessing Nine Risk Factors in Six Countries." *Environmental Health Perspectives*, vol. 122, no. 5, pp. 439–446, 2014.

[20] S. Stansfeld, "Noise Effects on Health in the Context of Air Pollution Exposure," *International Journal of Environmental Research and Public Health*, vol. 12, no. 10, pp. 12 735–12 760, 2015.

[21] M. Franklin and S. Fruin, "The role of traffic noise on the association between air pollution and children's lung function," *Environmental Research*, vol. 157, 2017.

[22] M. Franklin, X. Yin, R. McConnell, and S. Fruin, "Childhood psychosocial stress and the role of the built environment," *Journal of the American Medical Association, Pediatrics*, vol. In Review, 2019.

[23] M. Foraster, I. C. Eze, D. Vienneau, E. Schaffner, A. Jeong, H. Héritier, F. Rudzik, L. Thiesse, R. Pieren, M. Brink, C. Cajochen, J.-M. Wunderli, M. Röösli, and N. Probst-Hensch, "Long-term exposure to transportation noise and its association with adiposity markers and development of obesity," *Environment International*, vol. 121, no. June, pp. 879–889, 2018.

[24] H.-P. Hsieh, S.-D. Lin, and Y. Zheng, "Inferring air quality for station location recommendation based on urban big data," in *ACM SIGKDD*, 2015, pp. 437–446.

[25] W. Cheng, Y. Shen, Y. Zhu, and L. Huang, "A neural attention model for urban air quality inference: Learning the weights of monitoring stations," in *AAAI*, 2018.

[26] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[27] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[28] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," *ICLR*, vol. 2, no. 5, p. 6, 2017.

[29] M. Xin and Y. Wang, "Research on image classification model based on deep convolution neural network," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, p. 40, 2019.

[30] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in neural information processing systems*, 2016, pp. 3844–3852.

[31] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *ICLR*, 2018.

[32] X. SHI, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. WOO, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems 28*, 2015, pp. 802–810.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.

[34] A. Cavaliere, F. Carotenuto, F. Di Gennaro, B. Gioli, G. Gualtieri, F. Martelli, A. Matese, P. Toscano, C. Vagnoli, and A. Zaldei, "Development of low-cost air quality stations for next generation monitoring networks: Calibration and validation of $\mathrm{pm}_{2.5}$ and pm_{10} sensors," *Sensors*, vol. 18, no. 9, p. 2843, 2018.

[35] Y. Zhu, T. Kuhn, P. Mayo, and W. C. Hinds, "Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway," *Environmental science & technology*, vol. 40, no. 8, pp. 2531–2536, 2006.

[36] M. Fallah-Shorshani, L. Minet, R. Liu, C. Plante, S. Goudreau, T. Oiamo, A. Smargiassi, S. Weichenthal, and M. Hatzopoulou, "Capturing the spatial variability of noise levels based on a short-term monitoring campaign and comparing noise surfaces against personal exposures collected through a panel study," *Environmental research*, vol. 167, pp. 662–672, 2018.

[37] J. Tomić, N. Bogojević, M. Pljakić, and D. Šumarac-Pavlović, "Assessment of traffic noise levels in urban areas using different soft computing techniques," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. EL340–EL345, 2016.

[38] X. J. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2005.

[39] M. Guo, A. Haque, D.-A. Huang, S. Yeung, and L. Fei-Fei, "Dynamic task prioritization for multitask learning," in *ECCV*, 2018, pp. 270–287.

[40] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *CVPR*, 2018, pp. 7482–7491.

[41] T. VoPham, J. E. Hart, F. Laden, and Y.-Y. Chiang, "Emerging trends in geospatial artificial intelligence (geoai): potential applications for environmental epidemiology," *Environmental Health*, vol. 17, no. 1, p. 40, 2018.