Physics-Informed Deep Learning with Kalman Filter Mixture: A New State Prediction Model

Niharika Deshpande
Engineering Management & Systems Engineering
Old Dominion University
Norfolk, USA
ndesh002@odu.edu

Hyoshin Park
Engineering Management & Systems Engineering
Old Dominion University
Norfolk, USA
h1park@odu.edu

2024 58th Annual Conference on Information Sciences and Systems (CISS) | 979-8-3503-6929-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CISS59072.2024.10480207

***Abstract*—Harnessing a range of modeling approaches like Machine Learning (ML), Deep Learning (DL) etc. for analyzing spatiotemporal traffic data facilitates precise forecasts, optimizing transportation planning and congestion management for improved efficiency. Numerous studies in prediction modeling diligently incorporate spatiotemporal correlations into their analyses but fail to account for epistemic uncertainty which arises from incomplete knowledge across different spatiotemporal scales. This study aims to address this issue by capturing unobserved heterogeneity in travel time by considering distinct peaks in the probability density function, which we refer to as multimodal probability distribution while establishing causation through physics-based principles. The information obtained from this methodology is then employed in a new model called the Physics Informed-Graph Convolutional Gated Recurrent Neural Network (PI-GRNN). This DL model utilizes the inherent structure and relationships within the transportation network for capturing sequential patterns and dependencies in the data over time. The dynamic graph-based approach can utilize data from different locations and times to improve future travel time predictions at distant non-contiguous unobserved locations. We employ the PI-GRNN as the state-space model in the novel KF to obtain the evolution of the state with time. This approach will help in mitigating model drift caused by the data-driven approach by periodically correcting the PI-GRNN predictions with Kalman filter updates. To the best of our knowledge, this represents the pioneering data-driven multimodal multivariate learning approach to construct a dynamic graph of a traffic network. Furthermore, no other study has used the physics-informed data-driven technique as opposed to the mathematical model for the prediction step within the KF framework. Extensive experiments on real-world traffic data demonstrate that our model consistently outperforms the benchmark models.**

***Index Terms*—Kalman Filter, Physics-Informed, Graph Neural Network, Uncertainty Reduction**

I. Introduction

Advances in data intelligence and urban computing enable the extensive collection of traffic data, serving as vital indicators for assessing the state of the transportation system. This abundance of data plays a pivotal role in predicting future traffic conditions. Given the dynamic and unpredictable nature of road traffic, influenced by factors such as road closures, accidents, adverse weather conditions, it is imperative that any predictive model possesses the capability to account for these variables. This adaptability ensures the accuracy and reliability of traffic predictions, addressing the complexity of real-world scenarios.

Traditional choice theory assumes bounded rationality among travelers, leading to detour choices and increased congestion (1). Long-horizon optimization simplifies traffic states to unimodal distributions, neglecting the complexity of real-world dynamics. Gaussian processes struggle to integrate intricate prior knowledge (2). Recent advances like mixture density networks (3), handle multimodal output distributions, enhancing predictive accuracy. However, sequential learning still relies on unimodal distributions. This study proposes a novel information theory for actively sensing and learning sequential information, offering a more comprehensive understanding of evolving traffic conditions.

(4) developed a data-driven model for forecasting distant non-contiguous locations' conditions alleviating uncertainty and inter-traveler information transfer but fails to address coincidental correlations introduced due to pure data-driven approach. (5) addressed this issue using physics-regularized approach. They developed a Kalman Filter (KF) model by incorporating physics-regularized multimodal, multivariate correlations. But KFs use is limited to linear systems with white noise. Since traffic state is highly non-linear in nature, KF developed by (5) cannot account for it. To overcome this limitation of KF, different variations like EKF, UKF (6) and CKF (7) can be used but each has disadvantages associated with them. EKF relies on linearization, which means it assumes that the system dynamics and measurement models can be approximated as linear within the vicinity of the current state estimate. This assumption breaks down for highly nonlinear systems, leading to errors. While the UKF is more robust to non-linearities compared to the EKF, it may still struggle with highly nonlinear systems, and the choice of sigma point distribution can impact its performance. Like other nonlinear filters, the CKF can be sensitive to model errors and deviations from the assumed system and measurement models. If the models are significantly inaccurate, the filter's performance may degrade. To overcome the shortcomings of these non-linear filters, hybrid models that incorporate deep learning techniques can significantly improve their performance.

The studies developed algorithms that use Kalman filters and neural networks in combination, either to train the state-

NSF 1910397, 2106989 and NCDOT TCE2020-01

space model's equations or parameters with a neural network (8) or to update the neural network's parameters using a Kalman filter (9). (10) replaces the Kalman Gain (KG) component in KF with RNN but it may struggle with generalizing to unseen or different situations, making it less robust compared to traditional KF's KG computation, which relies on predefined mathematical models. To address this shortcoming, in our study, we replace the state-space model with a novel physics-informed deep neural network technique. Since the model has physics-informed component, it positively captures domain knowledge and non-linearity in dynamic systems. This modification in KF eliminates the need for accurate knowledge and modeling of the underlying dynamics. We develop a novel KF by incorporating these improvements which outperformed the benchmarks.

In DL, data correlations are effectively captured using techniques like RNNs (11), CNNs, and attention mechanisms. RNNs excel in sequential data addressing vanishing gradient issues through specialized versions like GRU and LSTM (12). For spatial dependencies, CNNs, GNNs (11) and attention mechanisms are employed. GNNs suited for graph-structured data show promise in forecasting applications but often use fixed graphs lacking adaptability to dynamic changes. This study introduces a physics-informed multimodal, multivariate approach for dynamic graph creation integrating spatiotemporal correlations into GNN and GRU. This innovative mixture algorithm ensures model training aligns with physical principles, offering a novel perspective on deep learning in traffic prediction.

The main contributions of our paper are summarized as follows:

- Our key contribution is the integration of a novel DL algorithm with a KF mixture, minimizing information uncertainty in traffic state estimation by leveraging a DL prediction algorithm within the KF framework for enhanced predictive accuracy.
- A novel DL model called Physics Informed-Graph Recurrent Neural Network (PI-GRNN) utilizes graph structures to encode meaningful representations of node attributes and their interactions. By capturing the temporal dynamics and evolution of graph data, PI-GRNN enhances information aggregation enabling more effective fusion and propagation across the graph through consideration of both current node states and historical contexts.

II. Kalman Filtering with Physics-Informed Deep Learning State-Space Model

The distinguishable aspect of the physics-informed and -regularized (PIR) model in the hierarchical update steps is the use of new information obtained from Temporal Multimodal Multivariate Learning (Figure 1). In this study, we employ a predictive model rooted in deep learning methodologies, as elaborated in the subsequent section, to anticipate the evolution of a selected state variable at the forthcoming time increment $t+1$. This chosen state variable represents a quantifiable parameter of interest, and our objective is to leverage the inherent

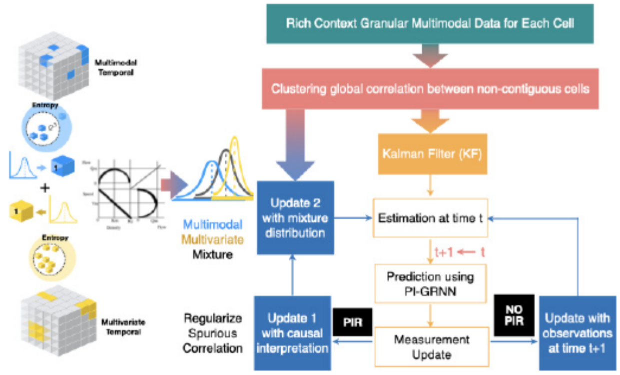


Fig. 1: Physics-informed and -regularized (PIR) KF in the hierarchical update steps with PI-GRNN as a state-space model

capabilities of deep learning techniques to generate accurate forecasts for this variable's values at future time points. In the update step, the predicted state is corrected using the noisy measurements at $t+1$. Clustering identifies similar travel time distributions. The global correlation between non-contiguous cells of an entire map is estimated by using Expectation Maximization. The optimal distribution of the data over K clusters is determined by maximizing the lower bound of the log of the likelihood. We decouple the spurious correlations first and then use the entropy method to estimate the mixture of multimodal and multivariate distributions. Since, the mixture is PDF with reduced entropy, providing an accurately estimated distribution rather than just mean and standard deviation, will increase the accuracy of updating the error covariance matrix.

A. Multimodal physics-informed deep learning as a state-space model

This section outlines the creation and refinement of the prediction model employed within the context of the KF framework. A traffic prediction problem can be formulated as a time-series forecasting problem with historical data and prior knowledge. The prior knowledge used in Graph Neural Network (GNN) is a pre-defined adjacency graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$. Here, $\mathcal{V}$ is a set of nodes that represent different locations (e.g., road segments) on the road network; $\mathcal{E}$ is a set of edges and $A \in \mathbb{R}^{N \times N}$ is the adjacency matrix.

Given the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ and its observed P step graph signals $\mathbf{X}_{(t-P):t}$, to learn a function f which can map $\mathbf{X}_{(t-P):t}$ and $\mathcal{G}$ to next Q step graph signals $\hat{\mathbf{X}}_{t:(t+Q)}$, represented as follows:

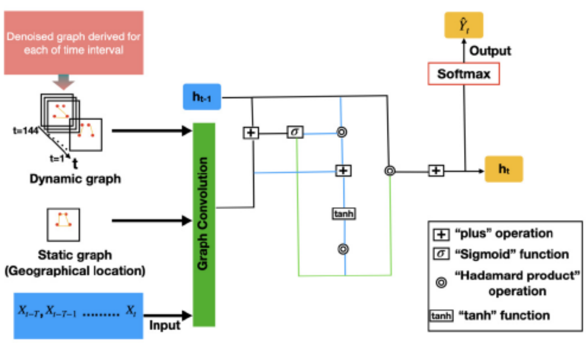
$$[\mathbf{X}_{(t-P):t}, \mathcal{G}] \xrightarrow{f} \hat{\mathbf{X}}_{t:(t+Q)},$$

where $\mathbf{X}_{(t-P):t} = (\mathbf{X}_{t-P}, \mathbf{X}_{t-P+1}, \ldots, \mathbf{X}_{t-1}) \in \mathbf{R}^{P \times N \times D}$, D is the number of features of each node (e.g., traffic volume, traffic speed, etc.) and $\hat{\mathbf{X}}_{t:(t+Q)} = \left(\hat{\mathbf{X}}_t, \hat{\mathbf{X}}_{t+1}, \ldots, \hat{\mathbf{X}}_{t+Q-1}\right) \in \mathbf{R}^{Q \times N \times D}$

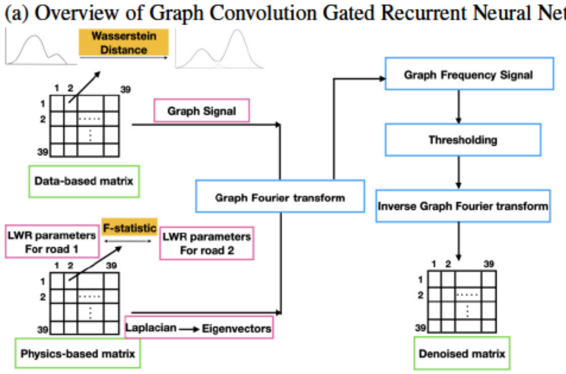
As shown in Figure 2a, Recurrent Neural Networks (RNN) are used in the case of sequential data as they retain the previous states in memory while accepting the current state. Therefore, it becomes a suitable means to solve time-series predictions. However, RNNs are capable of capturing only short-term temporal dependencies and have the issue of vanishing gradient. These limitations of RNN can be overcome by Long Short Term Memory (LSTM) (13) and Gated Recurrent Unit (GRU) (14). GRU has a less complex structure than LSTM as it has less number of gates, is easy to modify, and is faster to train. Therefore, we choose GRU for extracting temporal correlations from traffic time series data. We replace the matrix multiplications in GRU with the Graph Convolution (GC) module which is described using the following equations.

$$\begin{aligned} z_t &= \sigma(W_u.[GC(DA_t, A), h_{t-1}] + b_u) \\ r_t &= \sigma(W_r.[GC(DA_t, A), h_{t-1}] + b_r) \\ c_t &= tanh(W_c.[GC(DA_t, A), (r_t * h_{t-1})] + b_c) \\ h_t &= z_t * h_{t-1} + (1 - z_t) * c_t \end{aligned} \quad (1)$$

Where $\sigma(.)$ and $tanh(.)$ are the sigmoid functions, W and b are the weights and biases in the training, respectively. $*$ represents the matrix multiplication. DA_t denotes a dynamic adjacency graph at time interval t and A represents a pre-defined adjacency graph based on geographical locations.



(a) Overview of Graph Convolution Gated Recurrent Neural Network



(b) Flowchart of deriving dynamic adjacency matrix for each time interval

Fig. 2: The architecture of PI-GRNN

B. Weighted Dynamic adjacency matrix

The probability distribution function (PDF) is computed for individual road segments within a defined area of interest and for each designated time interval. This analysis is conducted by leveraging historical speed data, with the PDF derived through the utilization of a histogram. This methodology allows for a comprehensive understanding of the speed distribution across various road segments over time, providing valuable insights into the traffic dynamics within the specified geographical area. The obtained distribution is discerned to exhibit a multimodal nature indicating the presence of multiple peaks. This phenomenon is attributed to the diverse traffic patterns prevalent on the road.

To assess the semantic adjacency among road segments, it is imperative to evaluate the similarity of their PDFs. Various distance metrics, including KL divergence, Jensen Shannon entropy, Hellinger distance, and Wasserstein distance are considered. Notably, Wasserstein distance or Earth mover's distance (EMD), emerges as the most adept for assessing similarity in multimodal distributions. This metric captures differences in shape, location and spread between modes, making it particularly suitable for analyzing the multimodal distributions. This distance parameter preserves the distributional information of the data and hence can capture complex structure of multimodal distribution. It does not impose specific assumptions or constraints on the shape or type of the distributions being compared. This flexibility allows the Wasserstein distance to be used for comparing multimodal distributions that can exhibit various forms and structures. All these advantages make Wasserstein distance the most suitable parameter to measure the similarity between calculated speed distributions. A larger distance is assumed to indicate lower similarity, signifying reduced correlation between the road segments. The weighted adjacency matrix employing this assumption is calculated following given process. The Wasserstein distance is calculated using following equation.

$$W_1(P, Q) = \min_{\gamma \in \Gamma(P,Q)} \sum_{i,j} \gamma_{ij} \cdot d(x_i, y_j) \quad (2)$$

where, $W_1(P, Q)$ represents the Wasserstein distance between distributions P & Q, $W_1(P, Q) \in [0, \inf)$, Γ is a transportation plan that defines the amount of mass to be moved from each point in P to Q. It satisfies the constraints of being a valid joint distribution with marginals P & Q, denoted as $\gamma \in \Gamma(P, Q)$, x_i & y_j represent individual points (samples) from distributions P & Q, respectively, $d(x_i, y_j)$ is a distance metric (e.g., Euclidean distance or any other suitable distance measure) between x_i and y_j. Then, the distance matrix is normalized between values of 0 and 1 using the following formula. After that, the following equation is used to generate a weighted adjacency matrix. The adjacency matrix is a square matrix providing a succinct representation of semantic adjacency between pairs of road segments.

$$(x_{weighted})_i = 1 - z_i \quad (3)$$

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{4}$$

where, x_i is the value in square matrix to be normalized, and x_{min} & x_{max} is the minimum & maximum value in the matrix, respectively. This matrix is determined for each 10 minutes of time interval within 24 hours. These dynamic matrices helps to capture wide-range of spatial correlations. Figure 3a and Figure 3b shows the weighted adjacency matrix using multimodal data during time intervals 8-8:10 am and 8-8:10 pm, respectively. The spatial correlation patterns exhibit variability across different time intervals, as evident from the two figures.

As the weighted adjacency matrix as described above is entirely data-driven, it may exhibit spurious correlations due to the nature of its derivation. These coincidental correlations can negatively affect the prediction accuracy of the model, hence it is crucial to remove them. This research paper uses LWR (Lighthill-Whitham-Richards) which is a fundamental traffic flow theory to establish causal correlations and remove spurious ones. LWR is a macroscopic model that represents traffic flow based on the conservation of vehicles and the fundamental relationship between traffic density, flow rate, and speed. Mathematically, it can be expressed using the following equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{5}$$

where, ρ represents the traffic density (number of vehicles per unit length of the road), t is time, x is the spatial coordinate along the road and q is the traffic flow rate (number of vehicles per unit time). The LWR describes the evolution of traffic density and speed over time and space. If the parameters of the LWR are similar for two road segments, it signifies that the relationship between traffic density and speed is comparable for both of them. It suggests that drivers on these road segments experience similar congestion patterns, speed variations, flow characteristics and traffic flow capacities. Similar LWR parameters may indicate that congestion propagation between the two road segments is likely to be similar. Congestion in one segment may impact the traffic conditions in the other segment comparably. Hence, in this study, we compared the parameters for all segments using speed and density data collected using loop detectors over a year.

This study retroactively determines LWR traffic flow model parameters using a method of characteristics. The process involves solving differential equations backward in time from assumed initial conditions, utilizing collected speed and density data. Parameters include fundamental diagram parameters and traffic demand/supply parameters. Employing a least squares optimization method, the LWR equations are traced back in time, allowing for the determination of parameter values. Derived parameters for each road segment are compared using F-statistics and resulting F-values serve as weights in the adjacency matrix. These weights reflect the strength and significance of connections between segments, providing a quantitative measure of the relationship based on LWR parameter comparisons. Normalization is applied using equation 4 to ensure values in the adjacency matrix fall within [0, 1]. Figure 3c represents weighted adjacency matrix calculated using F-values between LWR parameters using multivariate data.

The spurious correlations are removed from data-driven matrix using the Graph Fourier Transform (GFT) method. GFT is computed using a physics-based adjacency graph, transforming the graph signal from the vertex domain to the graph frequency domain. The Laplacian matrix is constructed for data-driven adjacency matrix for each time interval. Eigenvalues and eigenvectors of the Laplacian matrix are computed, yielding the graph frequency signal. Thresholding is then applied to denoise the signal, with a threshold value of 0.01. The Inverse Graph Fourier Transform (IGFT) reconstructs the denoised graph signal in the vertex domain. This process is repeated for each of the 144 time intervals, resulting in denoised adjacency matrices used as input for the prediction algorithm.

III. Benchmark Analysis

The proposed model is developed using Pytorch 1.1.0 on a virtual workstation with an NVIDIA Quadro P2200 GPU. The model is trained using Adam optimizer. The learning rate is set to 0.001. The hidden state size is kept at 64. The batch size is set to 64 and the number of epochs is set to 100. To avoid overfitting, early stopping criteria are enforced. MAE is used as the loss function and if this metric doesn't improve for 5 number of epochs, the training is stopped. It takes 3 hours to train the model.

A. Evaluation Metrics of the Prediction

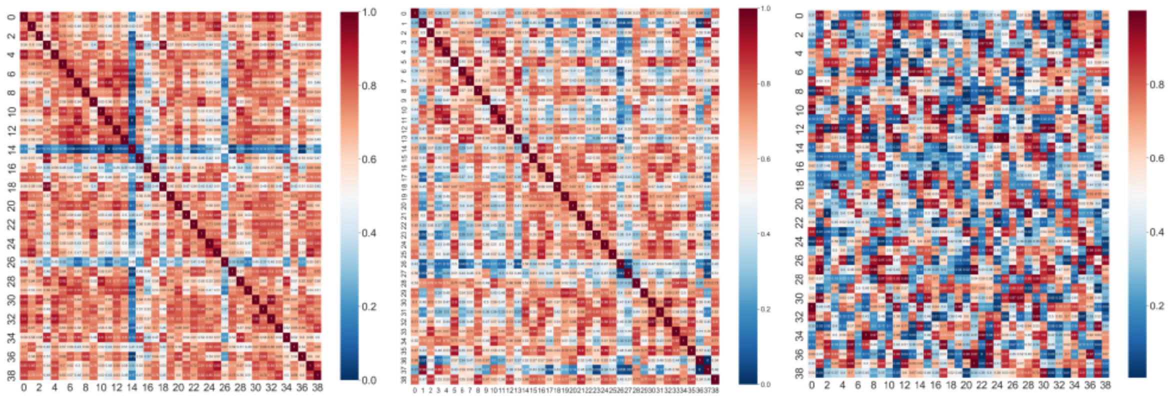
We evaluated the model performance based on three evaluation indicators, namely the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE). These metrics are defined as follows.

$$\begin{aligned} MAE &= \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \\ MAPE &= \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \\ RMSE &= \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} \end{aligned} \tag{6}$$

where n is the length of the time series, Y_t indicates the actual measurement, $\hat{Y}_t$ represents the predicted value from the model, and $\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$ denotes the forecast error. MAE reflects the absolute error of the prediction result. MAPE is a measure of the prediction accuracy of a forecasting method in statistics. RMSE can more accurately reflect the ability of the model to predict the values.

B. Benchmarks for PI-GRNN & KF models

The performance of the PI-GRNN model is compared with basic statistical models and with the latest hybrid GNN models using evaluation metrics. The prediction is determined for two



(a) Weighted adjacency matrix based on multimodal data at 8-8:10 am (data-driven)

(b) Weighted adjacency matrix based on multimodal data at 8-8:10 pm (data-driven)

(c) Weighted adjacency matrix based on a multivariate relationship across all times (physics-informed)

Fig. 3: Adjacency matrices from data-driven and physics-informed approaches

TABLE I: The evaluation metrics of the developed model and benchmarks

Model	30 minutes			1 hour		
	MSE	MASE	RMSE	MSE	MASE	RMSE
HA	4.20	7.85	13.05%	4.20	7.85	13.05%
ARIMA	5.18	10.5	12.75%	6.95	13.25	17.50%
DCRNN	3.20	6.50	8.85%	3.63	7.64	10.52%
AGCRN	3.25	6.70	9.03%	3.64	7.53	10.40%
DGCRN	2.99	6.05	8.02%	3.46	7.25	9.75%
PI-GRNN	**2.74**	**5.50**	**7.70%**	**3.38**	**7.19**	**9.68%**

horizons, 30 minutes (three time intervals) and 1 hour (six time intervals). The baseline models are as follows & Table I shows evaluation results. noitemsep

- **HA**: The Historical Average (HA) method predicts the future speed using an average of historical data.
- **ARIMA**: An autoregressive integrated moving average (ARIMA), is a statistical analysis model that predicts future values based on past values.
- **DCRNN** (15): Diffusion Convolutional Recurrent Neural Network is a fusion model of GCN with GRU for traffic data prediction.
- **AGCRN** (16): Adaptive Graph Convolutional Recurrent Network is a model that combines GCN with GRU employing an adaptive graph structure.
- **DGCRN** (17): Dynamic Graph Convolutional Recurrent Network model employs dynamic graph in GCN for spatial correlations and then uses the GRU model to gain temporal dependencies.

The performance of the KF-PIR & MIXTURE model is compared against the basic statistical model, traditional KF, KF-TML (4) data-driven model and hybrid KF-deep learning model KalmanNet (10). We used Mean Absolute Percentage Error (MAPE) as the measure of uncertainty. We assumed that lower value of MAPE implies lower uncertainty. The percentage uncertainty reduction is calculated against the ARIMA model using the following formula.

$$\% \text{ uncertainty reduction} = \frac{MAPE_{ARIMA} - MAPE_{after}}{MAPE_{ARIMA}} \tag{7}$$

where, $MAPE_{ARIMA}$ represents MAPE after applying the ARIMA model while $MAPE_{after}$ is MAPE following the application of the model under consideration to calculate the percent reduction in uncertainty.

The above formula is employed to calculate the percent reduction in uncertainty for each model. Figure 4 shows the significant percent reduction in uncertainty of predictions when the PIR + Mixture model is employed. Table II shows the percent uncertainty reduction of all the models. The results presented in the table show that our model performs significantly better than benchmarks.

TABLE II: Performance evaluation of the developed model and benchmarks (Part 2)

Model	Percent reduction in uncertainty
KF-PIR & Mixture	**19.3%**
KF-PIR	14.1%
KalmanNet (10)	13.5%
KF-TML (4)	5%
KF-traditional	2.1%

IV. Conclusion

This study addresses the limitation of suboptimal route suggestions by enhancing travel time predictions. Enhanced traffic predictions are attainable through data-driven models that capture unobserved heterogeneity by analyzing a mixture of multiple PDFs. However, the statistical transition of this knowledge across different times and spaces remains unexplored in prior studies. Additionally, the incorporation of physics knowledge serves to regularize potential spurious correlations within data-driven models.

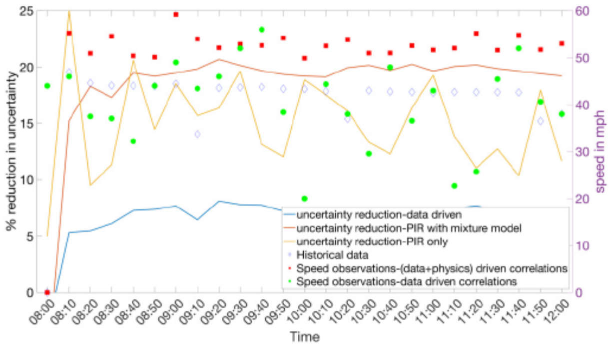


Fig. 4: Percent change in uncertainty for developed KF and benchmark models

This study develops advanced statistical deep learning model enriched with physics-informed regularization and cross-entropy-based mixture estimation that exhibit superior performance in minimizing travel time predictions compared to the author's earlier work (5). This paper introduces model that has ability to capture changing dynamics in system and utilize it to gain comprehensive spatiotemporal correlations. It unveils research prospects in physics-informed, information-theoretic statistical deep learning algorithms.

Acknowledgment

Funding for this research was provided by NSF [1910397, 2106989] and NCDOT [TCE2020-01].

References

[1] Q. Han and H. H. Timmermans, "Interactive Learning in Transportation Networks with Uncertainty, Bounded Rationality, and Strategic Choice Behavior: Quantal Response Model," Transportation Research Record, Jan. 2006, doi: 10.3141/1964-04.

[2] B. Alt, A. Šošić, and H. Koeppl, "Correlation priors for reinforcement learning," arXiv (Cornell University), vol. 32, pp. 14155–14165, Sep. 2019, [Online]. Available: http://arxiv.org/pdf/1909.05106.pdf

[3] F. Errica, D. Bacciu, and A. Micheli, "Graph mixture density networks," International Conference on Machine Learning, pp. 3025–3035, Jul. 2021, [Online]. Available: http://proceedings.mlr.press/v139/errica21a/errica21a.pdf

[4] H. Park et al., "Temporal multimodal multivariate learning," Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Aug. 2022, doi: 10.1145/3534678.3539159.

[5] N. Deshpande, H. Park, V. Pandey and G. Yoon, "Advancing Temporal Multimodal Learning with Physics Informed Regularization," 2023 57th Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 2023, pp. 1-5, doi: 10.1109/CISS56502.2023.10089632.

[6] S. Julier and J. Uhlmann, "New extension of the Kalman filter to nonlinear systems," Proceedings of SPIE, Jul. 1997, doi: 10.1117/12.280797.

[7] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," IEEE Transactions on Automatic Control, vol. 54, no. 6, pp. 1254–1269, Jun. 2009, doi: 10.1109/tac.2009.2019800.

[8] C. Chen, R. Xiong, R. Yang, W. Shen, and F. Sun, "State-of-charge estimation of lithium-ion battery using an improved neural network model and extended Kalman filter," Journal of Cleaner Production, vol. 234, pp. 1153–1164, Oct. 2019, doi: 10.1016/j.jclepro.2019.06.273.

[9] C. Guan, P. B. Luh, L. Michel, and Z. Chi, "Hybrid Kalman filters for very Short-Term load forecasting and prediction interval estimation," IEEE Transactions on Power Systems, vol. 28, no. 4, pp. 3806–3817, Nov. 2013, doi: 10.1109/tpwrs.2013.2264488.

[10] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. Van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," IEEE Transactions on Signal Processing, vol. 70, pp. 1532–1547, Jan. 2022, doi: 10.1109/tsp.2022.3158588.

[11] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18, Jul. 2018, doi: 10.24963/ijcai.2018/505.

[12] R. Fu, Z. Zuo, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," Youth Academic Annual Conference of Chinese Association of Automation (YAC), Nov. 2016, doi: 10.1109/yac.2016.7804912.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[14] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," arXiv (Cornell University), Sep. 2014, doi: 10.48550/arxiv.1409.1259.

[15] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion Convolutional recurrent neural network: Data-Driven Traffic Forecasting," International Conference on Learning Representations, Feb. 2018, [Online]. Available: https://openreview.net/pdf?id=SJiHXGWAZ

[16] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive Graph Convolutional Recurrent network for traffic forecasting," Open Publications of UTS Scholars (University of Technology Sydney), Jul. 2020, [Online]. Available: http://hdl.handle.net/10453/143566

[17] F. Li et al., "Dynamic Graph Convolutional Recurrent Network for Traffic Prediction: Benchmark and Solution," ACM Transactions on Knowledge Discovery From Data, vol. 17, no. 1, pp. 1–21, Feb. 2023, doi: 10.1145/3532611.