MetroBUX: A Topology-Based Visual Analytics for *B*us Operational *U*ncertainty E*X*ploration

Shishi Xiao, Qing Shi, Lingdan Shao, Bo Du, Yang Wang, Qiaomu Shen, *Member, IEEE*, and Wei Zeng, *Member, IEEE*

***Abstract*—In the public transportation system, punctuality benefits both bus operation and passengers' travel experience. However, uncertainty exists due to complex traffic conditions and heterogeneous driving behaviors. To analyze bus operational uncertainty, transport planners and bus operators need a tool that supports multi-granular modeling, spatio-temporal representation, and interactive exploration. To meet the requirement, we present *MetroBUX*, a visual analytics system for *B*us operational *U*ncertainty e*X*ploration. *MetroBUX* aligns daily bus trips and models stop-level uncertainty of bus arrival time. It has a consolidated interface with three main views: *Map View* for presenting the spatial distribution of uncertainty, *Temporal View* for tracking the evolution of uncertainty, and *Trip View* for inspecting uncertainty propagation. Specifically, *MetroBUX* enables integrated spatio-temporal analysis by connecting topological uncertainty distribution at different periods in a nested tracking graph. Furthermore, it supports interactive and hierarchical exploration, including region-, route-, trip-, and stop-level analysis. Case studies on real-world bus operational data and domain experts' feedback demonstrate the efficiency of *MetroBUX*.**

***Index Terms*—Bus operation, uncertainty modeling, uncertainty visualization, topology analysis, visual analytics.**

I. Introduction

PUBLIC transit plays an essential role in modern cities by providing commuters with an efficient and eco-friendly travel option. The bus system is a commonly used travel mode in multi-modal public transit systems [1], [2]. However, bus operation is volatile due to the dense bus network and mixed traffic environment with other vehicles. Ideally, the bus system should work as the pre-planned schedule to achieve maximum efficiency. But in realistic circumstances, bus operation always suffers from serious uncertainty which results in service inefficiency such as bus bunchings [3], [4]. Analyzing bus operational uncertainty can help domain experts to understand bus bunching and optimize the operation schedule. Nevertheless, complex traffic conditions pose a multitude of difficulties in the analytics of bus operational data. First, the phenomenon of bus bunching is observed to be stochastic in the real world. Ideally, bus operation shall follow the planned timetable or schedule, *i.e.*, punctuality. However, unforeseen circumstances like traffic accidents and passenger demand fluctuation widely exist, causing variations in bus arrival time and inregualarity of bus operations [5]. Second, it is evident that bus bunching happens in a variant of spatial and temporal dimensions [6], [7], [8]. The analytics tool needs to capture the spatio-temporal dynamics of bus operation, to help users identify when and where bus bunching may happen. Moreover, it is observed that heterogeneous driving behaviors cause travel time variations, even in the same route [9], [10]. As such, the analytics tool needs to support fine-grained analysis of bus operational data, allowing users to examine and compare different routes and stops in a public transit system.

Transportation researchers have made significant progress in improving bus operation through the use of comprehensive models [7], [8], [11] and deep-learning techniques [3], [12], [13]. Commercial products like INIT[1] and analytics platform like JP-DAP [14] offer dashboards to assist for the monitoring of bus operation. However, these solutions often lack the interactive analysis capabilities, which are crucial for meeting traffic operators' needs. Visual analytics that enables iterative, interactive and dynamic integration of human intelligence and data analysis, becomes a popular approach in intelligent transportation systems [15], [16], [17]. Many visual analytics systems have been developed for exploring passenger commuting data by public transportation (*e.g.*, [18], [19], [20], [21], [22]). The systems enhance the understanding of mobility or place connectedness via public transit, yet few are designed to assess bus operation variations systematically. On the other hand, some studies assess trip-level delay uncertainty, and suggest

Manuscript received 24 February 2023; revised 3 September 2023; accepted 27 November 2023. Date of publication 2 January 2024; date of current version 31 May 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62172398, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011700, and in part by the Shenzhen Science and Technology Plan Project (Shenzhen-Hong Kong-Macau Category C) under Grant SGDX20220530111001003. The Associate Editor for this article was S. S. Nedevschi. *(Shishi Xiao and Qing Shi contributed equally to this work.) (Corresponding author: Wei Zeng.)*

Shishi Xiao and Qing Shi are with Thrust of Computational Media and Arts, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China (e-mail: sxiao713@connect.hkust-gz.edu.cn; brantqshi@hkust-gz.edu.cn).

Lingdan Shao and Yang Wang are with the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Beijing 100045, China, and also with the University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: ld.shao@siat.ac.cn; yang.wang1@siat.ac.cn).

Bo Du is with the SMART Infrastructure Facility, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: bdu@uow.edu.au).

Qiaomu Shen is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: shenqm@sustech.edu.cn).

Wei Zeng is with the Thrust of Computational Media and Arts, and Thrust of Data Science and Analytics, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China, and also with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China (e-mail: weizeng@hkust-gz.edu.cn).

Digital Object Identifier 10.1109/TITS.2023.3338700

[1] https://www.initse.com/

visual designs to assist personal trip planning [23], [24]. However, these visual designs can not fulfill the requirements of domain experts in examining bus operational data, in a systematical manner considering spatial, temporal, and route-wise variations.

To overcome the above challenge, we present *MetroBUX*, a visual analytics system for exploring bus operational uncertainty. In the system, we model the stop-level uncertainty (Sec. IV) and derive the density contour through Kernel Density Estimation (KDE) (Sec. V-A). Then we employ the Reeb graph to model the topological structure and the Hilbert curve [25] to measure the spatial relationship between contours, thus the evolution of uncertainty distribution can be derived (Sec. V-B)). At last, several elaborately designed visualizations are integrated into a multi-view interface with flexible interactions to facilitate detail-oriented exploration of arrival time uncertainty (Sec. VI). We comprehensively evaluate the efficiency of *MetroBUX* in meeting the requirements of spatio-temporal and hierarchical (including region-, route-, trip-, and stop-level) analysis, with three case studies and expert feedback (Sec. VII).

In general, the contributions of our work are threefold:

- **Multi-granular uncertainty modeling.** We extract stop-level uncertainty from the fluctuation of arrival times in bus operational data, which are used to compute region-level uncertainty distribution over space, and derive nested uncertainty contours.
- **Integrated spatio-temporal visualization.** We leverage and improve the nested tracking graph for integrated spatio-temporal visual representation, and design a multi-view interface incorporating dedicated visual and interaction designs to support hierarchical analysis of uncertainty in the region-, route-, and stop-levels.
- **Insightful evaluations.** We conduct three case studies on real-world scenarios and reveal new insights including spatio-temporal variations of uncertainty, tidal phenomenon, and stop-level uncertainty reasoning. Experts' feedback also confirms *"user-friendly"* visual & interaction design as well as *"versatile"* applicability of *MetroBUX*.

II. Related Work

A. Bus Operation

Bus scheduling aims to achieve an optimal balance among various factors, including passenger demand [7], operating expenses [11], optimal route design [26], [27], and waiting time [8]. However, most approaches are susceptible to specified factors that only consider a subset of circumstances. To mitigate the uncertainty of bus arrival time and to improve public transit service, recent works leverage deep-learning techniques with historical bus operational data for bus arrival time prediction [3], [13]. However, these data-driven approaches also face challenges, including biased data collection, unpredictable traffic conditions, and heterogeneous driving behaviors.

Preliminary efforts have been made to investigate bus operational data with visual analytics. Palomo et al. [19] depicted train schedule by combining KDE with a graphical approach, allowing users to inspect and compare spatio-temporal patterns in transportation services. Weng et al. [20] developed BNVA, allowing users to inspect bus route networks and evaluate route candidates. Barabino et al. [5] introduced methods for the measurement of "inregularity" in transit services, which can be helpful for the identification of underlying causes. Zhao et al. [28] designed a knotted visualization method to present uncertainty information about public transit to support intelligent transit scheduling. However, these systems focus on one or several routes, failing to provide an overview of *when* and *where* bus operation is deficient. To mitigate the gap, we develop a hierarchical analytical framework assembled with region-, route-, trip- and stop-level analysis.

B. Urban Traffic Visualization

Understanding bus operation requires spatial and temporal analytics of bus operational data simultaneously, which has been comprehensively studied in urban traffic visualization [15], [16], [29], [30], [31], [32]. The widely used approaches include space-time cube framework (*e.g.*, [33], [34]), density map (*e.g.*, [22], [35]), linear cartogram (*e.g.*, [36], [37]), and space-time transformation (*e.g.*, [38], [39], [40]). All the approaches have certain limitations, such as increased navigation complexity in 3D for the space-time cube, map distortion for the cartogram, and reduced spatial information for the space-time transformation. The problem becomes more complex when the movements are constrained in the road network, which further requires depicting the underlying graph structure. A straightforward solution is to broaden the route in a map and embed the visualizations into the route [41]. However, this approach is cumbersome when the number of routes becomes large. Alternatively, novel visualizations that seamlessly integrate network information can be designed, such as bristle maps [42], and topology density map [22]. Coordinated views [43] are another strategy to support the visual exploration of traffic data, with different views presenting spatial and temporal patterns, respectively. For example, Zeng et al. [44] adopted space-time transformation for travel time visualization, complemented with an isochrone map for depicting spatial information. Advanced topological analysis (*e.g.*, [45], [46]) and interaction (*e.g.*, [18], [47]) methods can also be applied to facilitate visual exploration.

To the best of our knowledge, no visual analytics has been dedicated to uncertainty exploration in bus operation, which is essential for transportation planners and service providers in decision-making. To fill the gap, *MetroBUX* is designed to help domain experts explore bus operation uncertainty.

C. Uncertainty Visualization

Uncertainty exists inevitably in all stages of the analysis pipeline, including data acquisition, transformation, and visualization [48], [49]. Despite the diverse types and sources of uncertainty, uncertainty visualization can be generally categorized as *add glyph*, *add/modify geometry*, *animation*, *etc.* [48]. Some goal-directed visualization attempts to incorporate the underlying uncertainty for conveying more transparent data, thus decreasing the possibility of false decision-making [50]. The visualization community has designed uncertainty visualization tools for almost all application contexts, such as geodemographics [51], multidimensional projection [52], and hurricane prediction [53].

MacEachren et al. [54] summarized general strategies for uncertainty visualization in geospatial information: *static* representations such as iconography, visual variable coding, and their combinations; and *dynamic* representations including animated, sonic, and interactive ones. They outlined that understanding how knowledge of information uncertainty affects decision-making is a fundamental research challenge for uncertainty visualization. Gschwandtnei et al. [55] compared six different visual encodings of visualizing temporal uncertainty and reconfirmed that the best proper visual encoding depends on the task. Specifically, in the context of transportation, Kay et al. [24] investigated what type of uncertainty representation can produce reasonable judgment when taking the bus. They suggested using a dotplot to present probability distributions in a discrete way, whose variance is 1.15 times smaller than the density plot. Wunderlich et al. [23] proposed a visual design that depicts train trip planning with expected delay uncertainty and the potential impact of missing a transfer.

The exploration of bus operation uncertainty is complex as it involves multi-perspective analysis: *spatial* variation, *temporal* evolution, and *route* propagation. To address this challenge, we opt for KDE to provide an overview of uncertainty distribution over space, which has been widely applied in hotspot detection [45], [56] and POI analysis [22] that rely on the identification of high scalar values in a geographical space. On this basis, we further opt for nested tracking graph (NTG) [57], [58] to examine the evolution of uncertainty density at different levels over time, which has been successfully demonstrated in examining spatial data changes over time. Our interface also features a graphical schedule view for route-level comparison.

III. Requirement Analysis and System Overview

This section first introduces the requirements distilled from a collaboration with a transportation researcher (Sect. III-A), followed by an overview of the system (Sect. III-B).

A. Requirement Analysis

In the past ten months, we worked closely with a researcher (*CR*) with over 10 years of experience in public transport planning and operation. The *CR* mentioned that the latest deep learning approaches have improved prediction accuracy, but usually rely heavily on the training data and fail to capture dramatic variations. Understanding the variations can help domain experts analyze the failure of the prediction models in extreme cases and optimize the bus schedule.

In the first phase of the collaboration, we conducted several runs of semi-structured interviews with the *CR* to formulate the detailed design requirement. During the discussion, we modeled the variation of bus arrival time as *uncertainty* (Fig. 2 (*b*)) and aimed to develop a visual analysis system to explore the bus arrival time uncertainty from the bus operational data. The detailed analysis requirements are listed as follows:

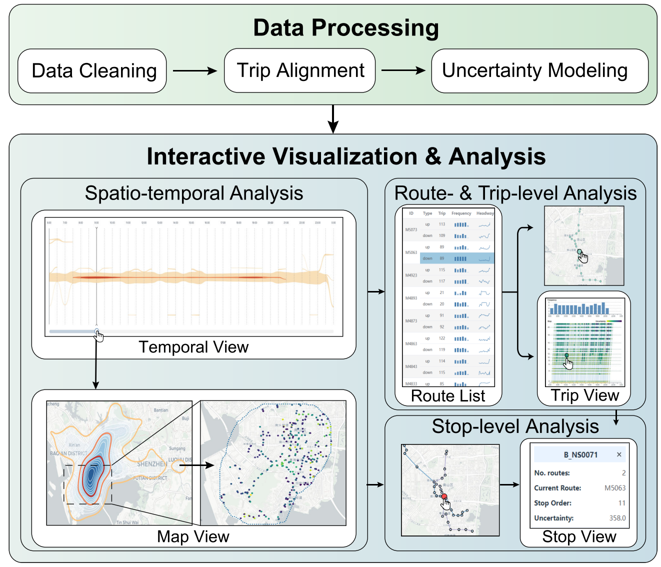


Fig. 1. Overview of the workflow, which mainly consists of *data processing* and *interactive visualization and analysis* stages.

- **R1**: **Multi-granular Uncertainty Modeling**. To facilitate the in-depth exploration of bus operational data, the system needs to provide uncertainty modeling at multiple granularities, including the *stop-* (*R1.1*) and *region*-level uncertainty (*R1.2*).
- **R2**: **Integrated Spatio-temporal Uncertainty Representation**. Traffic conditions and travel demands vary over space and time in a city. *CR* requests to get an intuitive visual representation of uncertainty distribution over space to locate regions of high uncertainty (*R2.1*), and evolution of uncertainty over time to identify periods of high uncertainty (*R2.2*).
- **R3**: **Adequate Auxiliary Information**. The system should provide rich auxiliary information to help users understand and reason the uncertainty patterns revealed from the visualizations. Specifically, the basic information (ID, name) and statistics of bus stops, routes (*R3.1*), and trips (*R3.2*) should be visualized to assist the analysis of spatial-temporal uncertainty.
- **R4**: **Effective User Interaction and Cross-view Linkage**. *CR* expects effective user interactions to support interactive exploration of bus operation uncertainty. The system should support the selection of region, route, trip, and stop of interest in different views (*R4.1*). Moreover, when using the system, the users always need to switch their attention among different views, thus the related visual elements in different views should be linked to make the interactive exploration easy (*R4.2*).

B. System Overview

We develop *MetroBUX*, a web-based visual analytics system which comprises two major phases: 1) *data processing*, and 2) *interactive visualization & analysis*, as illustrated in Fig. 1.

In the *data processing* phases, the raw data including route information and operational data are cleaned where duplicates are removed and missing values are filled. Then we perform the trip alignment and uncertainty modeling to get the stop-level uncertainty (*R1.1*) which will be passed to the next phase for in-depth analysis.

In the *interactive visualization & analysis* phase, the system provides several coordinated views to support flexible interactive exploration. All the routes are shown in *Route List* view which depicts *route ID, direction, number of daily trips, frequency histogram*, and *headway* information for each route (*R3.1*). Users can select a route of interest from *Route List* view and examine the details of the frequency histogram and headway information in the *Trip View* (*R3.2*). Based on the stop-level uncertainty, *MetroBUX* models region-level uncertainty distribution using kernel density estimation (*R1.2*), which is presented in the *Map View* (*R2.1*). We conduct a topological analysis on the derived contours at each time and connect the topology at different times in a nested tracking graph as in the *Temporal View* (*R2.2*). The *Temporal View* intuitively presents the uncertainty evolution over one day and allows users to examine detailed uncertainty distribution at a specific time range by dragging the time slider. Users can also zoom/pan the *Map View*, and the nested tracking graph will be updated accordingly (*R4.1*, *R4.2*). *Map View* and *Temporal View* work in an integrated manner to support spatial and temporal analysis simultaneously (*R4.2*).

IV. Data Processing

In this section, we first give a brief introduction of the input data (Sec. IV-A), followed by the description of data cleaning process (Sec. IV-B) and trip alignment and uncertainty modeling (Sec. IV-C.2).

A. Input Data

The data we used in our system include the **stop dataset**, **route dataset** and **operational dataset**, as follows:

Stop dataset includes the information of all bus stops. Each stop is represented as a triplet: $s :=< stop\ ID, lat, lon >$, where lat, lon specify the geographical coordinates of a given bus stop s, respectively. There are in total 1226 bus stops in the dataset.

Route dataset includes the information of bus routes. Each bus route is presented as $R := \{s_i^R\}_{i=1}^{|S_R|}$, where i indicates the index of bus stops along the route, while $|S_R|$ indicates the total number of bus stops in operation. s_1^R and $s_{|S_R|}^R$ are the starting and ending terminals, respectively. It's worth noting that several bus routes may pass the same stop, and bus operations are variant for different routes at the same stop. In this study, the route dataset consists of 92 bus routes.

Operational data include the daily bus operation data that record the times a bus arrives at and leaves a bus stop. Each record includes *route ID*, *stop ID*, *location*, *type* and *arrive time*, where *location* indicates geographical coordinates to collect this record; *type* indicate direction of the bus, and *arrive time* is the time arriving at the stop. Since a bus route is typically served by several buses, we can connect multiple

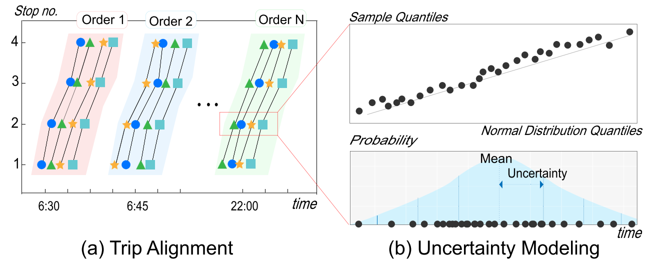


Fig. 2. Bus trips across multiple days are aligned via clustering (a), which are then are used for stop-level uncertainty modeling (b).

logs at different stops but with the same *bus ID* within a reasonable time period as a bus trip (Fig. 2 (*a*)). Note that bus drivers may choose not to stop at a bus stop if there is no waiting or alighting passenger, yielding missing values at certain stops. Hence, we fill in the missing values before the followup analyses. The static stop and route information and dynamic bus operational data are jointed using conjunct features of *route ID*, *stop ID*, and *type*.

B. Data Cleaning

Automatically collected data often exhibits quality issues that compromise its consistency, a challenge widely acknowledged in existing literature [59], [60]. These inconsistencies may arise from various factors, such as sensor instability, weak location signals, or errors committed by bus operators. To address these issues, we categorize data inconsistencies into three main types: duplicated records, abnormal data, and missing data. Initially, we eliminate duplicate records (*i.e.*, two close arrival times of the same bus ID at the same route and stop). Next, we filter out outliers such as the irrational departure times at the starting stop or one single recorded arrival time for a particular bus ID within one-hour window. After that, we arrange the remaining arrival times in order of their corresponding stops along each route. Arrival times linked by a common bus ID within a predefined time window are grouped and treated as a single bus trip. With a given trip, arrival times are missing when drivers do not stop by a bus stop. To model the uncertainty of all stops, we fill out the missing arrival times via linear interpolation.

C. Trip Alignment and Uncertainty Modeling

1) Trip Alignment: After connecting bus trips and completing the missing values, we further align bus trips spanning different days to model bus operation uncertainty. We use attribute *trip order* (denoted as tr) to indicate the order of a bus trip in one day, which is assigned according to the departure time of a bus trip at the starting terminal. Most bus trips in the operational data exhibit regular departure intervals, for which the trip orders can be assigned directly. However, certain bus trips have close departure times, which happens when the drivers start the engine earlier but do not leave the starting terminal. In such cases, we measure differences in arrival times between these uncertain trips and those certain trips on the other days at all stops along the route. The trip with the minimum difference is selected and assigned the trip order.

2) Uncertainty Modeling: With the trip alignment, we derive a list of bus arrival times at a stop s for a bus trip with trip order o spanning over one month, denoted as $\mathcal{T}_{s,o} = \{t^1_{s,o}, t^2_{s,o}, \cdots, t^m_{s,o}\}$, where m is the max index of day (*i.e.*, 30 in this work). An arrival time $t^i_{s,o}$ can be affected by factors like traffic and weather conditions, passenger demands, and driving behaviors. The arrival times at different days do not affect each other, hence $t^i_{s,o}$ and $t^j_{s,o}$ are independent variables. According to the central limit theorem, the probability of bus arrival times will closely approximate as a normal distribution. To validate this assumption, we randomly select 10 $\mathcal{T}_{s,o}$ and compare the values in $\mathcal{T}_{s,o}$ with theoretical values in normal distribution using a quantile-quantile plot. An example is shown in Fig. 2 (*b-top*). The linearity of the points in the plot suggests the proximity of theoretical and data distributions. Hence, the assumption that $\mathcal{T}_{s,o}$ follows the $\mathcal{N}(\overline{\mathcal{T}}_{s,o}, \sigma_{s,o})$, where $\overline{\mathcal{T}}_{s,o}$ is mean value and $\sigma_{s,o}$ is standard deviation of the arrival times.

Next, we model stop-level uncertainty based on historical bus operational data (*R1.1*). This work regards uncertainty as the extent of deviation from normal arrival time, a form of *statistical uncertainty* that calculates a numeric range with the likelihood of containing the target value determined by the model and the confidence associated with it [61]. We have high confidence that the arrival time will fall in the range $[\overline{\mathcal{T}}_{s,o} - \sigma_{s,o}, \overline{\mathcal{T}}_{s,o} + \sigma_{s,o}]$. Empirically, we find that the range will not cause overlapping issues between consecutive trips. As such, we set uncertainty of bus arrival times as $\sigma_{s,o}$, as illustrated in Fig. 2 (*b-bottom*).

However, trip orders between different bus routes are distinct due to varying departure frequencies. To compare arrival times among bus routes, we group arrival time uncertainty at each time interval ΔT (30 minutes in this work), and measure the average uncertainty. In this way, we generate a list of uncertainty values for each bus stop s over one day, denoted as $\left\{\sigma_s^{\Delta T_1}, \sigma_s^{\Delta T_2}, \cdots, \sigma_s^{\Delta T_k}\right\}$, where k is a fixed number for all bus routes. Our approach involves aggregating data for all bus routes at each stop, instead of focusing on individual route-stop pairs. This choice is rooted in the fact that spatial modeling of bus operation uncertainty is undertaken with a focus on uncertainty values rather than stop density, for which we adopt adaptive bandwidths to alleviate the effect by stop density. Computation per route-stop pair would further introduce the effect by route frequency, and cause potential increase in computational load during the KDE process (Sec. V-A.1). Another advantage of aggregating by stop is its capability to handle situations where specific bus routes have low-frequency service, especially at suburban stops during night-time hours. In this way, we can improve the identification of pronounced uncertainty patterns, particularly those related to punctuality issues, for suburban stops.

V. SPATIO-TEMPORAL UNCERTAINTY REPRESENTATION

After modeling stop-level uncertainty, we employ kernel density estimation for regional uncertainty modeling (*R1.2*) and visual representation (Sec. V-A). We further employ a nested tracking graph to represent the evolution of uncertainty over time (Sec. V-B).

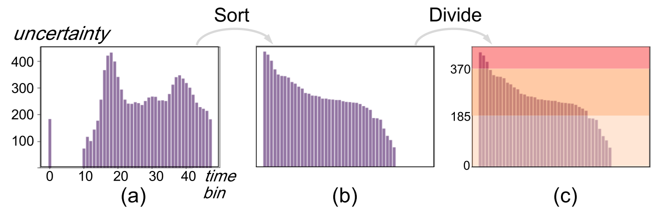


Fig. 3. Dividing uncertainty values into three intervals. (a) measure uncertainties for each time bin, (b) sort uncertainties in decreasing order, (c) divide them based on the biggest difference.

A. Spatial Modeling and Representation

1) Kernel Density Estimation: As a non-parametric approach, KDE is a compelling method to evaluate density probability at interactive rates. Given a set of virtual bus stops $\mathcal{S} = \{s_i\}_{i=1}^n$, where each stop s_i has a list of arrival time uncertainty $\left\{\sigma_s^{\Delta T_i}\right\}_{i=1}^k$. With a given time interval ΔT, we measure the regional uncertainty distribution at arbitrary spatial location x as:

$$\rho(\mathbf{x} \in \mathbb{R}^2 | t = \Delta T) = \sum_{s_i \in \mathcal{S}} \frac{\sigma_{s_i}^{\Delta T}}{h(s_i)} K\left(\frac{||pos(\mathbf{x}) - pos(s_i)||}{h(s_i)}\right), \tag{1}$$

where K denotes Gaussian kernel, $||\cdot||$ represents the Euclidean distance between two stops, and $h(s_i)$ is a function for adjusting bandwidth based on s_i. Conventional KDE uses a fixed bandwidth for all the sampling points, for which both the stop density and uncertainty values will contribute to the uncertainty density estimation. However, in this work, we are more interested in the uncertainty values than the stop density over space. As such, we adopt an adaptive bandwidth inspired by [62], as follows:

$$h(s_i) = h \times f(s_i), \tag{2}$$

where $h(s_i)$ is the bandwidth that corresponds to a spatial distance. $f(s_i) = |\{s \mid ||pos(s) - pos(s_i)|| < h, s \in \mathcal{S}\}|$ counts stops falling in a circle with s_i as a center and h as a radius, which increases as h increases. In this way, we can differentiate each stop's uncertainty density contribution and mitigate the artificial impact by stop density. Large h tends to over-smooth the density distribution, whilst small h yields many small fragmented regions. Fig. 4 shows a comparison of *Map View* and *Temporal View* generated by our adaptive bandwidth (top) and fixed bandwidth (bottom). Fixed bandwidth may generate some outliers, among which a notable one is the area at 12:00 marked in red (bottom-right), corresponding to Shenzhen North station, one of the busiest areas in Shenzhen. Due to the limited number of bus stops within this area, density map by KDE using fixed bandwidth exhibits small fragmented regions that are isolated from the other regions. This can lead to mistakenly-perceived low level of uncertainty in the red area, whilst the region exhibits significant uncertainty in

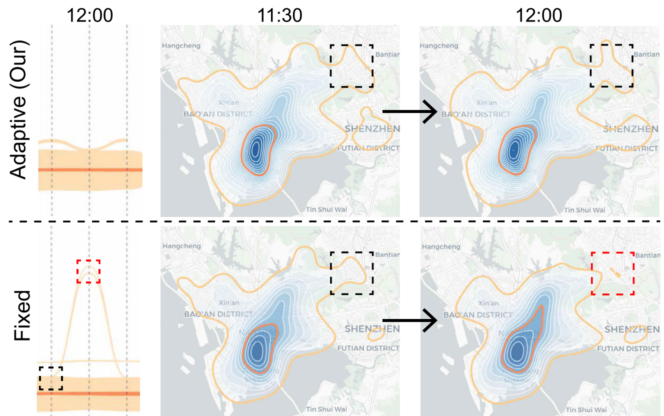
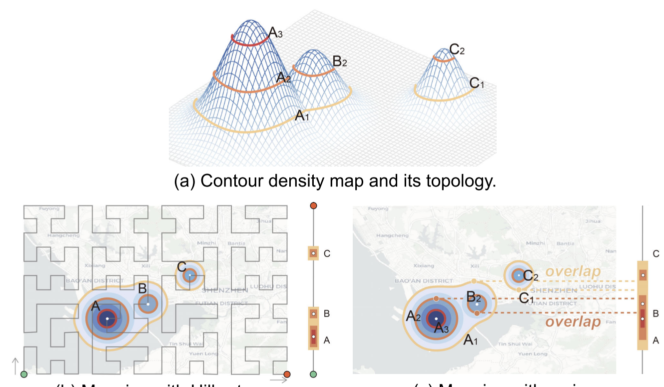


Fig. 4. Comparing *Map View* and *Temporal View* generated by adaptive (top) and fixed (bottom) bandwidth. Fixed bandwidth can cause many small fragmented regions as in the red area.



(a) Contour density map and its topology.

(b) Mapping with Hilbert curve.

(c) Mapping with y-aixs.

Fig. 5. Illustration of density map and the highlighted contours in 3D (*a*). All points on the 2D map can be mapped onto a 1D space using Hilbert curve (*b*) rather than positions along y-axis (*c*), with region sizes reflected as the length of the line segments.

reality. Our adaptive bandwidth can effectively address the problem, as shown in the top-right.

After consulting with the expert *CR*, we set a default bandwidth of 1000m which keeps a balance between too many fragmented and over-smoothing regions as a rule of thumb. Alternatively, users can examine and compare effects of different bandwidths using a bandwidth slider overlaid on the *Map View* (see Fig. 7 (*d*)). Moreover, conventional KDE computation with a fixed screen length as the bandwidth suffers from dramatic changes in density maps after zoom operations [63]. To alleviate the issue, we adjust the bandwidth by anchoring the bandwidth with a fixed distance when users zoom in/out on the map.

2) Contour Topology: Heatmap based on KDE is the most commonly used method to present the spatial distribution of bus operation uncertainty (R2.1) KDE computes uncertainty density distribution over space at a specific time interval. On this basis, one can visualize and compare uncertainty densities at different time intervals. However, using a fixed color mapping for all time intervals will introduce bias, as the minimum and maximum uncertainty values at different time intervals can vary significantly, particularly for peak *vs.* non-peak hours. Moreover, the density map adopts a continuously changing colormap with gradient blue, posing a threat of ambiguous interpretation of choropleth border. To enable consistent uncertainty comparison and remove the potential bias of the choropleth border, we further incorporate several highlighted isolines sampled from the blue contours in density maps. The sample strategy is based on the elbow method [64], which is a heuristic for finding the optimal threshold by locating the point with a sharp drop.

As illustrated in Fig. 3, we first divide a day into 48 time bins and measure average uncertainty in each time bin, then sort the uncertainty values in decreasing trend. The value with largest difference between the uncertainty values, indicating the severity of uncertainty, would be set as critical points. Following this criteria, we get three intervals: *low* interval is [0, 185]; *medium* level is [185, 370]; *high* level is [370, *max*], which are accordingly denoted as $\mathcal{L} := \{l_{low}, l_{medium}, l_{high}\}$. The levels are encoded by three-class OrRD sequential colors from orange to red, as shown in Fig. 5 (*a*). We further derive a topological structure from the contours on the map using the Reeb graph. The right of Fig. 5 (*a*) illustrates the corresponding topological structure of the contours shown in the left density map. These conspicuous isolines distinguish the value divergence in the uncertainty density field and enhance identifying the shape and size of regions with specific uncertainty density ranges.

B. Temporal Representation

Besides the spatial representation of arrival time uncertainty at a specific time interval, we also need to present the evolution of uncertainty along a temporal dimension (*R2.2*). Such a view requires addressing three sub-tasks: 1) mapping spatial space (2D) to temporal space (1D), 2) tracking the evolution of high-uncertainty regions, and 3) revealing the uncertainty density of different levels.

1) Mapping Mechanism: To associate uncertainty in spatial and temporal dimensions, an imperative step is needed to appropriately transform the geographical location into 1D space, such that the density distributions among different time intervals can be connected via nested tracking graph. We distill the problem into a projection from 2D pairs of latitude and longitude to 1D space. The intuitive way is to directly map a spatial location by only longitude and latitude as depicted in Fig. 5 (*c*). However, this method can only reveal the distance information of one direction and ignore the other one. To better preserve the spatial distance, we leverage Hilbert curve (Fig. 5 (*b*)) that traverses every point in a given graph and keeps them in a closed loop. Hilbert curves are drawn precisely by connecting these cells to a completed line, which means for any coordinate (x, y) in unit square $[0, 1]^2$, there is always a corresponding point in unit line segment [0, 1].

As shown in Fig. 5 (*b*), the curve ranging from the start point (green) to the end point (red), is about to fill up the map. The line on the right is the straightened Hilbert curve corresponding to the one on the map, which is a finite line whose length equals the number of pixels in the map view since we constrain the minimum unit of a single pixel. Take

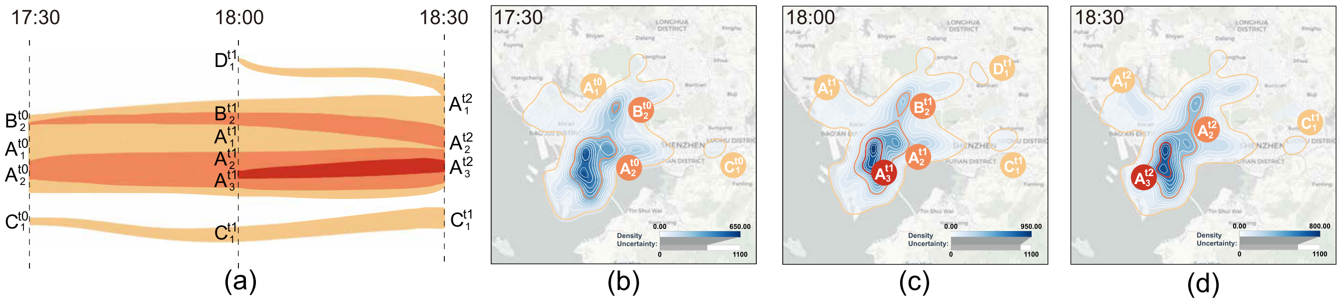


Fig. 6. Illustration of uncertainty evolution from 17:30 to 18:30 using nested tracking graph (a). Corresponding map views at 17:30, 18:00 and 18:30 are shown in (b), (c) and (d), respectively.

Fig. 5 as an instance, where several regions of different uncertainty levels are integrated to form $Topology := [[A_1 \rightarrow [[A_2 \rightarrow A_3], B_2], [C_1 \rightarrow C_2]]$, which has two nesting trees. We denote each region within a certain uncertainty as a node there. Specifically, the mapping algorithm can be summarized as follows. For each nesting tree, we first abstract the nodes N with nested order of mapping through Depth First Search (DFS). The properties of exploring as far as possible and backtracking along each branch enable us to traverse branches with nested structures one by one. In addition, we notice that the points of highest uncertainty are the same within the same branch. To alleviate the redundant computation, we only consider the $root$, including the root of the tree and the first node of new branches, which are the conclusive factors to decide the point. For instance, when dealing with the first tree, we get the order that $\{A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow B_2\}$, which traverses a deeper branch firstly and then backtracks to another branch from crossing region A_2. $\{A_2, B_2\}$ are included into set $root$ because they are determining nodes to locate the mapping points. Then we map the point of highest uncertainty and relative region size into the Hilbert curve following the order of node, respectively. For point projection, we find the point presenting the highest uncertainty within the region. For region size projection, we first compute the area of each region of that level. With the screen size and region size, the proportion between them is preserved and is simple to transform into a normalized Hilbert curve with a range of [0, 1]. Accordingly, the region is projected as a line segment depending on the predetermined point and proportion, as illustrated in Fig. 5 (*b*-right). All nesting trees in the $Topology$ will repeat point projection and region size projection iteratively. Existing works (*e.g.*, [45]) map 2D geographical contours to 1D axis directly based on their positions along y-axis (Fig. 5 (*c*)). This naive method omits distances along x-axis, causing overlaps for horizontally faraway contours such as B and C. Instead, our proposed Hilbert curve can separate them well.

2) Nested Evolution: Two components of uncertainty evolution are fully valued: uncertainty values across multiple levels within a region, dynamic variation within and between regions, including spatial movement, region size change, *etc*. We leverage the Nested Tracking Graph (NTG) technique to accomplish these two goals. The nesting hierarchy property contributes to presenting multi-level uncertainty. Moreover, the topology structure of NTG can easily capture the dynamic evolution of uncertainty over time.

As described in Fig. 6, the empirically defined thresholds for dividing the hierarchy of uncertainty density are encoded from yellow to orange to red, which show as isolines in the map view while presented as a nested graph in the temporal view. Based on such levels $\mathcal{L}$, we can obtain three main components to construct a nested tracking graph $G = (N, E_T, E_N)$. We follow the notation at Fig. 5 (*a*) to tag the nested hierarchical uncertainty as 1, 2, 3, respectively. Each $n \in N$ is a region at given uncertainty levels, for instance, we can get regions $\{A_1^{to}, A_2^{to}, B_2^{to}, C_1^{to}\}$ in Fig. 6 (*b*). E_T represents the tracking graphs of each level along the temporal dimension, as the regions $\{A_1^{to}, C_1^{to}\}$ at 17:30 evolve into $\{A_1^{t1}, C_1^{t1}, D_1^{t1}\}$ at 18:00, which are all region above l_{low} of uncertainty. E_N are nesting glyphs of the relationship between different levels within the same region at the same timestamp, as $\{\{A_1^{t0}, \{A_2^{t0}, B_2^{t0}\}\}, C_1^{t0}\}$ at 17:30. Since we maintain the relationship of region sizes during mapping, the width of n_{low} is greater than n_{high}, allowing us to visualize the hierarchy with successive levels nested within each other. As the left of Fig. 6 (*a*) shows, it is obvious to obtain multiple levels of uncertainty within the same region of the same timestamp. By observing the shape of the nested graph at different levels, we can clearly see when the high-level uncertainty begins and ends. Also, the relation among different regions is displayed, including split and merge, which provide valuable clues for tracking uncertainty from a geographical perspective.

VI. MetroBUX Interface

We develop the interface with five coordinated views: *Temporal View* and *Map View* present the general spatial distribution (*R2.1*) and temporal evolution (*R2.2*) of bus operation uncertainty. Use can further check the detailed uncertainty evolution along trips through *Trip View* (*R2.2*). Moreover, the auxiliary information is provided through *Route List* view and *Stop View* for users to reason the uncertainty (*R3.1*).

A. Route List

Route List view, shown as Fig. 7 (*b*), lists all routes for users to examine. The view is essentially a table with rows representing bus routes, and columns containing route information (e.g., route ID, route type, number of daily trips, frequency, and headway). Two bus routes sharing the same

Fig. 7. Interface of *MetroBUX* system for the input dataset (a). *Route List* (b) view displays general information of all routes, including route ID, route type, number of daily trips, frequency, and headway. *Temporal View* (c) provides an overview of and allows users to explore the evolution of uncertainty over time interactively. *Map View* (d) presents region-level uncertainty distribution over space and user-selected bus route. *Trip View* (e) displays detailed information of a specific route, including frequency (e1) and trip-level uncertainty information (e2). *Stop View* (f) displays the finest stop-level uncertainty information of the selected stops.

route ID but different route types (up *vs.* down) are grouped together. *Departure frequency* is shown in the third field with barchart. We count the departure frequency of each bus route in every three hours, then use a word-scale bar chart to show the frequency histogram in one day. *Headway* represents the amount of time between two consecutive bus trips arriving at a stop. We measure the average headway of all bus stops along a bus route in every 30 minutes and display the headway information as a word-scale line chart. From the route list view, users can select routes of interest for further analysis such as route comparison. The selected route will be displayed in the *Map View*, while more detailed departure frequency and arrival time information will be displayed in the *Trip View*.

B. Temporal View

Temporal View (Fig. 7 (*c*)) incorporates the nested tracking graph to show the uncertainty evolution with time (*R2.2*). We opt for three levels for the uncertainty contours, represented as three-class OrRD sequential discrete colors. The view also allows users to select a particular time period (*R4.1*), by dragging or clicking on the time slider to select the time period of interest. Once a time range is selected, the *Map View* will update to show the corresponding uncertainty density map.

C. Map View

Map View allows users to explore the spatial distribution of uncertainty (Fig. 7 (*d*)). The view adopts a sequential continuous colormap from white to blue to encode continuous uncertainty values estimated by KDE within a given time interval (Sec. V-A.1). Since the ranges of uncertainty density at different time intervals vary greatly, using a consistent scale for map views at all time intervals is not feasible. We choose to adapt the scale to fit the density range at each time interval. To remind users that the scale of colormaps is different at different time intervals, we put an additional progress bar beneath the colormap legend to indicate the current range within the global range; see Fig. 7 (*d*) for example. Moreover, we also highlight the three-level uncertainty contours, with colors corresponding to the color legend used in the *Temporal View*.

In addition to contour map, *Map View* also enables point-based visualization of stop arrival uncertainty (Fig. 7 (*d1*)) or route connections (Fig. 7 (*d2*)). When displaying the stop arrival uncertainty, we compute arrival time uncertainty for each stop per bus route and adopt the Viridis colormap for visual encoding of uncertainty values. For route connections, we adopt varying hues for different routes, and highlight selected stops by enlarging the dots. In both cases, the uncertainty density map will be disabled, allowing users to observe the spatial context. Users can interactively switch between the modes to meet different analysis requirements.

D. Trip View

Trip View (Fig. 7 (*e*)) visualizes the trip-level uncertainty with a given route. It consists of two sub-components:

frequency histogram (Fig. 7 (*e1*)) showing the departure frequency per hour by barchart; and *graphical schedule* (Fig. 7 (*e2*)) visualizing operation uncertainty for each trip through Marey's graph. Specifically, *frequency histogram* is a finer-grained visualization of departure frequency sparkline presented in *Route List* view. From this view, we can observe that there are typically more trip frequencies in the morning and evening peak hours, and trips seldom operate after 21:00. In the *graphical schedule* view, x-, y-axis indicate the operation time in one day and the stations along the route, respectively. Thus, each trip can be presented by a polyline connecting a sequence of dots indicating the temporal information to reach a specific stop. To visualize the stop-level uncertainty along a route, we use Viridis sequential colormap and size to double-encode the uncertainty value with each given stop, as measured by standard deviation, *e.g.*, the point with the lighter color and smaller size represents a stop with lower uncertainty.

E. Stop View

Stop View (Fig. 7 (*f*)) lists basic information of the selected stops with a sequence of widgets, to facilitate stop-level uncertainty exploration (*R3.3*). The stop id is shown at the head of each widget. The number of routes passing through this stop, the current route id, the order of the stop in the route and the average uncertainty of this stop are also shown.

F. User Interaction

MetroBUX provides a series of user interactions to meet the requirements of spatio-temporal, and route-, trip- and stop-level analysis of bus operation uncertainty, including:

- *Animation. MetroBUX* provides users with a quick overview of the spatio-temporal changes of uncertainty. Users can click the play button to start the animation. The *Map View* will display the corresponding Uncertainty Density Map one by one. Users also can pause and replay the animation.
- *Selection.* Users can select a specific time period of interest from *Temporal View*, and *Map View* will show the corresponding uncertainty density map (*R4.1*). If users are interested in some routes, they can select routes from *Route List* view and compare routes through *Map View* and *Trip View*. Users can also select stops in *Map View* and *Trip View*. The selected stop's details will be displayed in *Stop View*, and a route connection view formed by all lines passing through the selected site will be displayed in *Map View*.
- *Filtering.* Users can filter a subset of stops in the *Map View* with a *Lasso* tool. Upon filtering, *Map View* will display stops in the selected region and the stops' colors map the average uncertainty of all routes passing through the stop at the time period (*R4.1*). *Route List* view lists all routes that cross this region.

VII. Evaluation

We conduct a comprehensive assessment on *MetroBUX*, including three case studies exploring bus arrival time uncertainty from different perspectives (Sec. VII-A - Sec. VII-C), and feedback from three independent experts (Sec. VII-D).

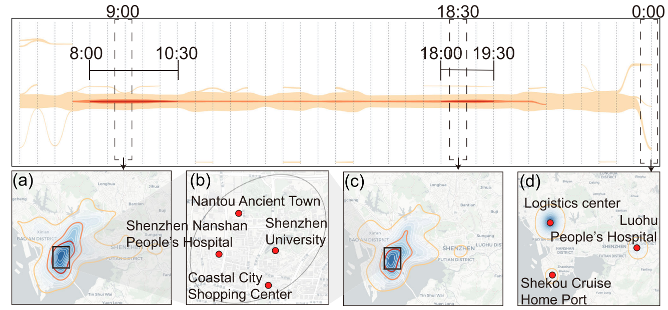


Fig. 8. Case 1: identifying periods and regions of high uncertainty. Corresponding density maps for the chosen time periods in *Temproal View* are presented below.

A. Study 1: Identify Periods and Regions of High Uncertainty

The traffic congestion status is directly correlated to the uncertainty of bus arrival time, which shows various patterns in both spatial and temporal dimensions, as described in *R2.1* and *R2.2*. In this study, we use *MetroBUX* to identify periods and regions of high uncertainty, as illustrated in Fig. 8. The temporal view (Fig. 8-top) depicts that the morning peak hours ranging from 8:00 to 10:30 and evening peak hours from 18:00 to 19:30, have three levels of colors, indicating their high uncertainty. Also, the three nested levels share a joint center for most of the day, and there are no significant fluctuations. This indicates that the corresponding region of high uncertainty exists for most of the day, as discussed in Sec. V-B.

Based on the observation, we select three typical timestamps to explore the spatial features: 9:00, 18:30 and 0:00, which are illustrated in Fig. 8 (*a,c,d*), respectively. The first two timestamps are peak hours, whose uncertainty density distributions are very close, particularly in the region of high-level uncertainty. By zooming in the *Map View* (Fig. 8 (*b*)), we find many sites featuring a mass passenger flow within the region, such as Shenzhen University, Nantou Ancient Town, Shenzhen Nanshan People's Hospital, and Coastal City Shopping Center. In addition, the common region of high uncertainty disappears at 0:00, while three branches appear simultaneously. Through inspecting spatial features in detail, we find that all three uncertainty regions have facilities that still operate at midnight, including a logistic center, the Shenzhen Luohu People's Hospital, and the Shekou Cruise Home Port. There are 24-hour bus routes to commute late-night workers at the logistics center and travelers at the cruise port.

Insights From Expert: MetroBUX efficiently troubleshoots traffic congestion by identifying peak hours and areas with high uncertainty within a district. To address these congestion challenges, our collaborating expert *CR* identified potential solutions by enhancing the frequency of bus services in the district, as shown in Fig. 8 (b). This would involve increasing the number of buses operating on existing routes during peak hours, thereby offering commuters additional transit options. Furthermore, introducing a new circuit line connecting the four stops with the highest uncertainty could distribute the transit

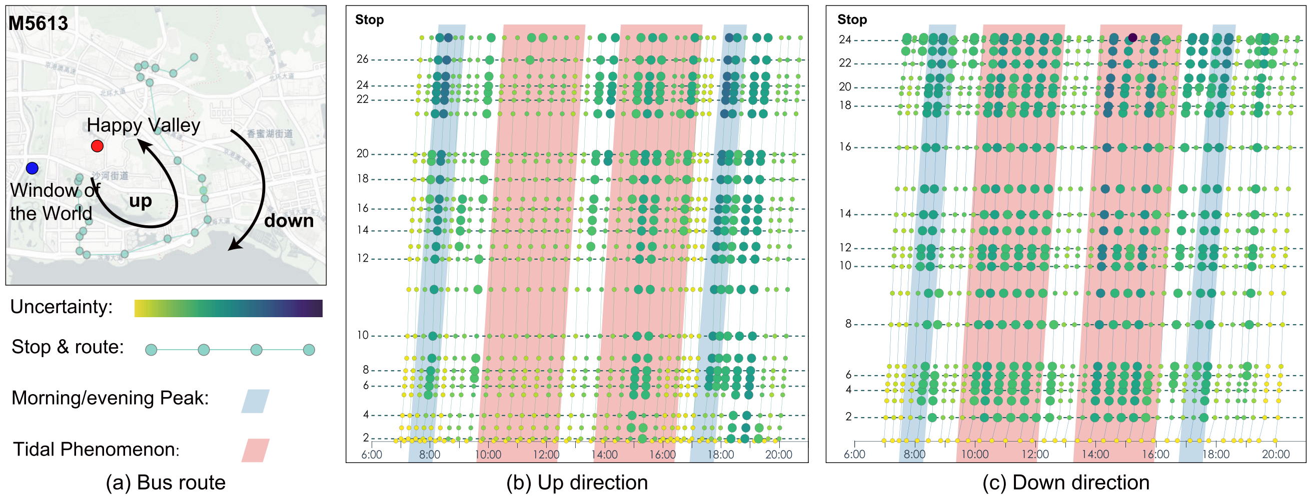


Fig. 9. Case 2: comparing route- and trip-level uncertainty. Route views of both up and down route *M5613* show trip-level uncertainty variations, including temporal difference and propagation of uncertainty along a route, and also differences between the routes that indicate tidal phenomenon.

load more evenly, which distributes the passenger load across multiple routes.

B. Study 2: Compare Route- and Trip-Level Uncertainty

In this case study, we use *MetroBUX* to make a comparison of both trip-level uncertainty within one bus route (*R3.2*), and route-level uncertainty between different routes (*R3.1*). Fig. 9 shows a selected route *M5613* which passes through several amusement parks and scenic spots. Owing to the well-marked string of large points, we can first identify the morning/evening peak hour (blue background) in both up and down route views after browsing all bus trips along the time dimension (Fig. 9 (*b* & c)). We observe the bus trips show a progressive increase of uncertainty following stop order in a route, as the node sides become bigger and darker. This demonstrates the propagation of bus arrival time uncertainty along bus routes. The occurrence of bus bunching tends to increase at these nodes with darker colors and bigger sizes, especially in the up direction's evening peak hours.

Then we try to identify directional differences by comparing the bus trips in up and down directions, as shown in Fig. 9 (*b* & c)). We find an interesting time range, between 10:00 and 12:00, when the uncertainty level is relatively low in the up direction, while the down direction shows much higher uncertainty across the routes, as marked with red background. Such occurrence conforms to one of the features of the typical tidal phenomenon [65]. However, both up and down directions of routes reveal evenly distributed uncertainty between 14:00 and 16:00, which slightly deviates from the definition of the tidal phenomenon that assumes the condition of the evening should appear opposite to the morning. We also notice that the ending stop of the down direction is close to two famous playgrounds: Happy Valley and Window of the World. Depending on the function of these venues, a great number of individuals visit there in the morning, and some arrive or depart in the afternoon, leading to a differentiated tidal phenomenon, as *CR* explained.

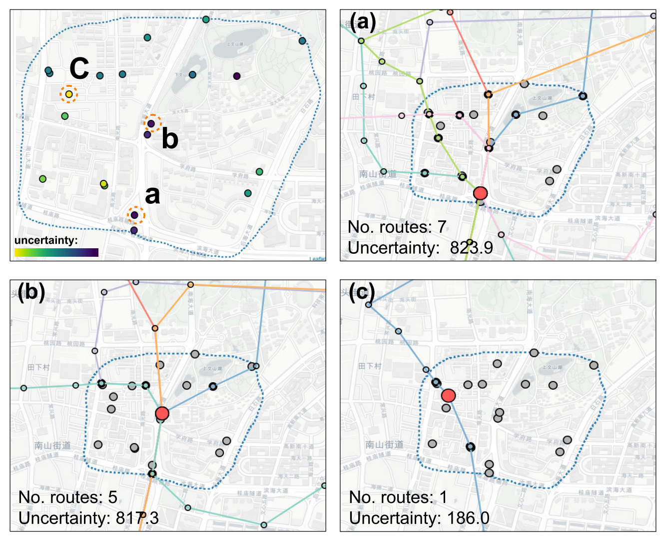


Fig. 10. Case 3: understanding reasons for high uncertainty. The bus stops with high uncertainty show a common feature of many routes passing through.

Insights From Expert: MetroBUX allows analysts to easily identify the tidal phenomena and the bus bunching incidents, and further analyze their causes as tourists commute between attractions. Based on the observations from this study, *CR* suggested two primary strategies to counter these challenges. First, the operator can create dedicated routes that serve solely tourist attractions, thereby reducing potential congestion along these paths. The second approach involves adjusting the signal light timings during peak hours for these routes, increasing the duration of green lights while reducing red light intervals. This adjustment aims to expedite bus operations and prevent bus bunching.

C. Study 3: Understand Reasons for High Uncertainty

In this case, *MetroBUX* is used to conduct fine-grained granularity analysis of reasons for high bus operation uncertainty at trip level (*R3.2*). As shown in Fig. 10, we first delineate a customized region with the lasso tool. All regional stops have a similar spatial context since the region is small. However, the stops still show large variations in arrival time uncertainty. The stops with high uncertainty are marked in dark colors, while those in low uncertainty are in light colors. By plotting stops with high uncertainty, one can notice that most of these stops (Fig. 10 (*a* & b)) have multiple routes passing through. For instance, stop *a* with the greatest uncertainty 823.9 has the most number of seven routes. A possible reason is that many people and buses will pass through the transfer stop, which will likely cause road congestion and increase the stop's uncertainty. In comparison, stop *c* (Fig. 10 (*c*)) is simple with only one route passing through, and its uncertainty is small. Moreover, one can notice that if two stops are close by and share the same routes, then the stop are likely to have similar uncertainty, which could be demonstrated in the uncertainty values of stops *a* &b. This is probably because of the delay propagation effect along the same route.

Insights From Expert: The phenomenon of uncertainty propagation becomes particularly noticeable along stops *a & b*, both of which are located at busy intersections, which experience a high number of passengers and vehicles. A common presumption is that the uncertainty of a stop would correlate with the number of routes traversing it, it is interesting to note that stop *b* has a similar level of uncertainty as stop *a*, despite having two fewer routes. To alleviate the pressure on stop *b*, *CR* suggested to redistribute some of its routes to nearby stops exhibiting lower uncertainty levels. As depicted in of Fig. 10 (top-left), the stop adjacent to stop *c* in green color and sharing the same intersection with stop *b* can be a potential candidate.

D. Expert Interview

We further conducted interviews with five independent domain experts (*E1, E2, E3, E4 and E5*) who are not the authors of this work, to evaluate the performance of *MetroBUX* and collect their feedback for improvement. *E1* has worked in the transportation domain, specifically in public transport management and data analytics, for over ten years. *E2* has extensive experience in applying data-driven approaches to intelligent transportation system. *E3* is a senior researcher in traffic visualization domain, mainly in traffic prediction and risk monitoring. *E4* is a mid-career researcher with expertise in public transport planning and operation. *E5* is a senior manager in a bus company with over 15 years' experience on bus scheduling, timetabling and operation management. The interviews were held in a semi-structured format. For each interview, we first introduced the research background and how we modeled bus operational uncertainty. Then, we explained the interface of *MetroBUX* and demonstrated the case studies through videos. At last, experts were asked to explore the system freely. Each interview lasted for about one hour. Their comments are summarized in four perspectives: *applicability, visual design, interaction design*, and *improvement*.

1) Applicability: All the experts were impressed by the comprehensive functionality of *MetroBUX*, which is *"user friendly"* and *"versatile"*. They all acknowledged that the system has been carefully designed and fulfills the requirements in terms of spatio-temporal analysis and multi-granular exploration ranging from the stop-level to trip- and route-level. The system is helpful for various high-level analytical tasks, including bus bunching examination and reasoning. Moreover, *E1* and *E3* praised the new findings that were not expected, such as the tidal phenomenon in Case 2, which is *"very helpful to understand traffic dynamics and realize intelligent traffic management"*. *E1* and *E5* were keen to use *MetroBUX* to *"explore data from different cities and validate its generalizability"*. *E4* was glad to see some advanced analysis capabilities (e.g., bus bunching, tidal phenomenon) in addition to the essential bus operational data analysis.

2) Visual Design: *E1* and *E2* spoke highly of the temporal representation of uncertainty due to its overview and capacity to show the relationships between regions of high uncertainty. *E1* commented that three-level contours in the density map are enough to identify regions of interest. *E2* expressed the same opinion and further pointed out that *"the hierarchical nesting graph in temporal view make it easy to identify peak/non-peak hour and some seemingly irrational but reasonable routes"*. For instance, the three branches at the end of the temporal view as in Fig. 8 are specific regions that need more attention at midnight. *E3* mentioned that the size and color encoding in trip view was straightforward and helpful to observe the propagation of uncertainty along a route, *"which helps to identify the starting stop of uncertainty increase"*. He also appreciated the intuitive representation that helped recognize bus bunching in the trip view. *E4* acknowledged the considerate design of visualizing and analysing the data from multiple dimensions (e.g., stop-/route-/trip-/network level), which provided comprehensive understanding of bus operation. *E5* appreciated the visual design in a user-friendly and easy-to-understand way, especially with the Map View and Trip View providing complementary perspectives to inspect bus operation.

3) Interaction Design: The experts thought the interaction design was helpful, especially for across-view linkage. *"It's flexible in selecting a time interval of interest in temporal view, and the map that shows corresponding detailed spatial information will appear intermediately"*, commented by *E1*. *E2* and *E3* were particularly interested in stop-level information, and they praised the design of stop selection that shows all routes passing through the same bus stop by simply clicking the stop. Furthermore, *E2* and *E5* deemed it useful to zoom in on Map view, showing fine-grained geographical features like parks or shopping malls. The spatial context helps verify why some stops present high uncertainty. *E4* found it helpful to see both the Map and Temporal views changing simultaneously when the user dragged along the timeline. *E5* acknowledged the lasso tool to be flexible to allow manual selection of any region of interest to analyse the covered bus stops.

4) Improvement: The experts also provided valuable comments for further improvement. *E2* and *E3* suggested to integrate real-time data into the system, thus it can both analyze historical uncertainty and assist in bus arrival time prediction for dynamic management, which significantly extends its applicability. *E1* mentioned that the *What-if scenario* analysis ought to be integrated into the current system, as it yields individual contribution *w.r.t.* change input dynamically. For instance, to analyze the impact of a specific route, the direct approach is to compare the current status with the output of what if the route disappears. Moreover, *E1* was impressed by the route comparison demonstrated in case 2, and encouraged us to further improve the usability by providing side-by-side comparison views to make the exploration more convenient. *E4* suggested to develop a new function in the future to enable various predictions (e.g., bus arrival time, bus arrival delay, bus bunching) based on both historical data and real-time data. *E5* would like to see an extension from standard bus services to including more options of services such as on-demand bus and shuttle bus services. Relevant bus routes can be categorized by their service types.

VIII. Limitation and Future Work

1) Limitation: There are two limitations in *MetroBUX*. From *application* perspective, the current system does not provide multiple Trip View simultaneously, hindering the ability to compare different bus routes. When comparing route-level uncertainty, *"it is laborious to remember trip information of the last route"*, as commented by *E1*. Nevertheless, this is due to the constraint of the limited display space, which encloses spatial and temporal views to provide more information from different perspectives. *E1* understood the trade-off, and expressed her vision of supporting route comparison more effectively in the future. Second, from *visualization* perspective, the Hilbert curve transforms the points in a 2D field into 1D space while not completely considering the relation of position, leading to spatial information loss to some extent. Moreover, the proposed system employs KDE and density map to reveal regional uncertainty distribution in the map view. We have adopted an adjustable bandwidth based on stop density to eliminate its contribution to the uncertainty density. However, KDE, by nature, averages uncertainty density that will omit individual differences among nearby stops.

2) Future Work: MetroBUX serves for exploring bus operation based on historical uncertainty analysis. To extend its applicability, one promising direction is to integrate real-time operational data to estimate the real-time traffic conditions and inform real-time decision making. *CR*, *E1* and *E4* also expressed the vision of integrating simulating or prediction models into the current system, which will enable users to conduct *what-if scenario* analysis. *E1* and *E5* expressed interest in the point, as many arrival time prediction models treat individual routes separately, which cannot capture the effect of other routes. *E1* expects the system to confirm the observation further. Additional geographical factors like population density and land use information can be added to explore the cause of arrival time uncertainty. Moreover, as mentioned in the limitation, we plan to add functionality with stop-/route-specific assessment by comparison of before and after removal/addition of a single stop or route. Such flexible functionality can assist users in detecting and verifying the influence of each specific stop or route, contributing to a more comprehensive understanding and management of the bus system. Last, incorporating passenger demand information (e.g., origin-destination distributions and patronage at each bus stop/route) into the system will be helpful to understand the level of service from both supply and demand sides. In addition to seeking additional information, we would also like to explore ways to maximize the utilization of the existing data, such as filtering out anomalies that cause significant fluctuations in time reliability. In the future, we will continue to refine the system and subsequently extend its usage to a broader user base. In this way, we can gathere feedback to enhance the effectiveness of our system.

IX. Conclusion

In this work, we propose *MetroBUX*, a visual analytics tool to explore uncertainty inherent in bus operation. *MetroBUX* makes threefold contributions including multi-granular uncertainty modeling, integrated spatio-temporal visualization design, and a multi-view interface complemented with user-friendly interactions. Specifically, the system supports the exploration of bus operation uncertainty, including spatio-temporal investigation and hierarchical analysis, namely identifying regions of high uncertainty, comparing routes of divergent behavior, inspecting trips within a route, and exploring stops in the bus route network. A series of case studies demonstrate our system's efficiency towards the detail-oriented exploration of uncertainty, which is confirmed by domain experts in collected feedback.

Acknowledgment

The authors wish to thank the anonymous reviewers for their valuable comments.

References

[1] C. Iclodean, N. Cordos, and B. O. Varga, "Autonomous shuttle bus for public transportation: A review," *Energies*, vol. 13, no. 11, p. 2917, Jun. 2020.

[2] X. Chen, L. Yu, Y. Zhang, and J. Guo, "Analyzing urban bus service reliability at the stop, route, and network levels," *Transp. Res. A, Policy Pract.*, vol. 43, no. 8, pp. 722–734, Oct. 2009.

[3] Z. Gong, B. Du, Z. Liu, W. Zeng, P. Perez, and K. Wu, "SD-seq2seq: A deep learning model for bus bunching prediction based on smart card data," in *Proc. 29th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2020, pp. 1–9.

[4] B. Du and P.-A. Dublanche, "Bus bunching identification using smart card data," in *Proc. IEEE 24th Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2018, pp. 1087–1092.

[5] B. Barabino and M. Di Francesco, "Diagnosis of irregularity sources by automatic vehicle location data," *IEEE Intell. Transp. Syst. Mag.*, vol. 13, no. 2, pp. 152–165, Jan. 2021.

[6] A. Ceder, S. Hassold, and B. Dano, "Approaching even-load and even-headway transit timetables using different bus sizes," *Public Transp.*, vol. 5, no. 3, pp. 193–217, Oct. 2013.

[7] D. Sun, Y. Xu, and Z.-R. Peng, "Timetable optimization for single bus line based on hybrid vehicle size model," *J. Traffic Transp. Eng.*, vol. 2, no. 3, pp. 179–186, Jun. 2015.

[8] Y. Wang, D. Zhang, L. Hu, Y. Yang, and L. H. Lee, "A data-driven and optimal bus scheduling model with time-dependent traffic and demand," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2443–2452, Sep. 2017.

[9] N. Andrienko, G. Andrienko, G. Fuchs, and P. Jankowski, "Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces," *Inf. Visualizat.*, vol. 15, no. 2, pp. 117–153, Apr. 2016.

[10] R. Beecham and J. Wood, "Exploring gendered cycling behaviours within a large-scale behavioural data-set," *Transp. Planning Technol.*, vol. 37, no. 1, pp. 83–97, Jan. 2014.

[11] S. Yan and C.-H. Tang, "An integrated framework for intercity bus scheduling under stochastic bus travel times," *Transp. Sci.*, vol. 42, no. 3, pp. 318–335, Aug. 2008.

[12] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Syst. Appl.*, vol. 120, pp. 426–435, Apr. 2019.

[13] J. Wu et al., "The bounds of improvements toward real-time forecast of multi-scenario train delays," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2445–2456, Mar. 2022.

[14] J. Mulerikkal, S. Thandassery, D. M. Dixon K, V. Rejathalal, and B. Ayyappan, "JP-DAP: An intelligent data analytics platform for metro rail transport systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9146–9156, Jul. 2022.

[15] W. Chen, F. Guo, and F.-Y. Wang, "A survey of traffic data visualization," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2970–2984, Dec. 2015.

[16] G. Andrienko, N. Andrienko, W. Chen, R. Maciejewski, and Y. Zhao, "Visual analytics of mobility and transportation: State of the art and further research directions," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2232–2249, Aug. 2017.

[17] Z. Deng, D. Weng, S. Liu, Y. Tian, M. Xu, and Y. Wu, "A survey of urban visual analytics: Advances and future directions," *Comput. Vis. Media*, vol. 9, no. 1, pp. 3–39, Mar. 2023.

[18] W. Zeng, C.-W. Fu, S. Müller Arisona, A. Erath, and H. Qu, "Visualizing waypoints-constrained origin-destination patterns for massive transportation data," *Comput. Graph. Forum*, vol. 35, no. 8, pp. 95–107, Dec. 2016.

[19] C. Palomo, Z. Guo, C. T. Silva, and J. Freire, "Visually exploring transportation schedules," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 170–179, Jan. 2016.

[20] D. Weng et al., "Towards better bus networks: A visual analytics approach," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 817–827, Feb. 2021.

[21] N. Andrienko, G. Andrienko, F. Patterson, and H. Stange, "Visual analysis of place connectedness by public transport," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3196–3208, Aug. 2020.

[22] Z. Feng, H. Li, W. Zeng, S.-H. Yang, and H. Qu, "Topology density map for urban data visualization and analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 828–838, Feb. 2021.

[23] M. Wunderlich, K. Ballweg, G. Fuchs, and T. von Landesberger, "Visualization of delay uncertainty and its impact on train trip planning: A design study," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 317–328, Jun. 2017.

[24] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson, "When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2016, pp. 5092–5103.

[25] D. Hilbert, "Über die stetige Abbildung einer linie auf ein Flächenstück," in *Dritter Band: Analysis · Grundlagen der Mathematik · Physik Verschiedenes*. Berlin, Germany: Springer, 1935, pp. 1–2.

[26] D. Weng, R. Chen, J. Zhang, J. Bao, Y. Zheng, and Y. Wu, "Pareto-optimal transit route planning with multi-objective monte-carlo tree search," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1185–1195, Feb. 2021.

[27] C. Chen, D. Zhang, N. Li, and Z.-H. Zhou, "B-Planner: Planning bidirectional night bus routes using large-scale taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1451–1465, Aug. 2014.

[28] W. Zhao, H. Jiang, K. Tang, W. Pei, Y. Wu, and A. Qayoom, "Knotted-line: A visual explorer for uncertainty in transportation system," *J. Comput. Lang.*, vol. 53, pp. 1–8, Aug. 2019.

[29] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli, and R. Nair, "AllAboard: Visual exploration of cellphone mobility data to optimise public transport," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 2, pp. 1036–1050, Feb. 2016.

[30] W. Zhao, G. Wang, Z. Wang, L. Liu, X. Wei, and Y. Wu, "A uncertainty visual analytics approach for bus travel time," *Vis. Informat.*, vol. 6, no. 4, pp. 1–11, Dec. 2022.

[31] H. Liu et al., "Visualization and visual analysis of vessel trajectory data: A survey," *Vis. Informat.*, vol. 5, no. 4, pp. 1–10, Dec. 2021.

[32] H. Wang et al., "Hierarchical visualization of geographical areal data with spatial attribute association," *Vis. Informat.*, vol. 5, no. 3, pp. 82–91, Sep. 2021.

[33] T. Kapler and W. Wright, "GeoTime information visualization," in *Proc. IEEE Symp. Inf. Visualizat.*, Jan. 2004, pp. 25–32.

[34] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2565–2574, Dec. 2012.

[35] N. Willems, H. Van De Wetering, and J. J. Van Wijk, "Visualization of vessel movements," *Comput. Graph. Forum*, vol. 28, no. 3, pp. 959–966, Jun. 2009.

[36] S. Hong, Y.-S. Kim, J.-C. Yoon, and C. R. Aragon, "Traffigram: Distortion for clarification via isochronal cartography," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2014, pp. 907–916.

[37] S. Hong, R. Kocielnik, M.-J. Yoo, S. Battersby, J. Kim, and C. Aragon, "Designing interactive distance cartograms to support urban travelers," in *Proc. IEEE Pacific Visualizat. Symp. (PacificVis)*, Apr. 2017, pp. 81–90.

[38] T. Crnovrsanin, C. Muelder, C. Correa, and K.-L. Ma, "Proximity-based visualization of movement trace data," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, Oct. 2009, pp. 11–18.

[39] N. Andrienko, G. Andrienko, L. Barrett, M. Dostie, and P. Henzi, "Space transformation for understanding group movement," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2169–2178, Dec. 2013.

[40] F. Kamw et al., "Urban structure accessibility modeling and visualization for joint spatiotemporal constraints," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 104–116, Jan. 2020.

[41] G. Sun, R. Liang, H. Qu, and Y. Wu, "Embedding spatio-temporal information into maps by route-zooming," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 5, pp. 1506–1519, May 2017.

[42] S. Kim, R. Maciejewski, A. Malik, Y. Jang, D. S. Ebert, and T. Isenberg, "Bristle maps: A multivariate abstraction technique for geovisualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 9, pp. 1438–1454, Sep. 2013.

[43] X. Chen, W. Zeng, Y. Lin, H. M. AI-Maneea, J. Roberts, and R. Chang, "Composition and configuration patterns in multiple-view visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 1514–1524, Feb. 2021.

[44] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1833–1842, Dec. 2014.

[45] J. Lukasczyk, R. Maciejewski, C. Garth, and H. Hagen, "Understanding hotspots: A topological visual analytics approach," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.* New York, NY, USA: ACM, Nov. 2015, pp. 1–10.

[46] H. Doraiswamy, N. Ferreira, T. Damoulas, J. Freire, and C. T. Silva, "Using topological analysis to support event-guided exploration in urban data," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 2634–2643, Dec. 2014.

[47] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl, "TrajectoryLenses—A set-based filtering and exploration technique for long-term trajectory data," *Comput. Graph. Forum*, vol. 32, no. 3pt4, pp. 451–460, Jun. 2013.

[48] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha, "Approaches to uncertainty visualization," *Vis. Comput.*, vol. 13, no. 8, pp. 370–390, Nov. 1997.

[49] G.-P. Bonneau et al., "Overview and state-of-the-art of uncertainty visualization," in *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, Midtown Manhattan, NY, USA: Springer, Sep. 2014, pp. 3–27.

[50] A. Kale, M. Kay, and J. Hullman, "Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–14.

[51] A. Slingsby, J. Dykes, and J. Wood, "Exploring uncertainty in geodemographics with interactive graphics," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2545–2554, Dec. 2011.

[52] H. Chen et al., "Uncertainty-aware multidimensional ensemble data visualization and exploration," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 9, pp. 1072–1086, Sep. 2015.

[53] L. Liu et al., "Uncertainty visualization by representative sampling from prediction ensembles," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 9, pp. 2165–2178, Sep. 2017.

[54] A. M. MacEachren et al., "Visualizing geospatial information uncertainty: What we know and what we need to know," *Cartography Geographic Inf. Sci.*, vol. 32, no. 3, pp. 139–160, Jan. 2005.

[55] T. Gschwandtnei, M. Bögl, P. Federico, and S. Miksch, "Visual encodings of temporal uncertainty: A comparative user study," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 539–548, Jan. 2016.

[56] R. Maciejewski et al., "Forecasting hotspots—A predictive analytics approach," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 4, pp. 440–453, Apr. 2011.

[57] J. Lukasczyk, G. Weber, R. Maciejewski, C. Garth, and H. Leitte, "Nested tracking graphs," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 12–22, Jun. 2017.

[58] J. Lukasczyk, C. Garth, G. H. Weber, T. Biedert, R. Maciejewski, and H. Leitte, "Dynamic nested tracking graphs," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 249–258, Jan. 2020.

[59] B. Barabino, M. Di Francesco, and S. Mozzoni, "Time reliability measures in bus transport services from the accurate use of automatic vehicle location raw data," *Qual. Rel. Eng. Int.*, vol. 33, no. 5, pp. 969–978, Jul. 2017.

[60] F. McLeod, "Estimating bus passenger waiting times from incomplete bus arrivals data," *J. Oper. Res. Soc.*, vol. 58, no. 11, pp. 1518–1525, Nov. 2007.

[61] C. Olston and J. D. Mackinlay, "Visualizing data with bounded uncertainty," in *Proc. IEEE Symp. Inf. Visualizat. (INFOVIS)*, Oct. 2002, pp. 37–40.

[62] C. Li, G. Baciu, and Y. Han, "StreamMap: Smooth dynamic visualization of high-density streaming points," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 3, pp. 1381–1393, Mar. 2018.

[63] O. D. Lampe and H. Hauser, "Interactive visualization of streaming data with kernel density estimation," in *Proc. IEEE Pacific Visualizat. Symp.*, Mar. 2011, pp. 171–178.

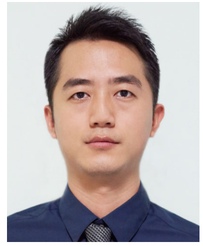
[64] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and K-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.

[65] M. Pei, P. Lin, and J. Ou, "Real-time optimal scheduling model for transit system with flexible bus line length," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 4, pp. 800–810, Apr. 2019.

Shishi Xiao is currently pursuing the master's degree with CMA Thrust, The Hong Kong University of Science and Technology (Guangzhou). Her research interests lie in the interdisciplinary field of data visualization and computer vision.

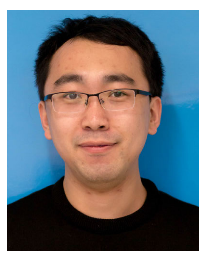
Qing Shi is currently a Research Assistant with the CMA Thrust, The Hong Kong University of Science and Technology (Guangzhou). His current research interests include visualization and visual analytics and HCI.

Lingdan Shao received the master's degree from the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. She is with the University of Chinese Academy of Sciences. Her research interests include data visualization and human–computer interaction. She received the ChinaVis'21 Best Paper Honorable Mention Award.

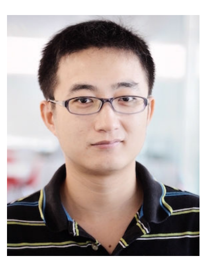


Bo Du is currently a Senior Lecturer and leading the Future Transport and Mobility Group, SMART Infrastructure Facility, University of Wollongong, Australia. His research interests include transport network modeling and optimization, transport data analytics and visualization, and optimal planning and operation of zero-emission and emerging mobility. He is an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Yang Wang received the B.Sc. degree in applied mathematics from the Ocean University of China in 1989, the M.Sc. degree in computer science from Carlton University in 2001, and the Ph.D. degree from the University of Alberta, Canada, in 2008. He is currently with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, as a Professor. He also with Xiamen University as an Adjunct Professor. He was an Alberta Industry Research and Development Associate (2009–2011) and a Canadian Fulbright Scholar (2014–2015). His research interests include service and cloud computing, programming language implementation, and software engineering.



Qiaomu Shen (Member, IEEE) received the Ph.D. degree in computer science from The Hong Kong University of Science and Technology in 2020. He is currently a Research Assistant Professor with the Southern University of Science and Technology (SUSTECH). Before joining SUSTECH, he was a Senior Developer with Noah's Ark Laboratory, Huawei Technologies. His current research interests include spatial–temporal visualization, urban computing, and visual analytics of complex systems.



Wei Zeng (Member, IEEE) received the Ph.D. degree in computer science from Nanyang Technological University in 2015. He is an Assistant Professor with The Hong Kong University of Science and Technology (Guangzhou). His current research interests include visualization and visual analytics, computer graphics, AR/VR, and HCI. He received various Best Paper Awards from conferences, including ICIV, VINCI, and ChinaVis. He serves as the Program Chair for VINCI'23; and a program committee for venues, including IEEE VIS, EuroVis STARs, and ChinaVis.