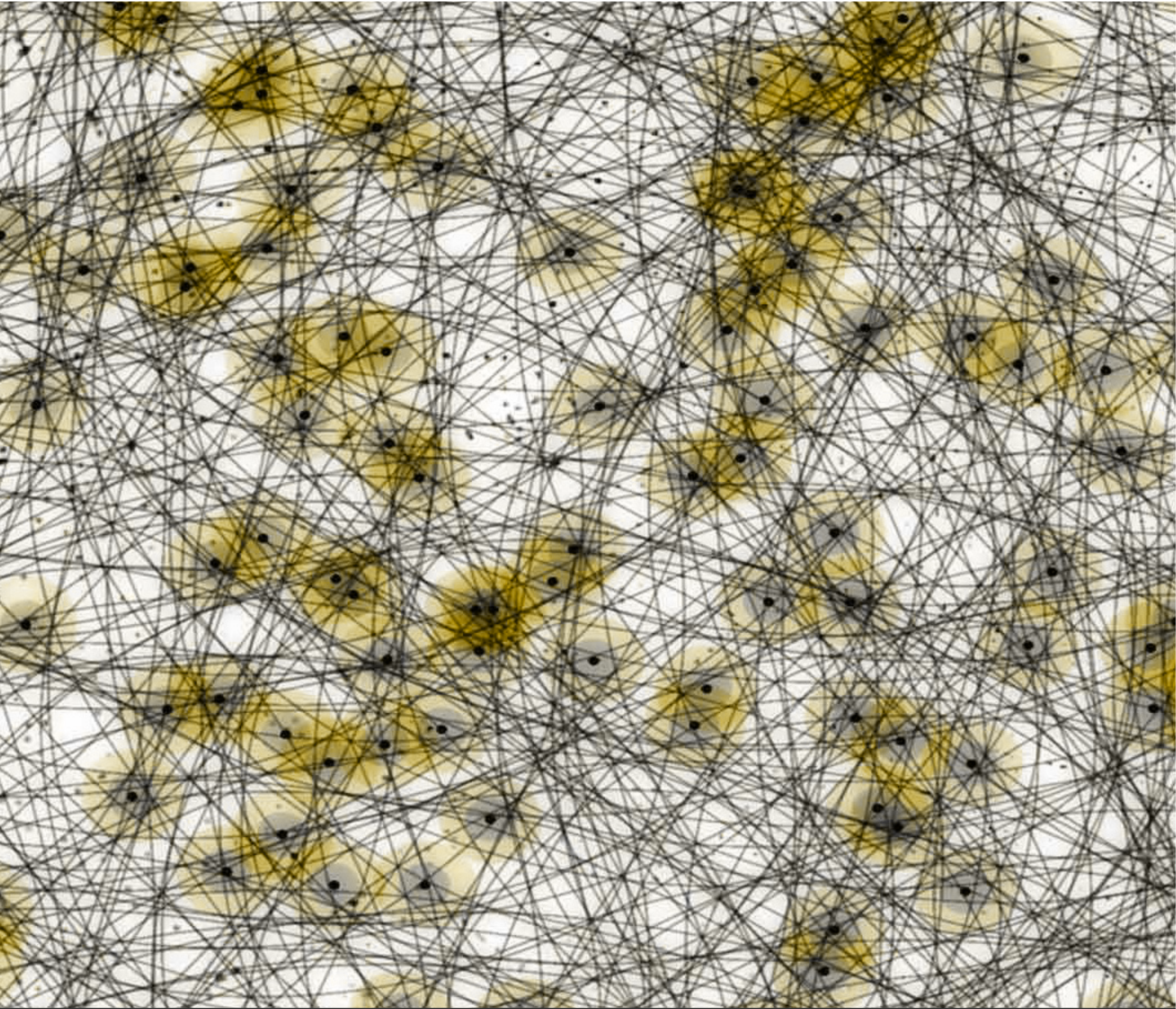


# UTS Final Report

UTS

Urban Traffic System

# Executive Summary

Rapid economic and population growth occurring throughout Saudi Arabia are posing new opportunities and changes for the Kingdom. In particular, that concomitant growth in the Kingdoms capital, Riyadh, is straining the citys road network to the point of becoming a major hindrance to socioeconomic activity. Further stressing the transportation infrastructure is the extraordi-nary growth in demand; between 1987 and 1995, vehicular trips have increased at the rapid rate of 9% per year. The Urban Trac Systems (UTS) project aims to create an alternative to traditional intelligent transportation systems by taking advantage of the digital traces of our everyday lives to create models for analysis, intervention and planning, for both policymakers and private citizens. The trac model is built with the exibility to shift spatiotemporally considering historic patterns and models and generating predictions for the future. The analysis performed within the larger urban context (macromodel) informs work at the local scale (micromodel)in other words, the predictions from the regional and urban scales augment local infrastructure to optimize the transportation system for the inhabitants passing through them.

At the end of year 2, origin-destination matrices (fundamental components of any urban and/or transportation planning initiative) have been constructed, alongside the completion of a wide range of regional human mobility analyses. Models have been obtained and have been calibrated for the Riyadh context; these models implement the ndings of the macroscopic investigations at a nite, local scale. Finally, a robust suite of visualization tools has been developed to make sense of and communicate the complexity of human trajectories through the region. Moving forward, the main objectives for year 3 are the further renements of these models and analyses, with the development of a transportation decision-making mechanism within the mobility data browser that will function at a variety of dimensions. This report summarizes the current state of the project, accomplishments for years 1 and 2, and the initiatives future directions.

# Table of Contents

## Table of Contents

# List of Acronyms

ADA: Ar-Riyadh Development Authority

CCES: Center for Complex Engineering Systems

CDR: Call Detail Record

DSS: Decision Support System

ITS: Intelligent Transportation System

KACST: King Abdulaziz City for Science and Technology

KSA: Kingdom of Saudi Arabia

MIT: Massachusetts Institute of Technology

OD: Origin-Destination

SO: Simulation-based optimization

UTS: Urban Trac SystemMotivation for UTS

# 4. Introduction

## 4.1 Context

Rapid economic and population growth occurring throughout Saudi Arabia is posing new opportunities and changes for the Kingdom. In particular, that concomitant growth in the Kingdoms capital, Riyadh, is straining the citys road network to the point of becoming a major hindrance to socioeconomic activity. Between 1987 and 1995, vehicular trips have increased at the rapid rate of 9% per year [1, 2]. The number of trips on freeways is expanding rapidly, resulting in longer commuting times. For instance, two major radial freeways, King Fahad Freeway and Makkah Freeway, reached their designed capacities of 160,000 vehicles per day two years after their completion in 1991 [2]. Left unaddressed, this strain on infrastructure threatens to weigh down the return on investment from the massive public projects and government-sponsored developments being constructed throughout the city, and adversely aect the quality of life of all residents. As many initiatives are still being planned and/or implemented, there is still substantial room for the implementation of urban and road network development, and the application of trac engineering techniques. For these reasons, a comprehensive study of the mobility situation and outlook, followed by suggestions for the most appropriate action, is crucial to healthy future development of Riyadh, the Kingdom, and its people.

## 4.2 Motivation

The motivation behind the project of UTS is derived by the demand for a system that addresses the exponential growth of urban environment in Riyadhby infrastructure, economics, and populations. The project will investigate current and expected ows of people and vehicles within the citys road network. The project takes advantage of the existing built infrastructures to sense the mobility of people, eliminating the nancial and temporal burdens of deploying a traditional ITS (cameras, computing facilities, etc.). The overall outcome of this project will assist planners and trac engineers in planning new interventions, and the public in making choices about their daily commutes.

New technologies are allowing for new ways to sense and track human activity in the city. While the physical sensors of traditional intelligent trac systems (ITS) provide a great level of accuracy and precision within a highly specied locale, the understanding gained with regards to transportation is geographically limited due to the technologys emphasis on being highly exact and highly local. Applying such technologies on a large scale is often prohibitively expensive. Our analysis has begun at two dierent scales: the macro level, which examines broad, human mobility ows across the city, and the micro level, which zooms into particular intersections of the city for detailed exploration.

The UTS project takes advantage of existing technologies that have been deployed for other purposes, such as for commerce and communication, in order to comprehend the dynamics and ows of material, capital, information and individuals [3]. Every time a credit card is used, a text message or an email is sent, or a purchase is processed on a major online retailer, an entry with the time and location of the transaction is added to a data set on a central server, administered and maintained by the entity providing the platform for these day-to-day operations [4]. Cellular call detail records in particular comprise a crucial information resource for the human mobility model, especially in cities where mobile penetration is high like Riyadh. This data provides the digital traces sucient to build an alternative ITS [5]. The trac model is intended to be able to shift spatiotemporallyconsidering historic patterns and models and generating predictions for the future. The analysis performed within the larger urban context (macromodel) informs our work at the local scale (micromodel). In other words, the predictions from the regional and urban scales augment local infrastructure to optimize the transportation system for the inhabitants passing through them.

Our research also considers going beyond the access or generation of data, which in todays highly connected society is increasingly pervasive. The creation and analysis is not an end in itself. Rather, we gain salience in our eort to make the data not only available, but also useful and understandable while remaining novel. This ethic is reminiscent of the Enlightenment era that supports the concept of witness, an analogy used by Christian Nold [4]. In this paradigm, one cannot just produce knowledge without having other people to see it and use it; knowledge is not attained for its own sake but as a practical asset. This explains the proliferation of publicly viewed or witnessed scientic experiments during this era, and why we believe that the urban data and analysis from our research has to be fed back to local residents, in addition to policy makers and researchers, to give it meaning.

We ground our proposal and research in the urban opportunities and challenges particularly to Riyadh. As a city experiencing rapid growth and containing a very young population, and whose economic diversication as part of the Strategic Plan for 2028 is changing the urban fabric of the city, we address the impact from this growth on the citys infrastructure. We believe our work in UTS helps to not only answer these lingering questions regarding Riyadhs existing infrastructure, but also as a resource by which planners and policymakers can evaluate their decisions and create new opportunities for the regions future.

## 4.3 Project Vision

The project aims to develop an urban traffic system (UTS) to address the human and vehicle traffic challenges in Riyadh. It contains both short-term and long-term components. For the short-term it attempts to investigate and improve the performance of the current road network by using traffic models that account for detailed driver behavior and vehicle technology models. The long-term approach addresses future development patterns across the Riyadh region, the added burden they will place on the road networks, and the major infrastructure improvements they will require.

The tools employed include macro- and micromodels that rely on data including mobile phone usage information and traffic flow/volume figures. By combining information on usage patterns, traffic congestion, and urban development, the project will follow an integrated approach to addressing the challenges of local traffic management and planning. The macromodel, is comprised of 3 components: the human mobility model, the vehicle flow model, and the planning model.

The human mobility model incorporates mobile phone spatial-temporal data, thereby mapping the locations of users within the city at a given time of day. It is thus a two-dimensional understanding of the city, with an additional time dimension to show the movement of people throughout the day within that space.

The flow model captures the road network and is concerned with linear flows as opposed to the spatial construct of the human mobility model that utilizes a general area map of the city.

On a longer trajectory is the planning model, which serves as the long-term component of the project. This anticipated model involves forecasting changes in population, economic factors pertaining to transportation development, and their affect on existing and planned infrastructure.

The purpose of the micromodel is to focus into areas where the macromodel identifies traffic problems, to understand how individual vehicle interactions lead to such problems, and to suggest solutions for alleviating the traffic pressures. The micromodel will yield detailed link specific metrics, such as energy consumption and emission patterns. It will then be embedded within an optimization framework, and be used to identify novel mobility strategies that improve the reliability, robustness, efficiency and sustainability of the network.

# 5. Development Process

## 5.1 Developing the Human Infrastructure

The UTS project is a diverse, multi-disciplinary team bridging MIT and KACST. These perspectives, ranging from computer science to urban planning, civil engineering to interaction design, together offer a rich environment of discourse and inspiration for the collaboration at hand. Indeed, the collaborative relationship established within UTS allows for multiple fingerprints on each of the projects and outcomes.

The structure of the UTS project, described under Project Vision, is based on the scale of proposed intervention. Within each scale, the project teams carry the same ethic of multiple fingerprints as members across disciplines, and from both KACST and MIT collaborate together. For the researchers, this pairing offers a chance to not only draw from the knowledge, expertise and disciplinary viewpoint of each member synergistically toward the research aims, but to learn from each other as well. This structure also allows for separate yet coordinated streams of research to occur simultaneously.

In addition to the multiple in-person workshops, the use of technology has allowed the teams based across the breadth of MIT and across continents, between Cambridge and Riyadh to collaborate seamlessly on projects. Regular WebEx teleconferences with the entire team allow for the sharing of outcomes and ideas, as well as the fostering dialogue across the scales and the centers. Individual team members have also had informal team meetings using various technologies. Anecdotally, there was a series of online conferences via Skype during MITs Independent Activities Period where team members collaborated from Singapore, Los Angeles, Cambridge and Riyadh, while simultaneously drawing and commenting over the same document.

The team also uses collaboration technology regularly to manage knowledge resources and outcomes. A Google Sites page serves as a common location for members to share documents and data. Information requiring higher levels of security is also stored on servers in a manner where collaboration and security are both accomplished.

Active relationships with various stakeholders and decision and policy makers have also added to the research by providing insight into the operations, challenges and opportunities for Riyadh. This dialogue allows the team to identify key research questions and policy priorities for the decision makers. The project has close collaboration, and on-going dialogue, with various Ministries across the Kingdom, as well as municipal agencies including the ADA and Transportation Police. The team is also in continued dialogue with data providers from both the public and private sectors.

## 5.2 Developing the Individuals

The UTS team has invested in its talent as well as the infrastructure for research and collaboration:

### Facilities

Joint centers are being created in Riyadh and on the MIT campus in Cambridge, Massachusetts to house researchers and provide a concurrent engineering capability to all projects in the CCES portfolio.

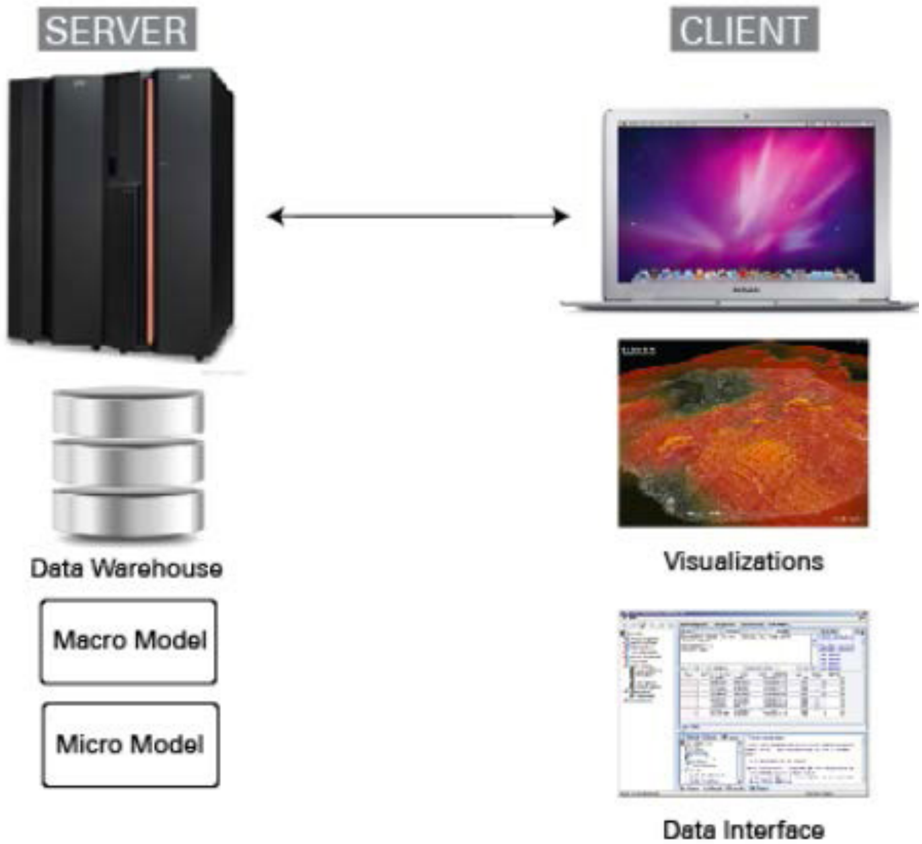
### Hardware & Software

Includes both servers and commercial software products purchased at KACST, including MATLAB, Processing, MySQL, Quantum GIS, Postgre SQL, Visual Studio, Eclipse, as well as UTS-specific software that is being developed by KACST & MIT researchers to support project objectives.

# 6. System Infrastructure

## 6.1 System Architecture

■ *Figure 1. The data warehouse, macro- and micromodels, visualization*




### Geospatial Vector Data

The overall architecture of the UTS system uses a system-of-systems approach to underscore subsystem modularity and enable smart component integration in a holistic decision support system. Most generally, the system will be build on a server/client model, with the server containing the data warehouse (a ready-optimized database), and executing the computationally intense macro and micromodels. The client end will encapsulate the visualization, sense-making products, and user interface to the data streams.

The data warehouse will contain various datasets, such as traffic volume measurements, call detail records, geographical data (shape files, land use maps), census/demographic data etc. Additional server-side components include the mathematical models of human mobility and population dynamics build from CDRs, as well as the micromodels of traffic simulation. These models will be developed using a mixture of software solutions (Aimsun, Matlab, R, custom applications, etc). These models will be large and complex, necessitating their placement on a centralized, easily accessible server. In comparison, the client side visualization models will be leaner and more flexible to facilitate user interaction.

## 6.2 Data Warehouse

The data warehouse houses several databases containing information of the structure of the city as well as the dynamics of it. It contains a geospatial database of the city including the lookup table of the locations of cell towers for the purpose of mapping mobile phone activity to locations. In addition, it contains information of the time series mobile phone usage data as well as traffic counts.

The data warehouse is essentially designed to be a database that is read-optimized i.e., designed to accommodate the needs mining the big data of mobile phone activity rather than the more general uses covered by databases. Several data retrieval techniques such as data partitioning were adopted for faster analysis and retrieval. Whereas the information stored in a database is normalizedi.e. the relations are organized to minimize redundancy to maintain timeliness and integrity in a data warehouse they are deformalized in order to provide faster and simpler queries. Moreover, data mining will help identifying problems and issues from various hidden data patterns [7].

### Road Network Traffic Counts

The data contained in the warehouse includes car counts (flow) conducted by the ADA using pressure tubes laid at points of reading throughout Riyadh. Car counts contain volume of cars passing through 500 points across Riyadh.

For each point of reading, the counting process takes a period of 3-5 consecutive days in which volume was recorded on a 15 minutes interval basis. The points where readings took place are of four main categories: expressways, city exit/ entry points, major and minor arterials, and turnings.

### Geospatial Vector Data

Call Detail Records. Data warehouse also includes call detail records (CDRs) which is data recorded for mainly billing purposes by a mobile phone operator. Cellular activity is one of the most powerful real-time sensing mechanisms currently available to us; the ubiquity of digital devices allow us to capture extremely high-resolution traces of humanity across a variety of dimensions. By filtering this social activity through the geography of the city we can begin to reveal previously inaccessible perspectives into how the urban fabric is used. Saudi Arabia's mobile phone penetration is above 198%an astonishing figure suggesting that many across the Kingdom own more than one mobile device. Documenting the technical details of an anonymized phone call, CDRs show great promise for applied research–they have recently been used to explore human communications, urban dynamics [8], and human mobility patterns [8, 9]. Furthermore, CDRs have been used to extract proxies of social networks. To safeguard privacy, the individual phone numbers are anonymized by the mobile phone operator and replaced with a unique, surrogate security ID. Each record consists of:

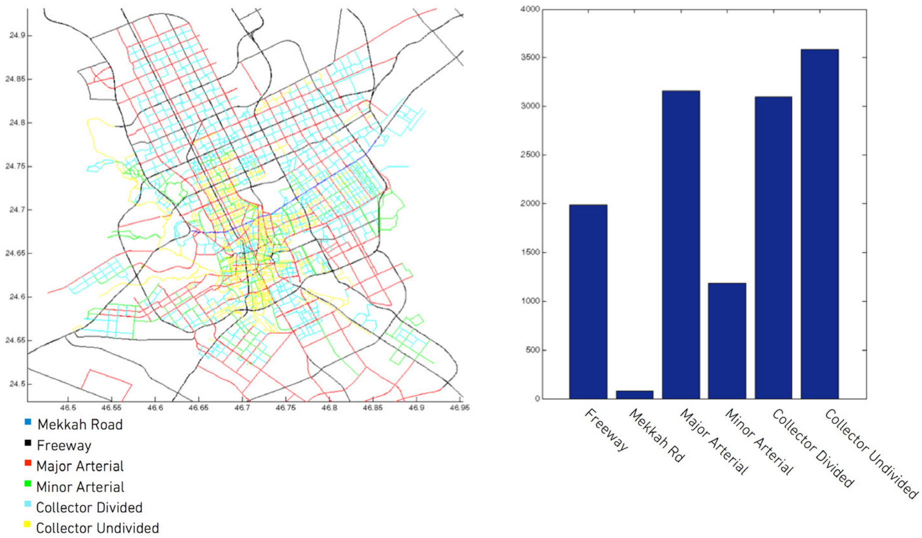
- When a call is placed or received
- When an sms is sent or received
- When the user connects to the internet

Through the established collaboration between KACST and STC, subsets of the CDRs covering intervals of interest around the year are going to be shared. Initially, STC are going to providing data for the first week of July. The data sets are large and can reach to hundreds of millions records per day. The originating and terminating phone numbers are hashed using a one way hashing function to protect the privacy of the people as well as maintain a unique ID for the purpose of analyses. The huge dataset undergoes several steps for it to be ready for querying by the mathematical models mentioned at a later stage.

Our dataset consists of 1 full month of records for the entire county of Saudi Arabia, with 100 million daily network connections to over 10 thousand unique cell towers. Each anonymized record held a precise time and duration measure for the connection, the caller's location, the type of connection (phone call, sms, internet query, etc.), and the user's type of service (subscription, pre-paid, etc.).

In addition to the CDRs, the geographic locations (coordinates) of the cell towers are needed so as to accurately map the cell phone activity within the urban space [10]. Within the proposed structure of the data warehouse, the anonymized CDRs are conveniently stored as an SQL (Structured Query Language) database on a dedicated, secure, server.

*Figure 2. Road Network Shapefile with Length Distribution*




## Geospatial Vector Data

In order to model the flows of vehicles in Riyadh, the vehicle flow model will need a database representing the road network of Riyadh. The database contains information about the spatial traffic network geometry and road parameters (speed limits, traffic lights, signs, et cetera).

The road network information was provided by the ADA, it contains two shape files that include nodes and links. Parameters of each node nodes are its latitude, longitude and unique ID while parameters of links include its starting node, terminating node, free flow speed, capacity number of lanes and length. Within the shape files of the road network of Riyadh, there are 10308 nodes and 146301 links, creating the need for optimized path-finding algorithms.

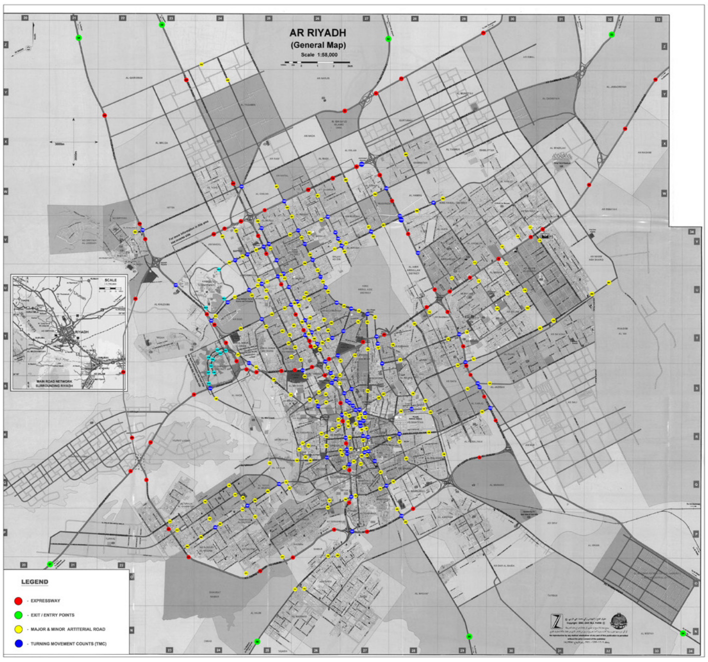



Figure 3. Car Count Intersections

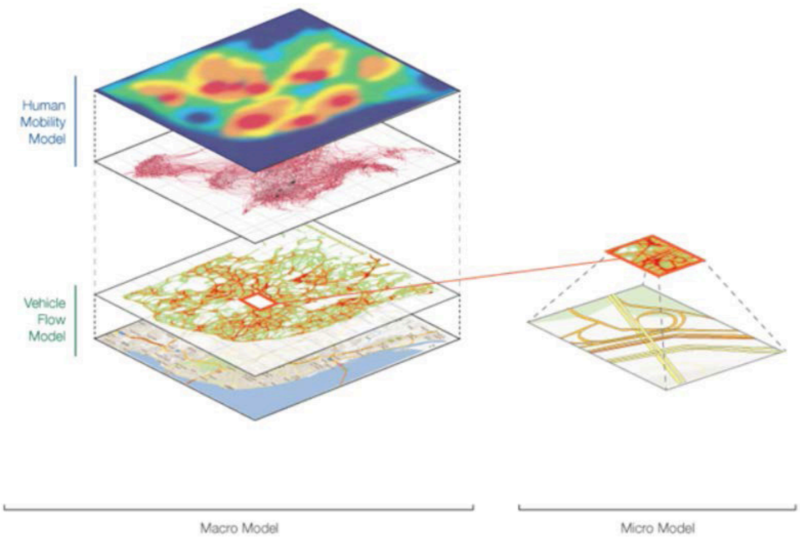
## Car Count Volumes

In addition to the various geospatial data, the ADA has supplied us with a set of transportation volume tables, collected over the course of roughly 2 years. These tables contain hourly vehicle counts for various intersections captured at 48-hour intervals, which were used to produce estimates of annual average daily traffic (AADT). The measures were taken by pneumatic-tube sensors placed at intersections across the city, as demonstrated in Figure 3.

## Additional Data

Among the data that are available and could be included within the data warehouse based on the demand of the models are census, demographics and land use data in addition to data from previously conducted surveys by ADA and the municipality of Riyadh.

Figure 4. The general architecture of UTS mathematical models

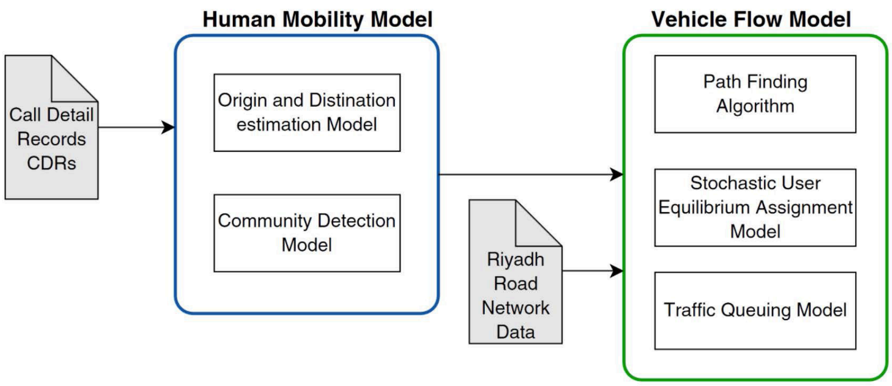


## 6.3 Mathematical Models

The domain of the UTS project incorporates several disciplinary perspectives on the urban transportation system of Riyadh, involving variables from a variety of interactions impacting the traffic system. On the macromodel, UTS attempts to understand the complexity of human mobility and its implications on the flows of vehicles around the road network of Riyadh. The micro-level, areas highlighted by the larger scale will be modeled on a finer level for an understanding of the dynamics within that space, and potential mechanisms for intervention. Figure 4 shows the general architecture of UTS on both macro and micro levels and how the models will be integrated.

The human mobility model will use CDR data as inputs to the model; it will infer the origins and destinations of daily trips as well as the densities of the aggregation of people around the city. The OD of trips is going to be fed into the vehicle flow model to simulate the status of congestions around the city. The model will provide dynamic breakdown of the flows of vehicles on the roads of Riyadh as well as identifications of the points of emergence of traffic congestions.

*Figure 5. Components within UTS macro-model*




## 6.4 Human Mobility Models

In order to unravel and quantify the patterns of human mobility based on CDRs, the UTS project makes extensive use of a large spectrum of existing data mining techniques. Moreover, recent approaches from network and complexity science are applied and further developed. More specifically, the set of mathematical models and tools involves, inter alia:

- Fine-grained and dynamic Origin-Destination (OD) matrices in both spatial and temporal scale to determine people flows at the urban scale. Details of the underlying mathematical model are given in [6].
- Decomposition and spatial clustering techniques to classify urban subcenters with respect to the degree of attractiveness.
- Social network analysis tools such as community detection algorithms or the gravity/radiation model described in [7, 8] to characterize the underlying interconnectivity and cohesiveness of different urban subcenters.

## Origin and Designation Estimation Models

Origin-destination (OD) matrices represent one of the most important sources of information used for strategic planning and management of transportation networks. A precise calculation of OD matrices is an essential component for enabling administrative authorities to optimize the use of their transportation networks, not only for the benefit of users in their daily journeys but also when considering the investments required to adapt these infrastructures anticipated future needs. Traditionally, urban planning and transportation engineering rely on household questionnaires or census and road surveys conducted every 5-10 years and develop methodologies for OD matrix estimation as conducted by the ADA and Municipality of Riyadh. This approach of calculating the OD matrix, from the initial data-gathering to the exploitation of the first results, is lengthy and may take years to get a mere snapshot of the travel demand. Moreover, the collected data has shortcomings both in terms of spatial and temporal scale.

Sensor-based OD estimation methods have also been developed in the past few years, making use of street sensors such as loop detectors and video cameras together with traffic assignment models. Analogous methods that have been developed using probe vehicles, whereby vehicles traces are used as data sources. Those methods are, however, limited by the fact that models are often underdetermined because the number of parameters to be estimated is typically larger than the number of monitored network links.

On the other hand, the wide deployment of pervasive computing devices (e.g. mobile phone, smart cards, GPS devices and digital cameras) provide unprecedented digital footprints that reveal where people are, and when they are there. In previous projects, different methodologies for detecting the presence and movement of crowds through their digital footprint (Flickr photos, mobile phone Call Detail Records (CDRs), smart card records) were developed. UTS adopts a similar approach that uses CDRs to model past behavior and will develop a GPS enabled mobile app to gather real time digital footprints of people roaming Riyadh to generate a fine grained analysis that can potentially make a big leap in terms of understanding the use of space and daily commuting flows for the purposes of urban mobility planning and management. The traditionally generated OD matrices produced by the ADA and the municipality can be incorporated to validate the precision of the generated OD matrices.

The procedure applied within UTS for estimating OD matrices based on CDRs consists of two steps: trip determination and origin-destination estimation. To alleviate the effects of localization errors and event-driven location measurements on individual trip determination, we apply a low-pass filter with a 10-minute resampling rate to the raw data. In addition, the method uses clustering to identify minor oscillations around a common location. The procedure used to handle location errors and identify meaningful locations in a users travel history comprises the following steps:

- We begin with a measurement series $M_S = \{m_q, m_{q+1}, \ldots, m_z\} \in M^{z-q-1}, q > z$, derived from a series of network connections over a certain time interval $\Delta T = t_{mz} - t_{mq} > 0$
- We define an area with radius $\Delta S$, such that maxdistance($p_{mi}$ , $p_{mj}$ ) < $\triangle S \forall q \leq i, j \leq, z$
- All consecutive points $p_j \in M_S$ for which this condition holds can be fused together such that the centroid becomes a virtual location

$p_s = (z-q)^{-1} \sum_{i=q}^{i=z} p_{m_i}$ (the centroid of the points), that is a trips origin or destination.

- Once the virtual locations are detected, we can evaluate the stops (virtual locations) and trips as paths between users positions at consecutive virtual locations. Each trip trip(u, o, d, t) is characterized by user ID u, origin location o, destination location d, and starting time t.
- Filter every trip trip(u, o, d, t) such that duration t is unreasonable to the distance for the trip.
- Once trips are extracted, we use the following procedure to derive OD flows:
  - o The geographical area under analysis is divided into regions: regioni, i = 1, ..., n
  - o Origin and destination regions, together with starting time, are extracted for each trip of each user trip(u, o, d, t)
  - o Trips with the same origin and destination regions are grouped together at different temporal windows tw, for example, weekly, daily, and hourly:

    m(i, j, tw) = $\sum_{0 \in region_i, d \in region_j, t \in tw}$ trip(u, o, d, t)

  - o The result is a 3D matrix M Є R³ whose element m(i,j,tw) represents the number of trips from origin region i to destination region j starting within the time window tw. [12]

### Community Detection Models

To assess the spatial clustering of people into highly interconnected communities (being at the root of trip generation), the project aims at applying partitioning (or community detection) algorithms to the mobile phone network at hand. A widely applied algorithm is based on the optimization of modularity, which is defined as follows [9]:

Consider a weighted, symmetric network of n nodes (representing mobile phone callers), with given weights of edges (representing call duration or number of calls) between node i and node j, denoted by $A_{i,j}$ (with $A_{i,j} = A_{j,i}$). Suppose that nodes also possess loop edges to themselves, i.e., we assume that $A_{i,i} \neq 0$ is possible. We refer to the symmetric matrix A of the values $A_{i,j}$ as a weighted adjacency matrix of the network. Introduce for each node i its strength as the total weight of all the edges connected with this node: $w(i) = \sum_j A_{i,j}$. Also define the total network adjacency matrix weight $M=\sum_j w(i)=\sum_{i,j} A_{i,j}$ .

Consider a suggested partitioning for which c(i) denotes the index of the community to which node i belongs. Then we perform our calculations by optimizing the modularity function defined as $Q = \frac{1}{M}\sum_{i,j}(A_{i,j} - \frac{w(i)w(j)}{M})\delta(c(i), c(j))$ , where δ is the Kronecker symbol, equal to 1 if c(i) = c(j) and 0 otherwise.

## 6.5 Macroscopic Traffic Models

Outputs from the human mobility model are fed into a macroscopic model that simulates traffic on the regional and urban scales. The macroscopic traffic model is going to be useful in the analysis the flow, and congestion, of vehicles through an area of interest. For example, the model may provide insights of a vehicle breakdown on congested street in terms of origins or destinations of cars traveling on that segment. This dynamic model is intended for a more rigorous analysis of these types of traffic incidents, such as when and from where congestion emerges.

The macroscopic traffic model will take the generated dynamic OD information and geometry shape files of the Riyadh road network as inputs. This model can also be understood as a macroscopic traffic queuing model, where road links represent queues. In modeling these queues, the model also takes into consideration the characteristics of vehicular flows. The model can be broken down as the following sub models:

## Path Finding Algorithm

The dynamic OD matrix contains the starting and ending points for a trip. However, this information lacks the paths those trips were taken, something vital to simulate road conditions. One may intuitively assume the shortest path is the one chosen by an individual, but we also know that factors like traffic may have the individual seek alternative paths in trying to reduce their travel time. As a consequence, individuals usually adopt alternate paths that might be perceived to have the shortest travel time. As the condition roads change dynamically throughout the day, we need to identify many paths that could be taken by travelers between two areas, regardless of their conditions at a specific time.

The macroscopic traffic model is going to house an improved version of Dials algorithm [18, 19]. The algorithm finds all possible paths from and origin and destination such that those paths consist of reasonable and efficient links. Efficient links are defined to be those links that travel farther from the origin and closer to the destination [18]. To identify those links, each link is labeled by O(i) and D(j) for a link that goes from i to j nodes, O(i) is the cost of reaching node i from the origin while D(j) the cost of reaching the destination from node j. Efficient links are defined by Dial by the following condition:

$$O(i) < O(j) \text{ and } D(i) > D(j)$$

However, Dials inequality check might filter out paths that have minimal travel time but dont satisfy the defined inequality[19]. Therefore, we are going to include Leurent [20] definition of efficient paths that states:

$$C_k^{od} \leq (1+F)\, C_{min}^{od} \ ,V\ k,o,d$$

Where $C_{min}^{od}$ is the minimum travel cost between o and d, F is a non-negative constant ε(0,1) called the route extension coefficient. Leurent definition of reasonable paths helps find more realistic paths than Dials, as noted by Li and Hunge [21]. The cost of a path is found by summing the time it takes to travel across every link within that path.

## Stochastic User Equilibrium Assignment Model

After finding the set of all reasonable paths, trips originating from origin O to destination D are assigned paths according to Wardrops concept of user equilibrium. The concept states that the journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route [22]. The concept leads to the conclusion that each driver is trying to use the path with minimal cost. As drivers are not fully aware of traffic conditions in real time, it doesnt make since to assume that they are going to use the current minimal cost path. Therefore, we infuse stochasticity into route decision making by incorporating the actual cost of a path with a random variable[18]:

$C^p = c^p + \varepsilon$

Cp reflect drivers perception of the cost in taking path p,$c^p$ is the actual cost of path p and ε is a random component following the Gumbel distribution. After finding the stochastic costs of each path, we can assign traffic flows from an origin to a set of links according to the following calculation[18]:

$$\phi^P = \frac{e^{(-\theta C^P}}{\sum_{X \in P_{od}} e^{(-\theta C^X}} \forall p \in P_{od}$$

where $\phi^P$ is the proportion of the total flow that is going to go through path p,Pod is the set of all reasonable paths that can be taken, $C^p$ is the cost of taking path p and θ is a positive dispersion parameter. Thus, traffic originating from a particular area that is assigned to a certain link p can be found as the following[23]:

$$T^p = \phi^p D^{od}$$

where $D^{od}$ is the total number of people flowing out the area and φp is the proportion of traffic flowing through path p.

## Queuing traffic model

After trips are assigned to paths following Wardrops concept of user equilibrium, the queuing model keeps track of that status of roads according to the flows of vehicles. For every link, the model keeps track of the exit time associated with each entering time [23]:

$$t_l(s) = s + f_t + \frac{\bar{x}_(l(s))}{/\delta_l}$$

where $t_l(s)$ is the exit time when entering link l at time s, $f_t$ is the free flow travel time, $x_l(s)$ represent the number of vehicles within link l at time s and $\delta_l$ is the capacity of road link l.

The exit time as shown above is affect by the nature of the link traveled shown in the free flow travel time in addition to the amount of cars stacked within a queue. This way of stating the exit time from a link is well observed daily by the behavior of cars on major arterials in Riyadh. For example on Mekkah Road, vehicles experience delays on links that takes inflows from service links because of the vast increase of x(l(s)) with regards to the capacity of that link. Therefore, the model has to maintain the law of conserved flows, which states that for every road link, the change in the number of cars within a link is determined by the difference between cars entering and leaving a link. Therefore, the outflow for a particular link can be found as the following:

$$g_l(s) = \begin{cases} \text{inflow to link } l \text{ at time } s - f_t & \text{if } x_{l(s)} = 0 \text{ and inflow less than capacity} \\ \delta_l & \textit{otherwise} \end{cases}$$

where gl(s) represents the outflow of link l at time s. Having the ratio of vehicles being stacked within a link higher than the capacity of a link will cause the queue to extend enough to start affecting subsequent links. Therefore, the queuing model is going to consider concepts of dynamics of traffic flows.

- Conservation law: the change in the number of cars within equals the difference between the inflow and outflow of a link.
- FIFO (First In First Out): cars flowing into a link first are the ones flowing out of a link first.
- Causality order: traffic conditions are affected by events in the past rather than future.
- Flow propagation: the inflow to a link is the same as the outflow at the

associated exit time.

The queuing model with the incorporated traffic flow concepts will provide a holistic view of patterns of flows emerging around the day. It would provide insightful information about the evolution of congestions and how they form. Furthermore, it will breakdown the sources and sinks of vehicles stacked within a congestion which is something that could not have been done without the traces of mobile phones.

## 6.6 Microscopic traffic simulation models

### Traffic Simulation Models

There are a variety of transportation strategies, such as traffic management strategies or network design strategies that have the potential to alleviate congestion, its impacts and costs. The impact and performance of potential strategies are typically evaluated based on detailed traffic models. The complexity of congested urban traffic flow has lead the traffic modeling community to focus mainly on the development of simulation-based traffic models.

Microscopic simulators embed the most detailed mathematical traffic models. They represent individual vehicles and can account for vehicle-specific technologies/attributes. They also represent individual travelers and embed detailed disaggregate behavioral models that describe how these travelers make travel decisions (e.g. how they choose their travel mode, their departure time, their route; how they respond to real-time traffic information; how drivers decide to change lanes). They also provide a detailed representation of the underlying supply network (e.g. variable message signs, public transport priorities). Thus, these traffic simulators can describe in detail the interactions between vehicle performance (e.g. instantaneous energy consumption, emissions), traveler behavior and the underlying transportation infrastructure, and yield a detailed description of traffic dynamics in urban networks.

Federal, state, regional and local agencies, transit agencies and transportation consultants develop and rely on microscopic simulation tools to inform their planning and operations decisions. For instance, the Ministry of Transportation of Ontario (MTO) has used microscopic simulators to evaluate planning and operational strategies in over 100 projects [25]. Other examples of cities that resort to microscopic simulators to evaluate the performance of predetermined transportation strategies include London, New York, Boston, Stockholm and Hong Kong [20,24,27,28].

These simulators are popular tools used in practice to evaluate the performance of a set of predetermined transportation strategies. For a given strategy, they can provide accurate and detailed performance estimates. That is, their use is therefore mostly limited to what-if analysis (also called scenario-based analysis) or sensitivity analysis.

### Simulation-based optimization methods

To fully benefit from both the current and the new generation of traffic simulation tools requires frameworks where these models can be efficiently used to identify strategies, rather than to evaluate a set of predetermined strategies. This involves embedding them within simulation-based optimization

(SO) frameworks, such that novel transportation strategies that are both economically viable and environmentally sustainable can be devised. The main challenge is to enable the use of these detailed and computationally inefficient multimodal traffic simulators to solve complex transportation problems in a computationally efficient manner. Efficient SO techniques respond to the needs of transportation practitioners by allowing them to address complex problems in a practical manner.

In the optimization problems, the objective function that is to be minimized or maximized is not defined explicitly; but it is derived implicitly by the simulation model (i.e., in our case; the microscopic traffic simulator). Furthermore, such a detailed simulator is stochastic and nonlinear which yields random performance measure with no closed form and thus it is extremely inefficient to evaluate. Alternatively, the simulation-based optimization can play a vital role for addressing problems that intend to improve the performance metrics.

Simulation optimization strategies can be sorted into four main methods:

(i) Random search and metaheuristics methods that aim to choose the decision variable from a set of possible values efficiently.
(ii) Ranking and selection methods by which a vector of decision variable is specified and the number of simulation replications to evaluate it is defined to provide a pre-determined probability for selecting the best value of the design variable.
(iii) Direct-gradient methods which use the simulation response to provide gradient estimates of the decision variable.
(iv) Metamodel methods which uses an indirect-gradient approach by which the gradient is computed for a deterministic function (metamodel) rather than for the simulation response.

In this paper, the metamodel methods will be focused on addressing the traffic management problem of signal plan control.

A metamodel can be simply considered as an analytical approximation of the objective function. Because of the stochasticity of the simulation model, computational cost, time, and other constraints, it is tough to experiment optimization directly. Instead, a surrogate model (metamodel) can be used.

Existing SO algorithms that embed microscopic simulators treat the simulator as a blackbox, using no a priori structural information about the underlying transportation problem (e.g. network structure). They are therefore not designed to yield good short-term performance, i.e. they require a large number of simulated observations in order to identify strategies with improved performance, and are thus not computationally efficient.

Efficient SO frameworks that embed less efficient traffic simulators can be developed by integrating other more tractable traffic models. In this project, we develop efficient SO frameworks to solve continuous nonlinear constrained and high-dimensional transportation problems. This project focuses on the development of SO techniques with good short-term algorithmic performance (i.e. computationally efficient techniques). In particular, it develops SO frameworks that can address complex problems within a tight computational budget (e.g. a fixed computational time or a fixed number of simulation runs).

In this project we use a detailed microscopic traffic simulation model of the downtown Al Batha area, and embed it within a state-of-the-art SO technique in order to address several types of traffic management problems, including large-scale traffic signal control problems that mitigate the spatial-temporal propagation of congestion, as well as its economic impacts.

The SO framework used in this work has been developed by Osorio and Bierlaire [30] and it has proved its capability in solving a variety of challenging traffic management problems (Osorio and Chong [31]; Osorio and Nanduri [32]; Chen, Osorio, and Santos [33]. The main idea of the framework is depicted in Figure 6. Information from the simulator is coupled with information from an analytical traffic model, the combined information is then used to perform optimization. This is known in the field of simulation-based optimization (SO) as a metamodel SO approach.

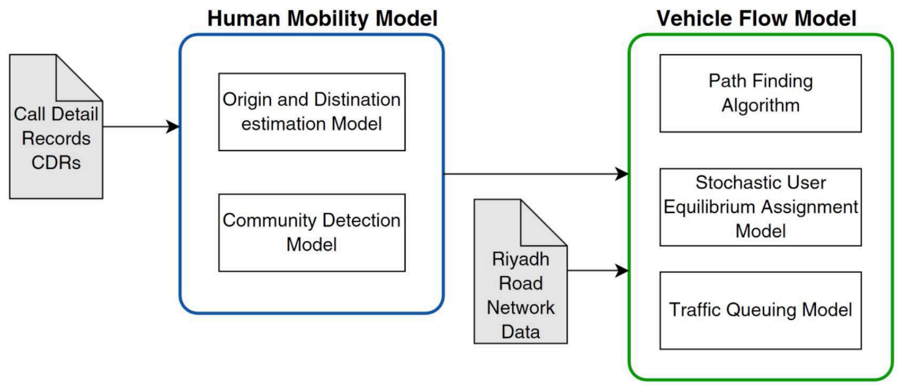



*Figure 6. Simulation-based optimization framework for solving urban traffic management problems using the microscopic traffic simulator.*

## Metamodel

Metamodels are classified as either functional or physical metamodels (see Conn et al. [34], Barton and Meckesheimer [35] or Sondergaard [36]. This project relies on the simulation-based optimization framework that was developed by Osorio and Bierlaire [30]. This framework proposes a metamodel that combines both the physical component and functional component. The functional component is represented by a quadratic polynomial which ensures asymptotic metamodel properties. A traffic model based on finite capacity queuing theory was developed by Osorio and Bierlaire [37] that is computationally efficient can be considered as the physical component of the metamodel. The parameters of the queuing model such as the space capacity, arrival rate, and the service rate have structural interpretation.

$$\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{q}) = \boldsymbol{\alpha}\mathbf{T}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{q}) + \boldsymbol{\phi}(\boldsymbol{x}; \boldsymbol{\beta})$$

Where,
*α, β* are the parameters of the metamodel.
*y* denotes the endogenous queuing variables
q denotes the exogenous queuing variables
T(x, y; q) is the approximation of the objective function provided by the queuing model
*ϕ* is a quadratic polynomial in

## Urban Traffic Network Queuing Model

Due to their complexity, microscopic traffic simulators require many parameters to be calibrated, i.e. estimate the parameters used in the simulators from measured data. Relating available data to the model parameters in a computationally efficient manner is a largely unresolved challenge. Most existing approaches to parameter calibration use black-box optimization routines, rather than exploiting the underlying structure of the problem by capturing the relationship between model parameters and the traffic data.

This project proposes a methodology of calibrating behavioral parameters from available traffic data in order to improve the quality of existing microscopic simulators. Our simulation based optimization (SO) approach combines information from the simulator with information from a macroscopic analytical model that relates traffic flow to travel time.

Improving the quality and use of existing microscopic and macroscopic models can help address urban congestion due to the low cost of implementation, i.e. updating signal control strategies, and the short-term solutions it can provide to existing transportation infrastructure.

It is the physical component of the metamodel which is differentiable and aiming to amend the computational tractability of the optimization method. This model provides closed form expressions for the performance measures and their derivatives. The model used in work was developed by Osorio and Bierlaire [30].

The queuing model is based on the finite capacity queuing theory. Each lane in the network is considered as a bounded queue that is the main element used for the calculation of the queuing parameters such as space capacity, arrival rate, service rate, and transition probabilities. The space capacity definition followed in this model is according to Heidemanne [38], Van Woensel, and Vandaele [39] as they defined the minimum length that a vehicle needs is the reciprocal of the jam density. The following expression can be used to calculate the space capacity:

$$\boldsymbol{k_i} = [\boldsymbol{l_i} + \boldsymbol{d_2}]/[\boldsymbol{d_2} + \boldsymbol{d_2}]$$

where $l_i$ denotes the length of the lane
$d_1$ is the average vehicle length, typically is 4 meters
$d_2$ is the minimum inter-vehicle distance, typically is 1 meter.

The demand of the network is illustrated by the turning flows through all intersections. Accordingly, the calibration of the queuing parameters is based on this demand throughout the network. The transition probabilities matrix give the proportion of flow goes from lane to lane . The external arrival rate is one of the exogenous parameters intended to characterize the demand throughout the network as it is defined as the rate of the vehicle hourly arriving at the queues from outside the network. The service rates of the queues are defined as the capacities of the underlying lanes. For signalized lanes, the service rate is the percent of time that a queue has a green time based on the saturation flow, which can be calculated as follows

$$\mu = \left({}^{g}/_{C}\right) s$$

where g is the green time duration for this lane in seconds
C is the cycle time duration in seconds which is the time required to complete one rotation of phases at a certain intersection
S is the saturation flow in vehicle per hour, which is typically 1800

## 6.7 Microscopic Model of Al Batha

The microscopic model that is at the core of our framework considers the downtown Al Batha area. Dornier Consulting [40] developed the model for the Municipality of Riyadh in 2008. This area is very active area (see Figure 7) of commercial importance due to concentration of numerous banks headquarters, shopping malls, governmental agencies, as well as residential areas for expatriates.

Figure 7. Al Batha Road as it seems commercially active with high associated congestion, Dornier Consulting [40].

The study area could be defined as three major zones; office/institutional civic north zone, commercial central zone, and residential/municipal south zone, see Figure 8. North of the main junction of King Saud Street, King Abdul Aziz Road and Al Batha Street is a much less dense and congested urban area made up of a series of Government buildings such as the Ministry of Education. Commercial zone starts at the junction of King Saud Street, King Abdul Aziz Road and Al Batha Street and extends to the end of the bulge forming a denser street pattern and a distinct land use made up of commercial activities mostly in the form of shops. The residential/municipal zone extends to Al Madinah Al Munauarah Road as it contains women's university, a large cemetery, and other municipal buildings.

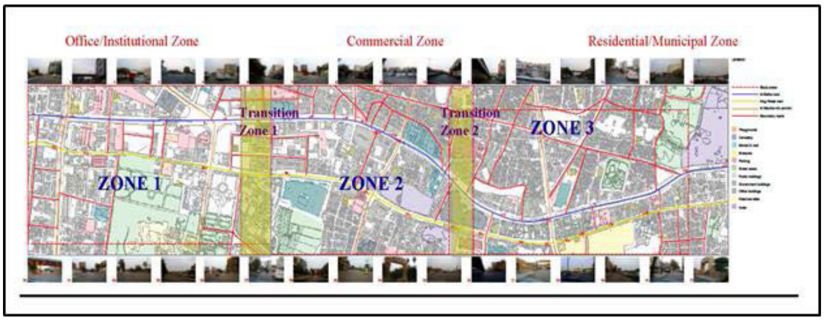



Figure 8. Three main zones of Al Batha Area [40].

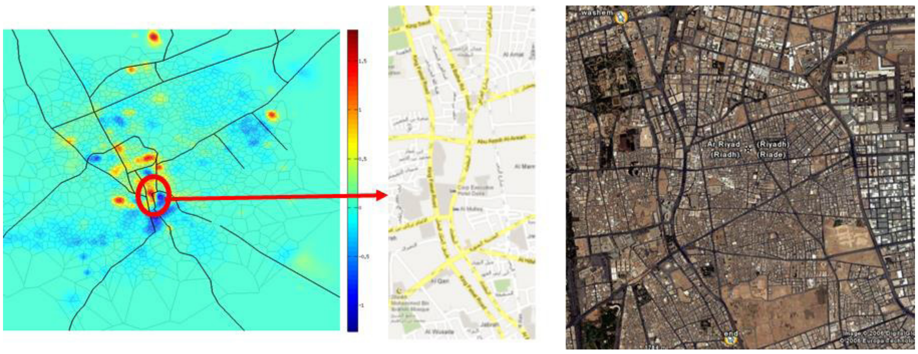


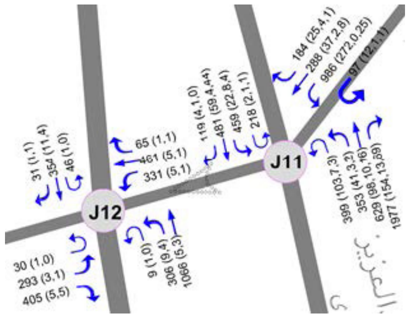
Figure 9. Al Batha network: location within Riyadh and detailed snapshot. (Google maps)

Dornier Consulting shared the microscopic traffic model with us. Snapshots of the considered network are displayed in Figure 9. There are two main arterials within this network: King Faisal Road and Al Batha Road. A total of 10 signalized intersections are accounted for, and are currently controlled by fixed-time signal controllers.

## Traffic Flow and Signal Plan Data

Traffic flow data was provided by Municipality of Riyadh MOR in consulting with Wilbur Smith Associates. This data had been collected during 2011 as the average traffic flow for the morning peak at the major intersections. It illustrates the turning flows on these intersections with vehicle classifications. The data given for the downtown area does not cover all intersections laid in that network and hence some of the turning proportions in the microscopic simulation model may be estimated.

*Figure 10. Sample of turning flow data at two intersections (Provided by Municipality of Riyadh in consulting with Wilbur Smith Associates)*

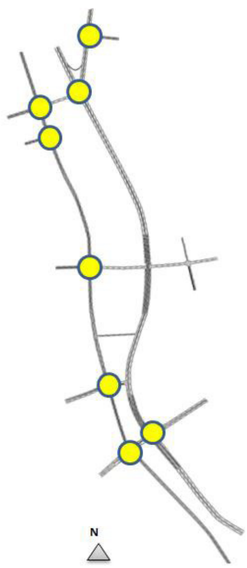



One of the main supply data within the microscopic traffic model is the signal plan data. For Riyadh downtown model, the signal plan data was given by Riyadh Traffic Police for all eight intersections in terms of the green time duration, the cycle time for each intersection, and amber/yellow time which is typically 3 seconds followed by all red phase that lasts for 3 seconds as well.

## 6.8 Traffic Signal Control Problem

The traffic management problem considered is a fixed-time traffic signal control problem. The signal plans of a number of intersections are determined jointly. The intersections that are controlled in our approach are highlighted in yellow in Figure 4. For a certain intersection and a given time interval, a fixed-time signal plan is a cyclic plan that is replicated throughout the time interval, where the cycle is defined as the time required to complete one rotation of phases. A phase is defined as any traffic signal display with its own set of timings that controls an individual vehicle movement.

*Figure 11. Subnetwork of Riyadh downtown*



In Riyadh subnetwork, the cycle times of the eight intersections range between 77 and 148 seconds. The cycle contains all-red periods, where all streams have red indications. The difference between the cycle time and the sum of the all-red times is called the available cycle time, where the ratio of the available cycle time and the cycle time is called the available cycle ratio. The decision variables of this problem consist of the green splits of the different intersections. The formulation of this problem is described in Osorio and Bierlaire [30] and Osorio and Bierlaire [37]. The following parameters can be introduced to formulate the problem:

| | |
|---|---|
| $b_i$ | available cycle ratio of intersection i |
| x(j) | green split of phase j |
| $x_L$ | vector of minimal green splits |
| $\mathscr{I}$ | set of intersection indices |
| $\mathscr{P}_I(i)$ | set of phase indices of intersection i |

The optimization problem is formulated as follows

$$\min_{x} E\left[F(x;q)\right]$$

subject to

$$\sum_{j \in \mathscr{P}_I(i)} x(j) = b_i, \forall i \in \mathscr{I}$$

$$x \geq x_L$$

The decision vector x represents the normalized green times for each phase, i.e., the green splits where the objective function is to minimize the expected value of the travel time. The available cycle time for each intersection and the green times of the phases are related using the linear constraints in the previous equations.

## 6.9 Visualizations / Sense-Making Apparatuses

The UTS project promises to unearth a massive amount of new urban data on Riyadh, and within this vast data stream lays a wealth of information. To explore, analyze and understand it, we must represent the data in an intelligent and comprehensible way. Therefore, one of the major components of the project will be its interactive visualization system. A strong system of representation is important for a wide variety of reasons: it communicates internal city dynamics to urban development, planning and policy specialists; it makes complicated urban phenomena more accessible and understandable to the public at large; and it highlights areas for further internal research and analysis. As such, a variety of visualization components have been, and continue to be produced across multiple scales. More information follows in the “Analysis and Visualization” section that follows.

# 7. Research Methodology

## 7.1 Data Acquisition

It goes without saying that data is an invaluable asset to the UTS project. However, with the complexities of transportation to the daily lives of individuals, there is no single type of data that answers all the facets of the research within the project. In fact, the team is actively seeking a diverse, and often disparate, suite of data. Various datasets have been obtained, and much is still in progress of being obtained:

- Several types of archived and static data have been already obtained including GIS maps of the street network, car counts at select locations within the city, planning maps, cell tower locations and census data.
- The team has acquired CDR data from STC, covering the entire month of December. Due to its size , a number of data management and organization techniques have been used to prepare it for various analyses.
- A microscopic traffic model has been obtained from Dorier Consulting, and is currently being analyzed and adapted for use within the project.

## 7.2 Stakeholder Meetings

The UTS team has benefitted from many fruitful meetings and interactions with the various stakeholders involved with transportation planning and operations. The team has also completed a survey across the government to understand the roles and tasks of each Ministry and agency with regard to their current transportation related practices. The survey sought to understand what problems each addressed in the short, medium and long-terms and how each contributed to the planning and implementation of mobility plans. This was implemented through a web survey.

Meetings have also taken place with members of agencies and ministries to identify the challenges facing Riyadh, and the policy priorities of these decision makers. Meetings and data sharing with the ADA have provided insight into priorities of urban planning, with special consideration for the future implementation of a public transportation system. Tours of the Riyadh transportation control center by the Riyadh Traffic Police offered insight to current practices by the agency in managing the city’s transportation network and the larger ambitions of technologies implemented by the SAHER system. Ongoing dialogue offers both insight and opportunities for real-world intervention by the project.

Conversations with a major telecommunications company have also been on going, both in person by the members in Riyadh and via Skype with the entire team. In addition to discussions regarding access to CDR information, conversations have also concerned how expertise may be shared in conveying the public findings of the research to the public at large.

We initially had several meetings with MOR and they provided us with some insight about the traffic management problems that they had handled to consultants. For instance, Dornier consultants developed a microscopic traffic model that was implemented in VISSIM for the purposes of beautification the downtown area and experimenting different designs of the roads. In addition, Wilbur Smith Associates have developed a macroscopic traffic model for the city of Riyadh to project the travel demand for different modes as some new projects and subcenters have been planned and to be built in the near future. This model was developed by TransCad that embeds a geographical information system to ease dealing with traffic counts locations and enhance the model performance for setting different scenarios.

Moreover, MOR had supplied us with essential documents and data needed to develop our own microscopic model including the subnetwork characteristics, roads and number of lanes, controlled intersections, and the traffic demand in terms of the turning flow at intersections.

We had been supplied with signal plan data by Riyadh Traffic Department in terms of the green time duration and cycle length for all the controlled intersections. Since the sequence of phases were not give, we had to spent some time in the field to check the sequence and make sure we model it correctly. Additionally, we inspected the allowable movements associated with lanes at all intersections.

## 7.3 Analysis and Visualization

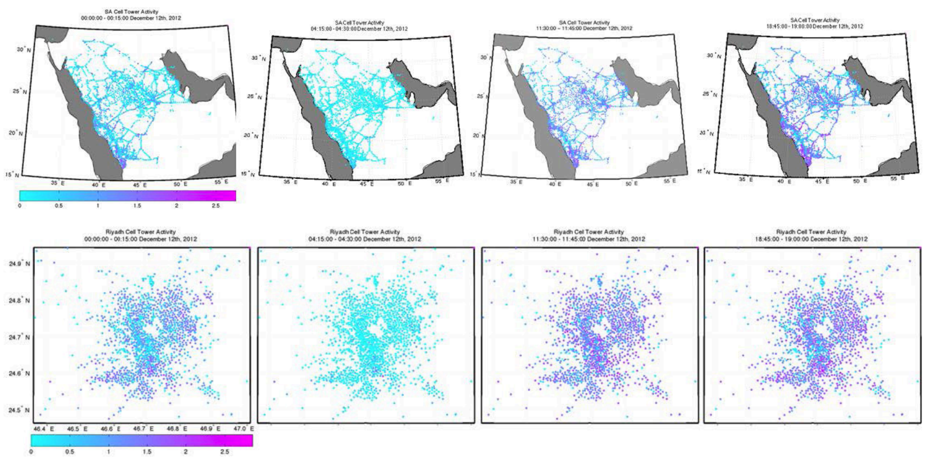
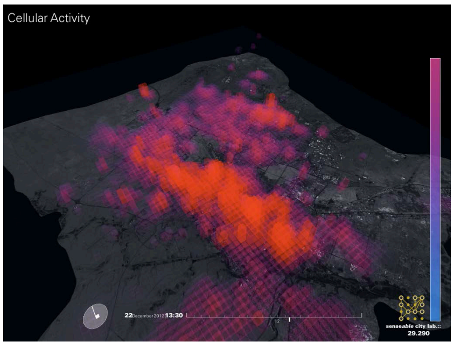


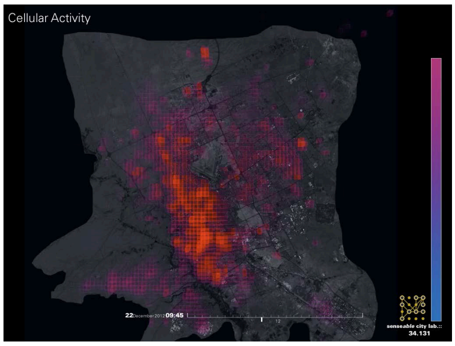
*Figure 12. Activity Snapshots across Riyadh and Saudi Arabia*

### Searching for the Social Pulse of the City

The first step in seeing and understanding our data, is in spatializing the activity patterns at different physical and temporal scales. The images above show basic mobile phone activity patterns over one day across (1) the city of Riyadh, and (2) all of Saudi Arabia. The aggregate cellular activity (number of calls, texts, and data queries) over an interval of 15 minutes is plotted by color (blue to pink on a logarithmic scale) for each cell tower. These static images are powerful, but they give us nothing about the rhythm or pace of life. Bringing more context to the forecity imagery, infrastructural form, and temporal dimensionality we arrive at a much richer representation of variation across the city.

*Figure 13. Activity Snapshots across Riyadh and Saudi Arabia*




The above visualization shows cellular activity through color, transparency, and height (again in logarithmic scale) gridded across the metropolitan expanse of Riyadh. As opposed to seeing the cell towers as discrete points in the city, we show network traffic interpolated over a 100 by 100 grid. In this sense, each grid cell is assigned an intensity based on its distance to surrounding antennas and their activity levels using a gaussian smoothing function. The temporal activity is interpolated in a similar manner, showing smooth transitions between each time-slice in the dataset.

This representation serves to illustrate Riyadh's social rhythm alongside its physical construction at a macro scale. With the inclusion of satellite imagery of the city as a base map, we arrive at a unique view of how activity is expressed over built urban form. As one would anticipate, the city's downtown core quickly becomes clouded in a smog of network activity early in the morning that hangs over region for the entire day. Clear subcenters emerge that follow construction density, and these subcenters appear to be partitioned by the roadway network itself.

The city's shifting activity profile also highlights a rich temporal signature of communication that is all Riyadh's own. Watching the oscillations of the activity landscape, a unique character emergeswe see that Riyadh really doesn't come alive before noon, with a peak in aggregate activity around 6:15. We also see strong regional delineation: the residential neighborhoods to the southwest and northeast of downtown core come alive well before the rest of the city, and experience the strongest inter hour fluctuations throughout the course of the day. Finally, we see some peculiar discontinuities in aggregate talk throughout the dayalmost as if all phone traffic was suddenly halved at strange intervals we'll return to this phenomenon later.

## Inferring Home/Work Locations

Expanding our time intervals to capture broader day and night variation we can begin to differentiate home and work locations over the citya methodological precursor to inferring urban land use. We define home locations as the most visited cell tower areas during weekday nighttime hours, and work locations as the most visited cell tower areas during weekday daytime hours. This essentially breaks down to filtering users who make the majority (60%) of nighttime calls in one place between the hours of 10:00 pm and 6:00 am, and the majority of daytime calls between 9:00am and 3:00pm in another place.

The process left us with approximately 2 million weekday home-work pairs, but the question became: how can we put these to use? Can the combination of these pairs tell us something new about the operational structure of city? Of course, each home-work dyad by its very nature defines the start and endpoint of a commute, making the process a fundamental step towards understanding travel demand, but is it possible to see something unique in the data?

We began by accumulating each home and work location bounded by the expanse of the city and geographically smoothing the results (shown below in two separate maps). These images show a very course-grained view of land use in two discrete dimensions. We then created an additional map to highlight the extremes, subtracting resident locations from work locations, shown in figure 3.

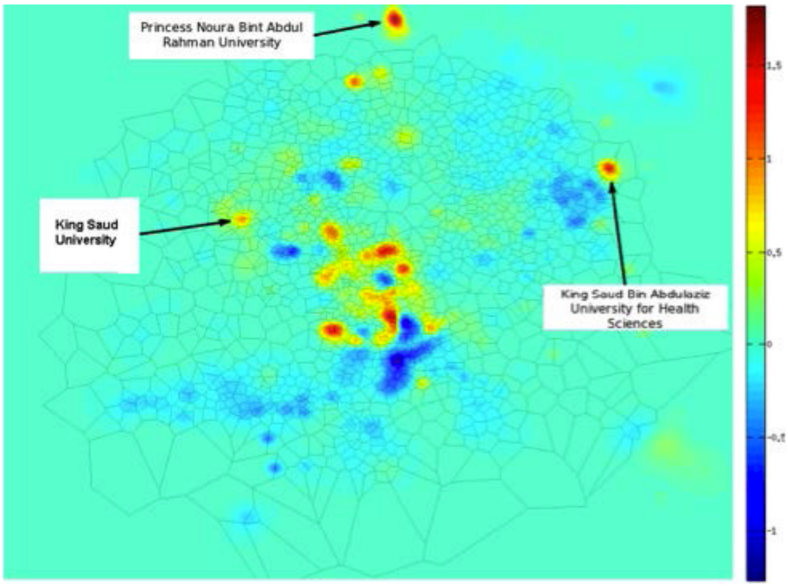



*Figure 14. Activity Snapshots across Riyadh and Saudi Arabia*

The map highlights the discrepancy between the purely residential and the purely worksite location, showing some mono-centrically clustered worker hotspots that follow the overall spatial logic of the city. At the periphery we also see a number of universities show up strongly as work locations. Lastly, we see high agglomerations of residences to both the south and east of the city, with smaller pockets scattered throughout. Again, this is in line with a subjective survey of the land through google maps and discussions with locals.

## Population Analysis

A detailed analysis of mobile communication data can be a very powerful tool for learning about complex systems centered on the activity of people. Archive data, spanning lengthy periods of time, allows one to extract accurate information about fields as diverse as human mobility [10], space occupation patterns [24], community mining [17], commuting behavior [14], and social networking [13], and to contribute to the ripening complex systems science in unexpected ways [25, 24].

Real-time communication data, on the other hand, is able to provide powerful and accurate input to Smart Cities' processing systems, with possible applications to congestion forecasting and exceptional events detection.

Currently, in-depth research is being performed to develop a powerful and reliable framework to obtain instantaneous population estimates from CDRs, in addition to exploiting origin-destination flows to sample instantaneous actual commuting patterns of citizens [27, 28], keeping in mind possible applications to housing planning.

The population estimates capture a very different picture than census data, as they reflect how a region is used by all of the country's population at any given time, including random shoppers, workers officially employed elsewhere, tourists, etc. The origin-destination flows complete the frame by indicating critical trajectories that determine, among other things, traffic congestion spots, and devising ad hoc solutions.

Obtaining reliable population estimates is a delicate process that needs to take into account disparate factors, including the uneven distribution of mobile phone coverage in different areas and the non-linear increase of telephonic traffic per person in more densely populated areas. Preliminary results are highly correlated to actual population figures, and bode well for the applicability of the method to Riyadh's urban environment, and consequently for its use as a powerful tool for planning.

The identification of regions, which attract a large share of people, comprises the following steps:

- The whole data set (CDRs) is scanned, and for each ID the cell tower which is most frequently visited in the working hours (9 AM to 5 PM to be definite) becomes the home location;
- To give an estimate for population, each user i is weighed by a factor 1/ci, where ci is the coverage of the home location of the user; this is to keep into account that users from areas underserved by the mobile phone provider actually represent more human beings than users from areas well served. A few simple devices have to be employed in order not to over-weigh cells with too low a coverage;
- The values so obtained for each cell can be rescaled to the overall population, to yield a local estimate of working population;
- For each cell tower area, one can obtain the local activity as the ratio between the working population and the night population.

The activity so obtained is a strong indicator of how attractive an area is. Given the fine-grained presence of mobile phone towers in large cities, the activity can draw a precise layout of the actual city center, or centers.

This method does in fact systematically highlight urban centers (see Figure 3), and for large cities it gives strong indications as to where the most productive zones lay (see Figure 4 and Figure 5).

## Land Use Inference Based on Human Activity

Building off of the home/work and population analyses, UTS will begin to develop complex estimates of land use patterns across the city of Riyadh. As demonstrated in literature, the bandwidth pattern of a particular cell tower holds a strong correlation with the area's land use [25, 26]. Thus, UTS plans to use the data of mobile phone usage for cell towers to identify land use in the different areas in Riyadh. A snapshot of the heat map representing the utilization of mobile network at a particular time is shown in Figure 10. The heat map has a temporal dimension as usage of mobile phones changes throughout the day. Areas with similar mobile phone usage distributions have land use parameters in common [10]. UTS will seek to categorize areas based on land use at a given point in time.

The process of identifying land use will integrate with the generated OD matrix to identify trip purposes [11]. This is important because flows of people correlate with the type of activities within a destination area at a given time. Identifying trip purposes will help in providing semantics to the output OD matrices that are fed into the vehicle flow model, where flows of cars can be distinguished by purpose.

## Detecting Mobility Communities

The home/work visualizations above point to an organizational logic of the city. Can empirical analysis confirm the strong regional clusters we see in the preceding maps? And if so, how can we visually explore the implications of this inherent structure? If we conceptualize the totality of home/work commutes as a city-wide mobility network, we can conceivably break this network into sub-communities by applying a regional delineation algorithm.

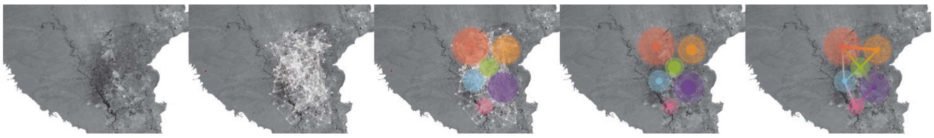


*Figure 15. Mobility Community Detection Process*

The process begins with the city-wide network of connected cell tower locations, where a weighted directed edge between two nodes is defined as the cumulative trip flow between them. The algorithm then uses a modularity optimization scheme, such that subnetworks are clustered in a way that minimizes internal arc disruption. Each resulting sub-community represents an area where the majority of commuters live and work. In total, the algorithm found to 17 distinct communities in Riyadh.

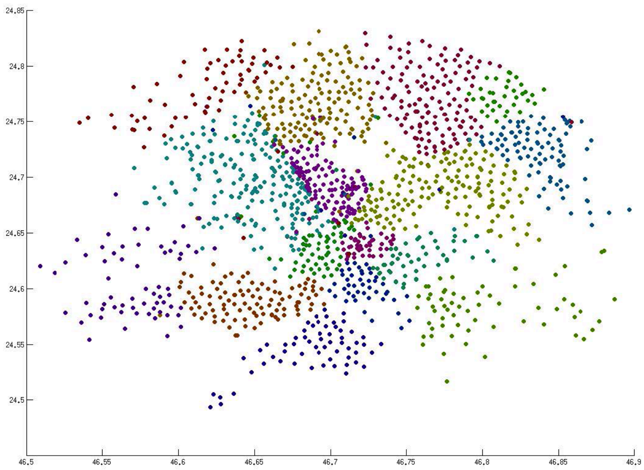



*Figure 16. Mobility Community Detection Process*

By overlaying this data on geography of the city, a number of interesting relationships are revealed between the detected communities and the built form of the city. Most strikingly, the resulting clusters closely correlate to the main arterials of city's roadway infrastructure. Mobility communities seem to be partitioned by the street network itself, underscoring the city's dependence on highway infrastructure, while also supporting the commonly held belief that heavily trafficked streets, on many levels, are instruments of segregation and control, or, perhaps more optimistically: good streets make good neighbors.

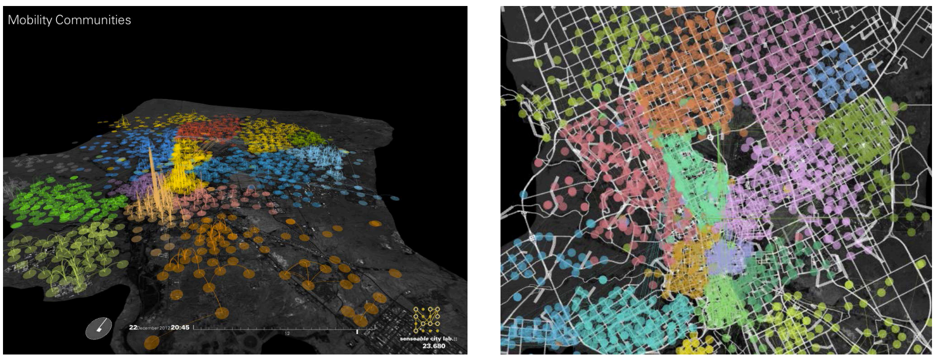



*Figure 17. Visualization Output*

## Transforming Cellular Network Traffic to Movement and Flow

Can CDRs tell the stories of urban inhabitants in real time? Is it possible to create a real-time OD census? Origin Destination (OD) matrices are one of the most vital facets of a strategic transportation planning or management initiative. Constructing accurate, precise OD matrices is a crucial component for transportation network optimizationnot only for individuals' daily commutes, but also for the forecasting and modeling of the city's future needs. Traditionally, ODs are constructed on census surveys that are conducted every 5 to 10 years. However the process is long and costly, and when completed, it only provides a basic snapshot of overall travel demand.

While some have proposed the installation of exhaustive sensor networks to bypass these inadequacies, our approach is to leverage the ubiquity of the sensing apparatus already in our environments, namely cell phones. By collecting and filtering each user's mobile activity as sequence of cell tower locations, we are able to estimate a populations travel demand in terms of origins and destinations of individual trips. We've shown that these approximated OD flows hold a strong correlation to Census estimates, however this approach includes the added benefit of capturing and travel demand at highly dynamic time slices ranging from seasonal variations to hourly fluctuations. Such a high temporal resolution has the potential to transform our understanding of urban mobility.

### The process of estimating OD flows over a particular time slice:

The process of estimating the aggregate flows of people across the city from the CDRs is a three step process that has the CDRs as inputs and the aggregation of flows of people between locations at every time step t as its result (i.e. Origin Destination matrix). The process starts by arranging data on a user level and considering each location transition as a potential trip. After that, the resulting potential trips go under a filtration process that filters out noise in the data from the potential trips generated. Finally the last step aggregate the resulting trips on both the spatial and time dimensions to generate what is called Origin Destination matrices based on the provided time slice of interest.

The first step in the process looks at phone activities on a user level and gathers all activities generated for each user in an iterative manner. On every iteration, we merge consecutive records pertaining to the same user into pairs of location records with their associated time of activity for further filtering. Because the generated couples of location records pertaining to the same user do not always capture the trips generated by the user in a precise manner, we apply further filtering on the results.

The goal of the filtering process is to eliminate all captured pair of location records that are considered as noise in terms of trips capturing. The filtration process eliminates all records that are considered as localization error, have very long time intervals or no movement detected. Localization error occurs when a user is served by close by cell towers where no movement actually occurred. Entries in the data that corresponds to localization error are filtered out by eliminating all trips that are less than 1km in distance between cell towers; the choice of 1km as a cut off for trips distances was due to the fact that the maximum distance between neighboring cell towers in the city of Riyadh is around 0.5km. In addition to that, each pair of records having a time difference between them that is more than 2 hours is filtered out of the data for the purpose of reducing the uncertainty in capturing the actual departure and arrival times for trips. Finally, the filtration process eliminates all pairs of location records that correspond to the same towers because they indicate no movement.

The result of the filtering process is going to constitute of all pairs of locations where movement was detected and reasonable time duration for the trip is captured. After that, the final step towards the generation of OD matrices is to aggregate the trips according to the specified time slice. For every time window, we aggregate the all the flows that have their starting and arrival locations in common. Doing that iteratively on every time window produces a dynamic OD matrices that correspond to different times.

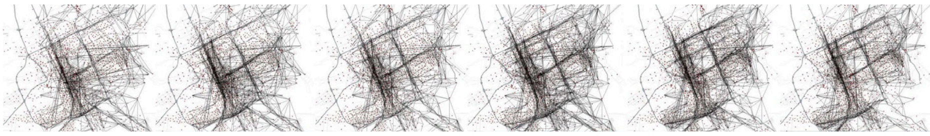


*Figure 18. OD Matrices across Riyadh from 6am to 9am, 9am to 12pm, 12pm to 3pm, 3pm to 6pm, 6pm to 9pm, and 9pm to 12am*

We first constructed OD matrices on an hourly basis in an effort to show them alongside the network activity. In visualizing the result, we represent each "trip" as an arc that rose from originating to terminating cell tower. Each arc embodies a variable number of trips, and to illustrate this we altered its thickness and height in correspondence to the intensity of activity along that route (on a logarithmic scale). To further highlight directionality, a color scheme has been applied that shows origins in blue and destinations in green. The OD arcs are drawn over the same city base geography, on top of the social interaction mesh from above, in an effort to reveal unseen connections between the two datasets.

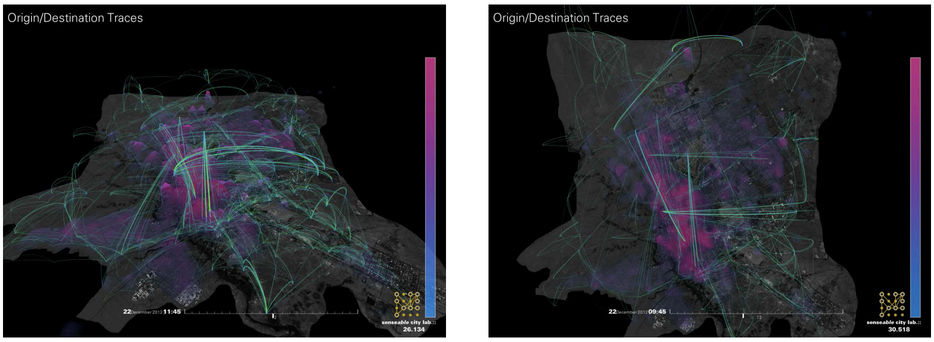



*Figure 19. OD Visualization*

The resulting dynamic maps held a striking similarity to the local intuition of vehicular flows across the city. Overall OD flows correspond quite closely to the underlying street network. Most notably, the visualized results show intense activity along the city's main arterials; King Fahd Road and the Northern and Eastern Ring roads. This is agrees with the local community's subjective understanding of commute patterns across the city, but to further validate our results, we compared them against the best ground truth measurements of roadway activity we: car count volumes captured by pneumatic-tube sensors placed at multiple intersection across the city.

These counts were built into the visualization as half-spheres place at their respective intersections. Each sphere changes shape and color at an hourly rhythm in line with the measured volume. Again, the main sections of the city line up quite nicely, however, some intriguing OD activity can be seen to the southeast of the city center that, unfortunately, has no corresponding car count figures to compare against. This remains an open area of exploration for the future.

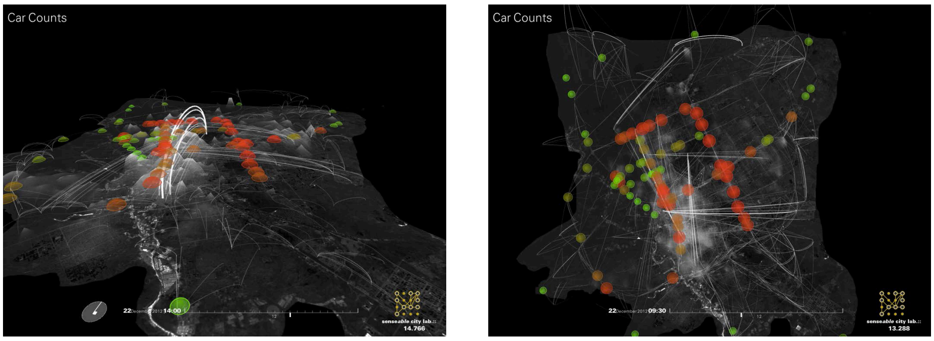



*Figure 20. ODs with car count data*

The final step of this line of analysis will be transforming these meso-scale commuting flows to activity on the road network itself. By probabilistically transposing our collected OD trips to a detailed Geographic Information System (GIS) database of road segments, we are able to explore the impact various mobility communities have on the transportation network. Demonstrating the potential for an exhaustive real time representation of commuting. A tool such as this has the ability to completely transform infrastructure planning, by acutely pinpointing the demand of particular neighborhoods at an aggregate city scale. The technique also has the power to quantitatively identify overburdened arterials, along with the specific congestion points that initiate the city's daily gridlock. Similar studies have found that its just a small number of drivers from a small number of neighborhoods who are responsible for tying up the key roads [29].

## Prayer Time Disruption and Mobility

Prayers affect in a striking way communication patterns in the Kingdom. The communication intensity sharply drops, and arguably prayer times coincide with an average increase in the duration of telephone calls. We advance no hypothesis as to the reason for this behavior.

Connections between prayers and mobility are very hard to come by. We have next to no data supporting the expectation that prayers have some effect on how people move. The negative correlation we can see between number of trips and their average duration and distance is in all likelihood a trivial effect of the decreased communication.

Even if an actual correlation exists, it is unlikely we'll be able to observe it with the data granularity we have.

### Data Source

We are using telephone data provided by a large Saudi Arabic telecom company. All the analysis below refers to data extracted for the sole day of December 10, 2013, within 20 km of the centre of Riyadh.

### Prayers and Communication

At and around prayer times, the intensity of communication decreases abruptly, as show in the figure below.

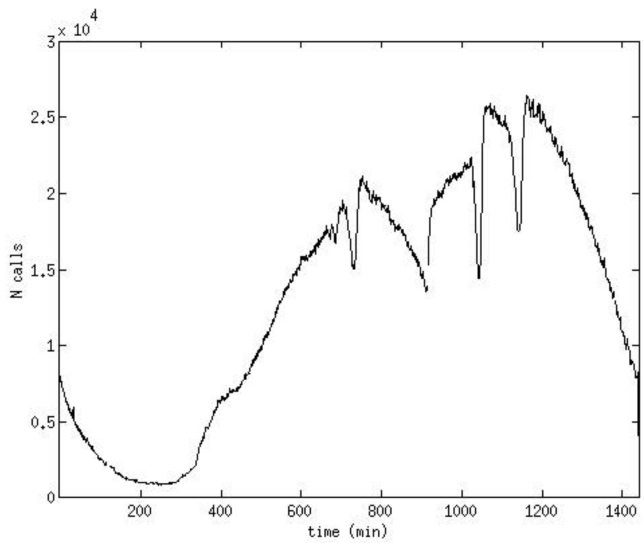



Figure 21.

Two main factors contribute to this behavior:

- Muslims involved in prayers neither make nor receive phone calls;
- commercial activities are closed during the prayers.

Even communications involving non-muslims is thus affected.

We have tried to assess subtler effects of prayer times on communication patterns, which have proved to be hard to disentangle from communication noise.

We were able to observe a small variation in the average duration of phone calls at prayer times. It is shady, however, that average call duration often increases until the end of the communication decrease, and steeply decreases immediately after (as can be seen in the close-ups below).

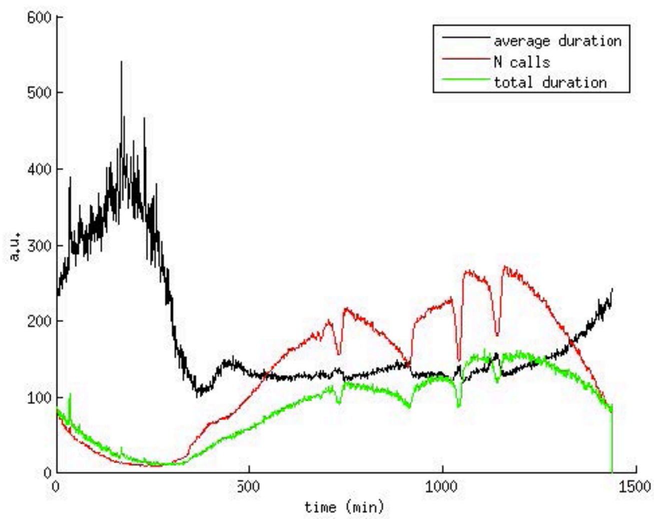
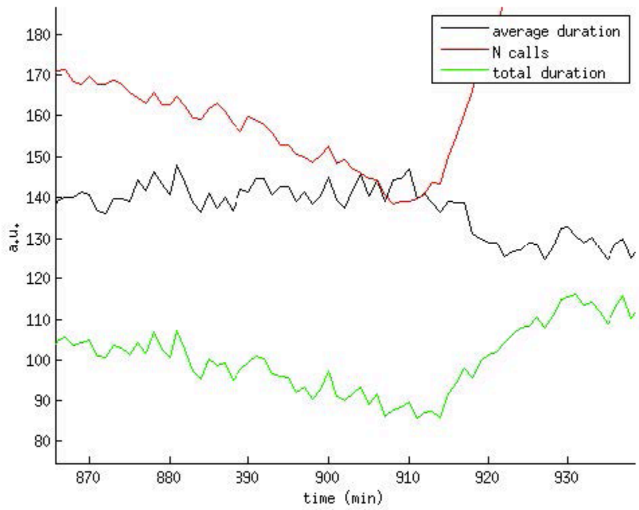


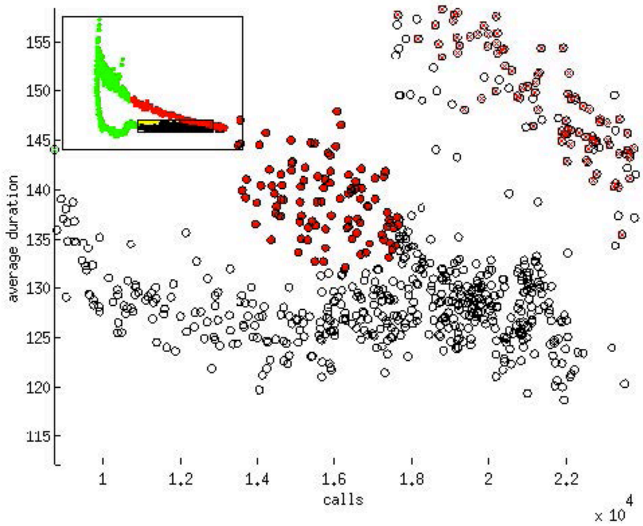

Figure 22.

Figure 23.




By looking at the relation between number of calls and their average duration at each hour (as shown below), one can surmise that prayer times are qualitatively different from all other moments, thus supporting the hypothesis that the variation of average duration is not a mere effect of noise.

Figure 24.




## Trip Detection: Method.

Method. To evaluate the effect of prayers on mobility, we estimated trips made by individuals, looking for variations in their statistical distributions astride prayer times.

One trip is made by a set of successive calls, made by the same user, such that:

(1) in each segment, the computed average speed lies a minimum and maximum value;
(2) two consecutive calls are separated by no more than a set time interval.

We recorded the number of minutes after midnight at the start of the trip, its duration, beginning-to-end distance, and number of legs.

Note: this definition of trip is of course arbitrary. The introduction of minima and maxima for speed and inter-call intervals serves the purpose of keeping the uncertainty about starting time and ending time of each trip as small as possible. However, it is hard to balance the interplay of these factors and, by pushing for more sharply defined starting times, one incurs worse sampling of real trips, which are then divided into shorter tracts.

### Trip Detection: Results.

First of all, no effect sizable effect of prayers on the number of trips is evident; the few traces visible are of the same order of magnitude as the noise.The analysis of the distribution of distances didn't highlight any effect either, probably because of systematic sampling biases. A study of local mobility, measured as the ratio of the traveled distance to travel time, didn't show any discernible patterns.

More research is needed to find an angle which could highlight such effects. Several sources of bias in the data are not controlled for, and could explain the totality of the deviations from regular behaviour seen at prayer times.

All in all, no conclusion can be drawn at the moment about differences in the characteristics of trips astride prayer times.

## 7.4 Signal Plan Optimization Results and Analysis

The Al Batha network model was implemented with the AIMSUN traffic simulation software [41]. The model network consists of 88 roads and 13 intersections as shown in Figure 13. There are 222 lanes, 8 of which are signalized. The total number of phases is 28 which are considered variables, i.e., the dimension of the decision vector is 28. The trust region problem consists of 694 variables with their corresponding lower bound constraints, 452 nonlinear equality constraints, 222 linear equality constraints and 1 nonlinear inequality constraint.

In the implementation of the SO framework, we consider three different initial points, an existing signal plan that is used for the Al Batha area, and two randomly drawn signal plans. Each run for each initial signal plan will be subjected to a tight computational budget, i.e., the maximum number of simulation runs that can be carried out is set to 150 replications.

In order to evaluate the performance of a new proposed signal plan, 50 replications of the simulation model are conducted to simulate in more details the influence of that signal plan on the network average travel time. The empirical cumulative distribution function (cdf) of the average travel time over these 50 replications is plotted for each signal plan. For each plot, three curves represent the cdf of the average travel time in the network according to: the initial plan which is plotted in blue, the signal plan identified by using a macroscopic traffic model (red curve), and the signal plan proposed by our micro approach (black curve).

Figure 14 shows the corresponding cdf curves, when using the existing Al Batha signal plan as the initial plan. The average trip travel time for the initial plan ranges between 7.3 to 8.4 minutes and the plan proposed by a macro traffic model gives similar results where the average travel times range between 7.1 and 8 minutes. The signal plan proposed by our approach shows a significant enhancement as the average network travel time is now of the order of 5.2 – 6 minutes.

Figure 14 considers the plans proposed by each method, when initializing the algorithms with uniformly drawn signal plans. The right plot of Figure 14 leads to signal plans with similar trend as for Figure 13, i.e. the micro approach significantly outperforms the macro approach and outperforms the initial signal plan. The left plot of Figure 14, the macro approach leads to a signal plan with improved performance when compared to the initial plan, and with similar but still worse performance than the plan of the micro approach.

*Figure 25. Empirical cdf's of the average travel times where the existing signal plan is considered as initial plan*

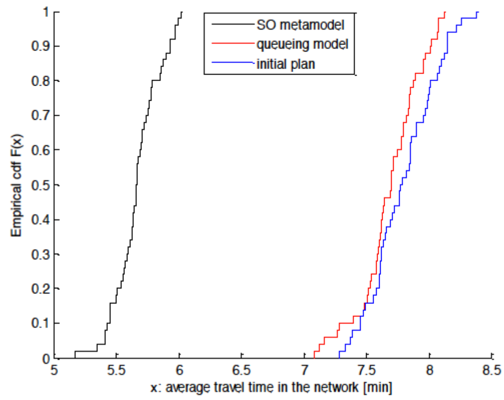



*Figure 26. Empirical cdf's of the average travel times where the initial plan is drawn uniformly from the feasible region*

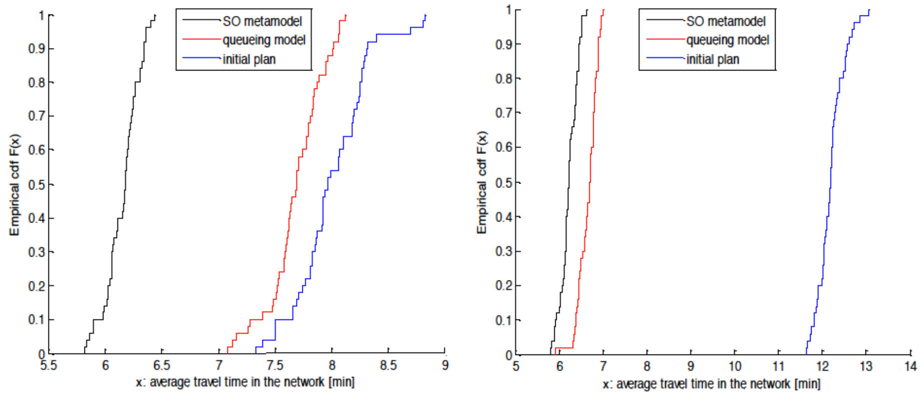



*Figure 27. Average link travel times using the initial signal plan (left) and the signal plan proposed by the micro approach (right), where the averages are taken over 50 simulation replications.*


Figure 15 displays the link-level performance of the existing Al Batha signal plan (left) and of the signal plan proposed by the micro approach. The maps color the links according to the average travel time per vehicle for each link. The average is obtained over 50 simulation replications. The green color indicates an average link travel time of less than 100 seconds, yellow indicates values between 100-200 seconds, and red refers to values greater than 200 seconds. This figure shows that the proposed approach leads not only to trip travel time improvements, but also to improvements of link travel time.

## 7.5 Traffic Management Methodologies

### Novel methods to mitigate urban congestion

As part of this research we have also formulated novel methods to address complex urban transportation problems: 1) we have proposed a simulation-based optimization (SO) algorithm that can tackle problems that are considered complex and of large-scale both in the field of SO and in the field of traffic control; 2) we have proposed a macroscopic analytical model that describes the between-link dependencies of an urban network, we have used it to illustrate the added value of accounting for between-link dependency for traffic operations.

### Large-scale Simulation-based Optimization

In Osorio and Chong (2012), we present a detailed mathematical formulation of a computationally efficient simulation based optimization (SO) algorithm suitable to address large-scale generally constrained urban transportation problems. The algorithm is based on a novel metamodel formulation.

We embed the metamodel within a derivative-free trust region algorithm and evaluate the performance of this SO approach considering tight computational budgets. We address a network-wide traffic signal control problem using a calibrated microscopic simulation model of evening peak period traffic of a network with over 600 links and 200 ntersections. We control 99 signal phases of 17 intersections distributed throughout the entire network. This SO problem is a high-dimensional nonlinear constrained problem. It is considered large-scale and complex in the fields of derivative-free optimization, traffic signal optimization and simulation-based optimization. We compare the performance of the proposed metamodel method to that of a traditional metamodel method. The proposed method systematically and efficiently identifies signal plans with improved average city-wide travel times. Ongoing work focuses on the formulation of a dynamic SO technique suitable for the real-time control of large-scale networks.

### Using between-link dependency information to improve the control of transportation networks

In Osorio and Wang (submitted), we present an analytical probabilistic model of urban traffic, that approximates higher-order (i.e., beyond first-order) distributional information of the main between-link dependencies. The traffic network is modeled as a Markovian finite capacity queueing network. The main challenge to such an approach remains the dimensionality of the joint qeueue-length distribution, which is exponential in the number of queues (i.e. roads). We have proposed an analytical approximation of the joint distribution with a dimension that is linear in the number of queues. The method decomposes the network into overlapping subnetworks. The state of each subnetwork is described aggregately, i.e. in terms of a reduced state space, while ensuring consistency with the disaggregate, i.e., full state space, distribution. This aggregation-disaggregation technique is proposed for the analysis of Markovian tandem finite capacity queueing networks. The model is validated. We present its use to address an urban traffic control problem, and show the added value of accounting for higher-order spatial between-queue dependency information in the control of congested networks. As part of ongoing work, this methodology is being applied to identify enhanced signal plans for the Al Batha network.

*Figure 28. Total Network Activity Viewer*

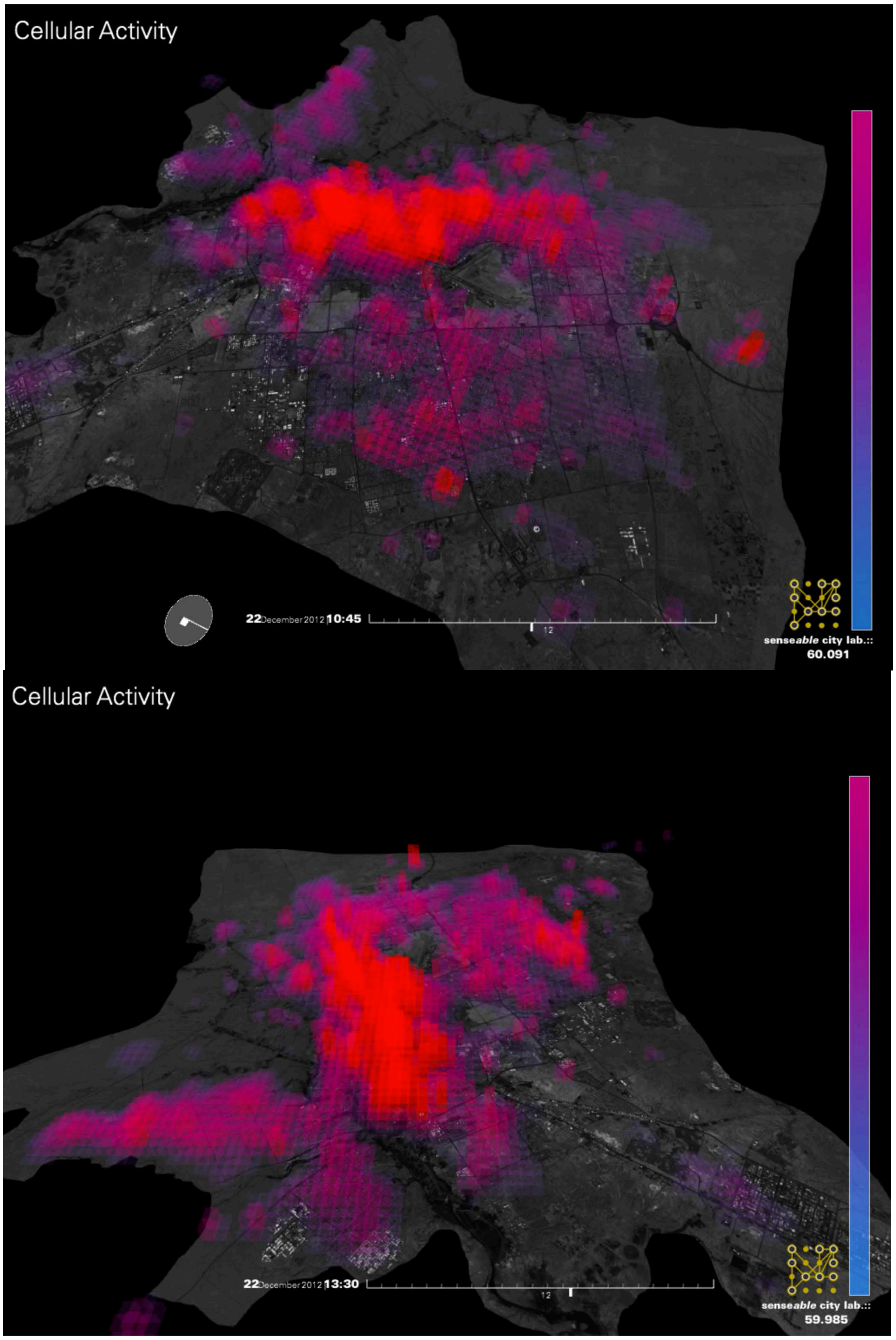



*Figure 29. Origin Destination Viewer*

Figure 30. ODs and Car Count Volumes

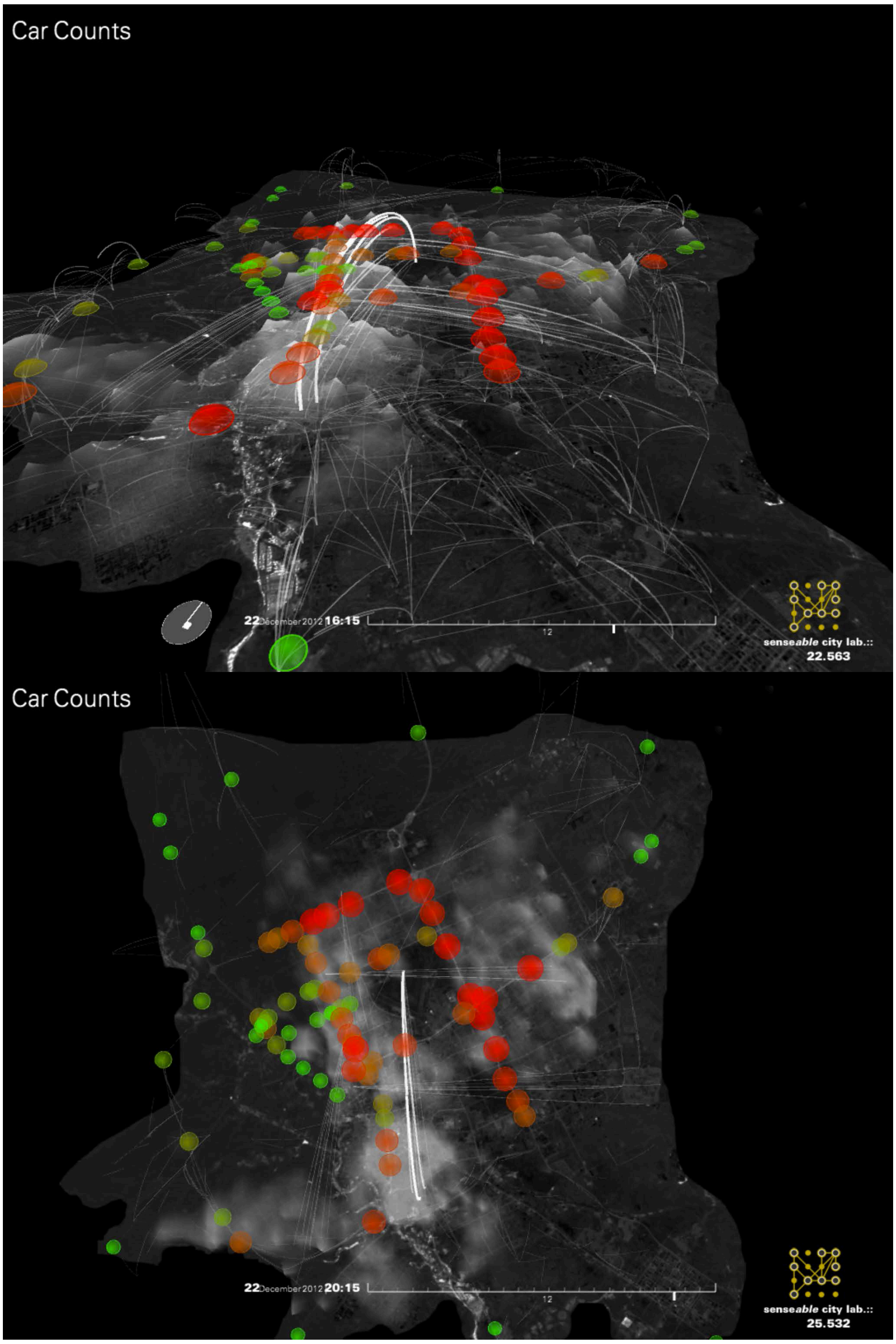



*Figure 31. Mobility Communities*

# 9. Future Directions

## 9.1 Human Mobility Scale

### Continued Analysis

- Acquirement of new data sources [on-going]: we plan to continue coordination with STC in the hopes of growing our CDR database, both in terms of the overall timespan captured, as well as the resolution of each record's meta information (e.g. including user demographics, increasing service description attributes, etc.). We are also beginning talks with Mobily, Saudi Arabia's second Telecommunications company, to increase our sample population for Riyadh.
- Transformation of OD matrices to congestion measures on the actual roadway network.
- Continued work on land-use inference analysis.
- Continued work on land-use inference analysis.

### New Analyses

- Spatial Clustering studies, a la the techniques proposed by Schlapfer et al. in "Human Spatial Organization and the Emergence of Urban Centers" (working paper) [March, 2014]
- Dynamic population, Saudi Arabia [March, 2014]
- Dynamic evaluation of urban planning initiatives [November, 2014]
- Advanced population analyses (toward public transport, etc.) [March, 2014 on-going]
- Demographic implications on mobility. [March, 2014]

### Data Browser Platform

- Fully-featured data browser control interface [March, 2014]
- Integration of complete mobility dataset, with flexible data loading back-end [September, 2013]
- Development of new representational components that reflect new analyses [March, 2013]
- Integration of new, contextual datasets (weather, demographics, etc.) [September, 2013]
- User-centric visualization resources [March, 2014]

## 9.2 Micro-Scale Analyses

### Micro-models Analysis/Adaptation

- Completion of the micro model translation and validation [September, 2012]
- Integration of the micro model within the SO framework. [September, 2012]
- Use of the micro model to solve complex large-scale traffic signal control problems for the Al Batha area. [on-going]
- Evaluate the potential improvements in terms of traditional traffic metrics

(e.g. travel times, speeds), yet also novel metrics (e.g. energy consumption, travel time reliability). [March, 2013]

### Simulation-based optimization techniques

- Development of dynamic SO techniques that can be used to solve traffic-responsive traffic management problems [March, 2013]
- Use of this dynamic SO technique to identify suitable traffic management strategies for the Al Batha area. [March, 2013]
- Development of efficient calibration techniques that enable the development of more accurate microscopic simulators for Riyadh. [on-going]

### Mitigation of spatio-temporal propagation of congestion

- Address a traffic management problem for the Al Batha area that accounts for detailed between-link spatial dependencies. Investigate the potential of this approach to mitigate both local and network-wide congestion impacts. [on-going]

# 10. Project Deliverables

## 10.1 Conference papers

Several conference papers are currently in various stages of writing, editing and submission. These include topics of network partitioning and community detection; contributions to urban visualization theories; large-scale SO; dynamic SO; spatio-temporal between link dependencies and their potential to contribute to mitigating congestion; detailed analysis of the potential of various traffic management strategies to mitigate congestion as well as its negative economic, energy and environmental impacts in the Al Batha area; and microscopic calibration techniques.

Additionally, a paper on the City Mobility Browser is in the process of being submitted. The paper presents the aforementioned data browser as a universal tool to aid in the understanding of human mobility at the urban scale. The browser framework operationalizes UTS research into a powerful tool to provide city development, planning, and policy professionals insights into the commuting patterns of a city, using mobile phone activity as a proxy for human travel. The paper describes the system's architecture, internal algorithms, and visualization features. The algorithms and processes utilize Call Detail Records (CDRs) to unearth insights pertaining to population distribution, trip directionality, home versus work locations and mobility communities. The visualization interface is used to provide a comprehensive view of the results of the processes. Finally, the paper presents the work in Riyadh as a case study.

## 10.2 Journal papers

In addition to the above conference papers, two papers related to the network partitioning strategies employed in the UTS research initiative are currently being prepared for submission. The first, "A General Optimization Technique for High Quality Community Detection in Complex Networks," by Sobolevsky et al., details the network partitioning method put to use in our study, showing an effective general search strategy for the optimization of various objective functions for community detection purposes. When applied to modularity, on both real-world and synthetic networks, it substantially outperforms existing modularity-optimizing state-of-the-art algorithms in terms of final scores of the objective function; for description length, it demonstrates an overall performance similar to the original Infomap algorithm, in certain cases actually showing improved results however. The second, "Delineating geographical regions with networks of human interactions in an extensive set of countries," by Sobolevsky et al., then explores application of this technique to various country-level phone call networks, ultimately showing that the partitioning compares well with politically and socially defined regions.

## 10.3 Visualization Platform

We continue to move forward in the creation of a flexible and robust visualization platform, through which a wide range of data streams can easily be integrated, filtered, recombined and distributed in new, imaginative ways. As new data specific to Riyadh is obtained, it can be added to the visualization platform for individuals to interact, mix and explore. While specific media have not been decided upon (smartphone application versus web application, for instance), the intent is to make the data accessible to a diverse audience.

The *City Mobility Browser* is currently in an alpha state, as many more visualization components have yet to be integrated, but a guided tour that demonstrated the platform's features was performed during the March 2013 workshop. Once the browser is stable and feature-complete, the development team will narrow in on its human-computer interaction aspects.

# References

[1] Arriyadh Development Authority (ADA) (1987) Demographic, transportation, land use, and economic studies for the city of Arriyadh. Technical report. September.

[2] Arriyadh Development Authority (ADA) (1996) MEDSTAR, Phase one: transportation deliverable.

[3] F. Calabrese, M. Colonna, P. Lovisolo, P., D. Parata, and C. Ratti, "Real-Time Urban Monitoring Using Cellular Phones: a Case-Study in Rome" IEEE Transactions on Intelligent Transportation Systems (2010).

[4] N.Nabian, D.Offenhuber, A.Vanky, C.Ratti. "DATA DIMENSION: Accessing Urban Data and Making It Accessible" (2012) From a forthcoming publication.

[5] Annual Report," Saudi Communication and Information Technology Commission" 2010.

[6] F. Calabrese, G. Di Lorenzo, L. Liu and C. Ratti, Estimating Origin-Destination Flows Using Mobile Phone Location Data. Pervasive Computing, IEEE , vol.10, no.4, pp. 36-44 (2011).

[7] G. Krings, F. Calabrese, C. Ratti and V. D. Blondel, Urban gravity: a model for intercity telecommunication flows. J. Stat. Mech. 2009, L07003 (2009).

[8] F. Simini, M. C. Gonzlez, A. Maritan and A.-L. Barabsi, A universal model for mobility and migration patterns. Nature, vol. 484, pp. 96100 (2012).

[9] C. Ratti et al. Redrawing the map of Great Britain from a network of human interactions. PLoS One, vol. 5, e14248 (2010).

[10] Calabrese, F., Reades, J., and Ratti, C. (2010). "Eigenplaces: segmenting space through digital signatures". IEEE Pervasive Computing, 9(1), 78-84.

[11] Phithakkitnukoon, S. and Ratti, C. (2011)."Inferring Asymmetry of Inhabitant Flow using Call Detail Records". Journal of Advances in Information Technology, 2(2), 1-11.

[12] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, Carlo Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data" IEEE Pervasive Computing, 10, 4, (2011), doi:10.1109/MPRV.2011.41

[13] Calabrese, F., Dahlem, D., Gerber, A., Paul, D., Chen, X., Rowland, J., Rath, C., Ratti, C.: "The Connected States of America: Quantifying Social Radii of Influence". In SocialCom/ PASSAT(2011) 223-230

[14] Phithakkitnukoon, S.; Calabrese, F.; Smoreda, Z.; Ratti, C.; "Out of Sight Out of Mind–How Our Mobile Social Network Changes during Migration," Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom), pp.515-520, 9-11 Oct. 2011

[15] Simini, Filippo and Gonzlez, Marta C. and Maritan, Amos and Barabsi, Albert-Lszl, "A universal model for mobility and migration patterns", Nature (2012)

[16] Hernn D. Rozenfeld, Diego Rybski, Jos S. Andrade, Michael Batty, H. Eugene Stanley, Hernn A. Makse, ”Laws of population growth”, Proceedings of the National Academy of Sciences 105, 48 (2008), doi:10.1073/pnas.0807435105

[17] Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, et al. (2010) ”Redrawing the Map of Great Britain from a Network of Human Interactions”. PLoS ONE 5(12):e14248.doi:10.1371/journal.pone.0014248

[18] Dial, R.B. (1971). A probabilistic multi-path traffic assignment model which obviates path Enumeration. Transportation Research, 5, 83 111

[19] Bing-Feng Si, Ming Zhong, Hao-Zhi Zhanga and Wen-Long Jinc. (2010) An improved Dials algorithm for logit-based traffic assignment within a directed acyclic network. Transportation Planning and Technology. Vol. 33, No. 2, March 2010, 123 137

[20] Leurent, F. (1997). Curbing the computational difficulty of logit equilibrium assignment model . Transportation Research, 31B, 315 326

[21] Li, Z.C. and Huang, H.J. (2003). Determining the efficient paths in stochastic traffic Assignment. Journal of Transportation Systems Engineering and Information Technology, 3 (1), 28 32

[22] Wardrop, J. G.; Whitehead, J. I. (1952). ”Correspondence. Some Theoretical Aspects of Road Traffic Research”. ICE Proceedings: Engineering Divisions 1 (5): 767.

[23] Yongtaek Lim, Benjamin Heydecker (2003). ”Dynamic departure time and stochastic user equilibrium assignment ”. Transportation Research Part B 39 (2005) 97118

[24] Traffic Technology International (2012a). Admirable Solution, Traffic Technology International. February/ March.

[25] Traffic Technology International (2012b). In the Frame, Traffic Technology International. April/May.

[26] Papayannoulis, V., Marsico, M., Maguire, T., Strasser, J. and Scalici, S. (2011). An integrated travel demand, mesoscopic and miroscopic modeling platform to assess traffic operations for Manhattan, New York, Proceedings of the Transportation Research Board (TRB) Conference, Washington DC, USA.

[27] Toledo, T., Koutsopoulos, H. N., Davol, A., Ben-Akiva, M. E., Burghout, W., Andreasson, I., Johansson, T. and Lundin, C. (2003). Calibration and validation of microscopic traffic simulation tools: Stockholm case study, Transportation Research Record 1831: 6575.

[28] Hasan, M. (1999). Evaluation of ramp control algorithms using a microscopic traffic simulation laboratory, MITSIM, Masters thesis, Massachusetts Institute of Technology.

[29] Wang, Pu, Timothy Hunter, Alexandre M. Bayen, Katja Schechtner, and Marta C. Gonzalez. (2013) ”Understanding Road Usage Patterns in Urban Areas.” Nature.com. Nature Publishing Group, 20 Dec. 2012. Web. 04 May 2013.

[30] Osorio, C. and Bierlaire,M. (2010). A simulation-based optimization approach to perform urban traffic control, Proceedings of the Triennial Symposium on Transportation Analysis (TRISTAN), Tromsø, Norway.

[31] Osorio, C., and L. Chong. 2012. “An efficient simulation-based optimization algorithm for large-scale transportation problems”. Winter Simulation Conference (WSC), Berlin, Germany.

[32] Osorio, C., and K. Nanduri. 2012. “Energy-efficient traffic management: a microscopic simulation-based approach”. International Symposium on Dynamic Traffic Assignment (DTA), Marthas Vineyard, USA.

[33] Chen, C. Osorio, and B. F. Santos. 2012. “A simulation-based approach to reliable signal control”. International Symposium on Transportation Network Reliability (INSTR).

[34] Conn, A. R., Scheinberg, K. and Vicente, L. N. (2009b). Introduction to derivative-freeoptimization,MPS/SIAMSeries on Optimization, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA, USA.

[35] Barton, R. R. and Meckesheimer, M. (2006). Metamodel-based simulation optimization, in S. G. Henderson and B. L. Nelson (eds), Handbooks in operations research and management science: Simulation, Vol. 13, Elsevier, Amsterdam, chapter 18, pp. 535–574.

[36] Søndergaard, J. (2003). Optimization using surrogate models by the Space Mapping technique, PhD thesis, Technical University of Denmark.

[37] Osorio, C. and Bierlaire, M. (2008). Network performance optimization using a queueing model, Proceedings of the European Transport Conference (ETC), Noordwijkerhout, The Netherlands.

[38] D. Heidemann. A queueing theory approach to speed-flow-density relationships. In Proceedingsof the 13th International Symposium on Transportation and Traffic Theory, pages 103–118, Lyon, France, July 1996.

[39] T. Van Woensel and N. Vandaele. Modelling traffic flows with queueing models: A review. Asia-Pacific Journal of Operational Research, 24(4):1–27, 2007.

[40] Dornier Consulting (2008). Strategic Context, a report developed for Municipality of Riyadh.

[41] TSS 2008, May. AIMSUN NG and AIMSUN Micro Version 5.1. Transport Simulation Systems.

# People

## Principal Investigators

Carlo Ratti, PhD
Carolina Osorio, PhD
Anas Alfaris, PhD

## Researchers

Riccardo Campari, PhD
*Postdoctoral Fellow*

Markus Schläpfer, PhD
*Postdoctoral Fellow*

Linsen Chong
*Phd Student*

Anthony Vanky
*Phd student*

Abdullah Abu Khwejah
*Research Specialist*

Eyas Alfaris
*M.S. student*

Fahad Alhasoun
*Research Specialist*

Kael Greco
*M.S. student*

Carter Wang
*M.S. Student*

Center for Complex Engineering Systems
At KACST and MIT

http://www.cces-kacst-mit.org/

CCES

مدينة الملك عبدالعزيز
للعلوم والتقنية KACST

MIT ESD