

# CityScope Andorra Data Observatory: A Case Study on Tourism Patterns

Arnaud Grignard<sup>1</sup>, Luis Alonso<sup>1</sup>, Núria Macià<sup>2</sup>, Marc Vilella<sup>3</sup>, Kent Larson<sup>1</sup>

<sup>1</sup> MIT Media Lab - City Science - Cambridge, USA

<sup>2</sup> Universitat d'Andorra - Sant Julià de Lòria, Andorra

<sup>3</sup> Observatori de la Sostenibilitat d'Andorra - Sant Julià de Lòria, Andorra

**Abstract.** This paper presents a data-driven agent-based simulation of individual mobility based on spatio-temporal data from mobile phones. The model developed is embedded within the CityScope framework, a platform used as decision support system for urban planning. This work analyzes the Andorra visitors' flow and traffic congestion through an agent-based visualization using different representation and abstraction features.

## 1 Introduction

Telecom data coupled with other data sources—such as social media—can help us understand human behavior at spacial, temporal, and social level. These unprecedented rich sources of data allow us to study how people move and, thus, how our society behaves. Previous research [1] [2] shows that such insights can be used to design interventions to improve our daily lives and even visitors' experience in the scope of tourism strategies. However, the results of the analyses are not always comprehensible for non-experts. CityScope is a visualization framework, developed by City Science at MIT Media Lab, that serves both as an urban data observatory and laboratory for decision making in urban planning. CityScope is a next generation, tangible, augmented reality platform that helps to (1) visualize and understand the meaning of complex urban data and inter-relationships, (2) simulate the impact of multiple interventions, and (3) support decision making in a dynamic, iterative, and evidence-based process. CityScope helps non-experts to engage into conversation through visualizations that synthesize the analyses in a coherent manner on the physical model of their cities. An Agent-Based Model (ABM) has been used for simulating the actions and interactions of autonomous agents in this work. The central idea of the model is to show emerging patterns in visitors' behavior during specific events and understand how this may affect and coexist with regular activity. This model leads to insightful visualizations that show how visitors move across the country.

The remaining of this paper is organized as follows. Section 2 gives a general overview of the framework. Section 3 describes the input data. Section 4 describes the model implemented. Section 5 shows the visual results of the simulation and, finally, Section 6 discusses further research.

## 2 Overview

Andorra, located between Spain and France in the middle of the Pyrenees, is a country with a population around 78,000 people that welcomes more than eight million visitors a year. According to the statistics provided by the *Departament d'Estadística d'Andorra*, the tourism sector accounts for 80% of the GDP of the country. Andorra has two types of visitors: (1) tourists, which stay over at least one night and (2) same-day visitors, which enter and leave the country the same day. The presented model has mainly been developed to simulate the movement of visitors across the territory and gain understanding on this industry. Modeling people's flow can help us assess the actual impact of visitors in terms of traffic congestion, energy consumption, consumer spending, among others. The current model focuses on visitors' attendance at the events held in the country as well as traffic congestion levels. The following two events in 2016 have been analyzed: (1) *Cirque du Soleil: VISION* and (2) *Le Tour de France*.

Scalada by *Cirque du Soleil* is a series of indoor, summer shows specifically designed for Andorra by 45 DEGREES—the global events company from *Cirque du Soleil*. Since 2013, the company has combined art, technology, and Andorran elements in their performances and attracted many visitors to the country. *VISION* was a 60-minute event performed on Tuesdays, Wednesdays, Thursdays, Fridays, and Saturdays from July, 2 to July, 30 2016 at 10:00pm. The venue had a capacity for 5,000 people per performance. *Le Tour de France* is the annual multiple stage bicycle race held in France. However, the event occasionally makes passes through nearby countries. In 2016, Andorra hosted the arrival of the Stage 9 up in the mountains (Arcalís, Ordino) and the departure of the Stage 10 in the city center (Escaldes, Escaldes-Engordany).

The model has been implemented using different environments: Processing [3] and the GAMA platform [4]. Processing is both an open source programming language and integrated development environment built for the visual design community. It is well suited for accurate visualizations and provides a seamless feature interaction integration. However, it has not been designed to develop complex ABM. In this sense, GAMA is a modeling and simulation-development environment for building spatially explicit agent-based simulations. This multiple-application domain platform uses a high-level and intuitive agent-based language that allows users to undertake most of the tasks related to modeling, visualization, and simulation exploration with dedicated tools. GAMA includes a uniform cost search pathfinding algorithm that allows to work out custom pathfinding logics—helping overcome issues encountered with Processing. The resulting model is projected on the Andorra CityScope table, which is a 3D model of the two main cities of Andorra<sup>4</sup>.

---

<sup>4</sup> It has been presented at the Smart City Expo World Congress 2016 in Barcelona. <https://youtu.be/hdL0aundHL4>

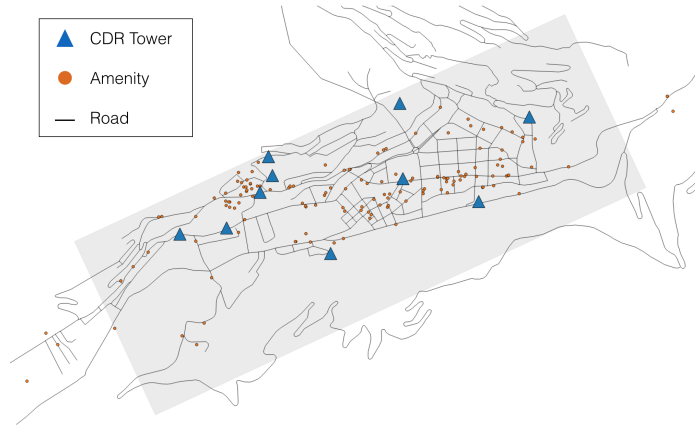
### 3 A Data-driven Model

Figure 1 represents the different elements obtained from the input datasets: (1) telecom data, (2) amenities, and (3) road network. The ABM uses these different types of data to characterize both static and dynamic agents.

**Telecom data:** We use cell phone communication data to understand human mobility patterns. Andorra Telecom provided a three-year collection—from 2014 to 2016—of anonymized Call Detail Records (CDR), which represents a total of 450GB of data. Observations in these records have a spatial component and are triggered by any kind of action with a mobile phone (i.e., phone call, text message, cellular data). From the features, we obtain the location of the cell towers involved in the action and, thus, compute the origin and destination of each agent. We can also assign the country of residence to the agents.

**Amenities:** Amenities are places where agents may go, such as restaurants, hotels, or points of interest. Their geolocation have been gathered from TripAdvisor, Yelp, and the *Andorra Turisme* office.

**Road Network:** Agents do not move in a straight line; their trajectories are constrained by the actual road network. Therefore, agents move along a graph topology, which is provided by Open Street Maps. Roads can be of different type (primary, secondary, residential, and pedestrian) allowing only certain behaviors. Roads can be either one-way or bidirectional, but not all agents can go in both directions. The congestion level is updated during the simulation according to the number of agents present on the road and it can be modified to emphasize specific patterns such as traffic congestion.



**Fig. 1.** Input data: cell towers (triangle), amenities (circle) and roads (line). The grey area represents the scope of the Andorra CityScope table.

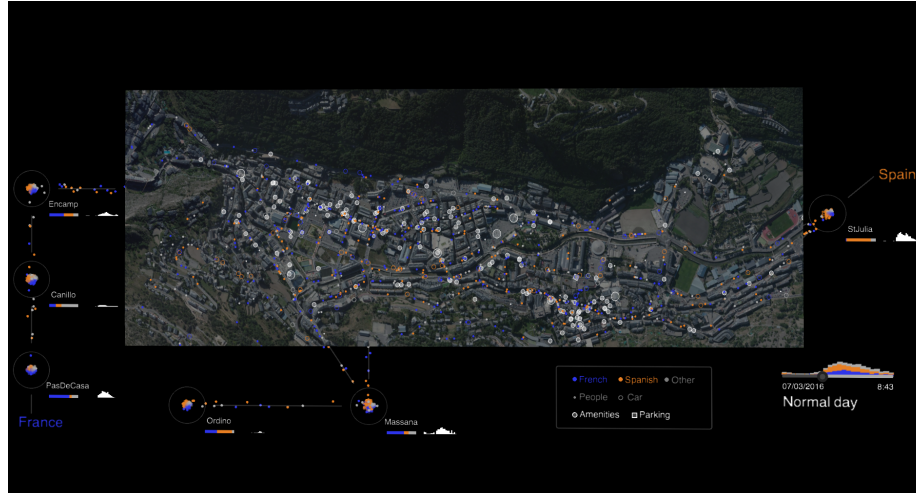
## 4 Model Description

Figure 2 corresponds to the simulation of a regular day—used as a benchmark in the analyses. Every simulation represents a full day and runs until all the observations of the day from the CDR data set are processed. People are represented by solid circles and vehicles by stroke circles; their color varies according to the country of residence—red refers to people from Spain, blue refers to people from France, and white refers to people from other countries. For the sake of clarity, we follow the classification provided by the *Departament d’Estadística d’Andorra*: Spain, France, and Others. In 2016, Andorra received eight million visitors approximately divided into 4.2 M Spanish people, 3.2 M French people, and 0.6 M from other nationalities. However, the telecom data allows us to provide more detailed classification. By extracting the origin of the SIM card, we identified visitors mainly from Belgium, The Netherlands, United Kingdom, Italy, Norway, United States, and Germany.

At city scale, the two main, central cities of the country—Andorra la Vella and Escaldes-Engordany—are explicitly displayed using Geographic Information Systems (GIS) data (see map in Fig. 2). The rest of the territory is conceptually represented by clusters, which correspond to the two cities located near the border (i.e., Sant Julià de Lòria near the Spanish border and Pas de la Casa near the French border) and the parishes of Canillo, Encamp, Ordino, and La Massana (see pie charts in Fig. 2). The emerging structures show people’s flow from one city to another giving a general view of the activity at a country level.

ABMs have successfully been applied to study emergences from a wide range of adaptive system made of individual entities, contributing to an easier and deeper understanding on local interactions, variability among entities, adaptive behaviors, and environmental states [5]. Lately, ABMs have also been used as a data visualization tool since they give the possibility to interact with the representation [6]. In the presented model, dynamic agents have a set of variables assigned that influence their behavior whenever a change occurs, either in its own state (e.g., when the agent arrives at its destination, it stops) or in the external environment (e.g., when a road is full, the agent can take an alternative path). The set of variables is composed of (1) country of residence, (2) origin location—defined randomly or using telecom data—, (3) preferred destination—generated by a decision making submodule—, (4) distance traveled, (5) speed of movement, and (6) passable streets. Agent’s trajectory is determined by an Origin-Destination (OD) matrix. The OD matrix is computed using the location of the cell towers where the action with the mobile phone was originated and terminated. The destination of the agent is set to the closest amenity to the cell tower where the action terminated. Depending on its speed (time difference between origin and destination location), the agent will be considered as a walking person (solid circle) or as a vehicle (stroke circle). The model is implemented with an enriched GIS data that provides information to the dynamics agents in order to adapt their movement such as amenities’ capacity and working hours, events, and direction of roads. Agents adapt themselves to both (1) congestion traffic and (2) amenity occupancy.





**Fig. 2.** Andorra CityScope overview. The map provides a detailed city view whereas the clusters provide an abstract view of the country. Each cluster grows according to the number of users. The temporal evolution of the simulation is displayed hourly on the slider and the volume of people is displayed by the histogram at the bottom right corner.

**Congestion traffic.** If congestion is too high, a pathfinding is called to recalculate an alternative route. If a road is busy, then the agent will recompute the shortest path to its destination avoiding this road.

**Amenity occupancy.** Once agents reach their destination, they stay there for a few iterations. The number of iterations is defined by the average time spent on those places. The amenity size increases (or decreases) according to the number of agents currently in the location. Depending on the amenity occupancy, the agent might recompute its destination. If the amenity assigned as destination is full, the agent will select another amenity close to its initial destination. The chosen amenity also depends on the agent's country of residence and the language affinity of the amenities.

## 5 Results

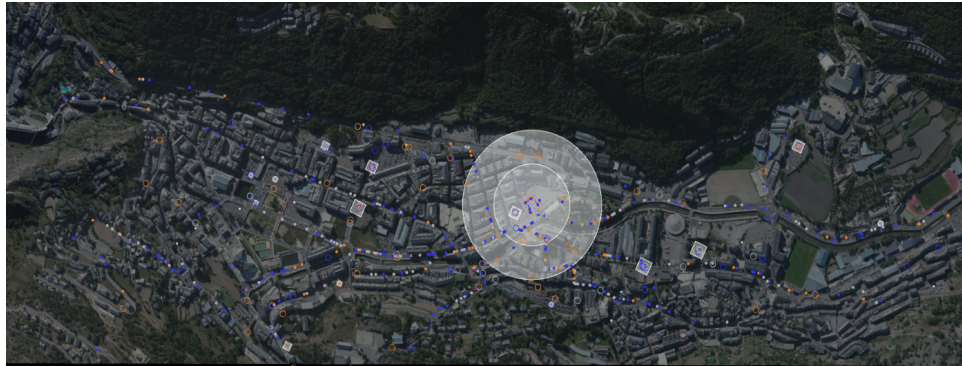
The emerging patterns display the actual dynamics of the city providing a urban planning tool that goes beyond the traditional ones that are usually focused on land uses and sociological static data extracted from surveys. The ABM visualization shows different patterns of movements from visitors revealing the structure of the city as a complex system.

The following subsections describe (1) raw and (2) aggregated results, which highlight helpful information regarding *Cirque du Soleil* and *Le Tour de France*, and refer to (3) online results.

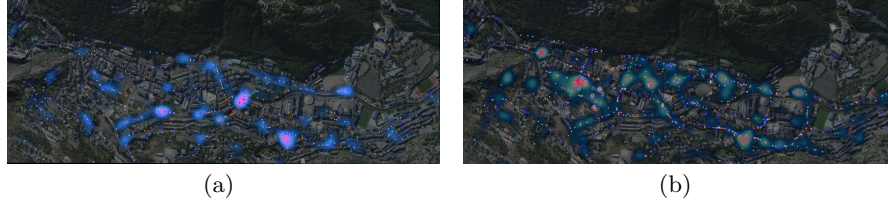
### 5.1 Raw Results

When running a simulation on the Andorra CityScope table, one can immediately identify three main elements: (1) city representation defined by buildings, amenities, cell towers and roads, (2) people’s movement defined by dynamics agents, and (3) amenities’ density.

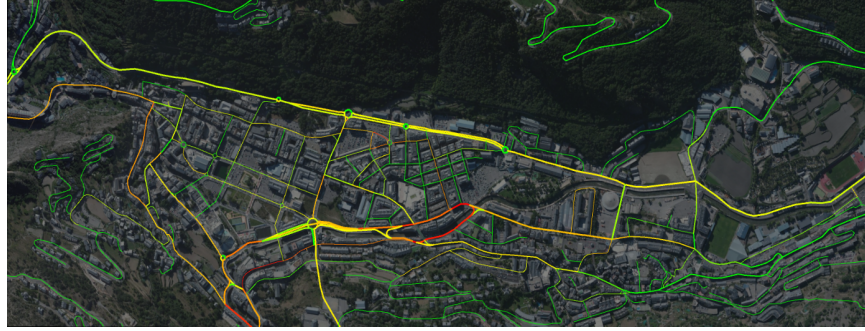
As mention in Section 4, the number of people that are present in the amenities evolves during the simulation. The amenity size increases (or decreases) according to the number of agents currently in the location. This helps identify which and when places are popular or busy and isolate them. For instance, Fig. 4 shows the activity during the *Cirque du Soleil* on July, 16. The wide white circle spots the location where the show was taking place. This was the densest place at that time; ticketing for the event was 5,174 attendees. According to the statistics from the *Andorra Turisme* report, the average attendance per performance was 4,540 people in 2016. Overlapping layers from different days and/or editions of the event—while dynamically running them—is a useful way to display and discuss the numbers.



**Fig. 3.** Simulation for the *Cirque du Soleil* on July, 16.



**Fig. 4.** Aggregated results. (a) Instant congestion heatmap for a specific time during *Le Tour de France* and (b) heatmap that summarizes the congestion for the whole day.



**Fig. 5.** Traffic congestion during the *Cirque du Soleil* on July, 16. From green to red, the traffic congestion increases.

## 5.2 Aggregated Results

Differently from the *Cirque du Soleil*, *Le Tour de France* is an outdoor event and there is no ticketing process to assess attendance. To this end, aggregated data can be visualized on the CityScope table resulting in heatmaps that summarize global activity in the city and provide an attendance estimate. Figures 4(a) and 4(b) show occupancy levels for *Le Tour de France* on July, 12. The starting line was in Escaldes-Engordany, which corresponds to the hottest area in Fig. 4(b)—large red concentration on the left side of the image. Comparing this kind of visualization to the activity of a regular day, one can understand which events bring more people to the country. In addition, we can identify where these visitors go and what they do. This could be used to efficiently plan events or find new ones that help spread visitors across the territory. Figure 6 shows congestion levels. Focusing on the roads only, the movement of agents representing vehicles can be translated into another view based on traffic density.

## 5.3 Online Results

A video displaying the ABM visualization of the Andorra CityScope Data Observatory is available at: <https://youtu.be/fLikAuFvVyg>

## 6 Discussion and Further Research

Data collected from wearable could lead to more accurate human behavior studies. However, CDR data can cover larger groups for longer periods of time in a non-invasive way and studies can be easily scaled/replicate. CDR data only allow to trace visitors movement based on the cell towers geolocation. To improve the accuracy of the ABM, Andorra Telecom is collecting a new source of data from the Radio Network Controller that can now provide the geolocation of the devices. The geolocation has an error of 50-100 m in urban areas and up to 200 m in rural areas. Besides, observations for this data source are triggered (1) by any kind of action with the phone (phone call, text message, cellular data)—like CDR data, but also (2) when the user moves and the network detects the device changes cells or technology (i.e., 2G, 3G, 4G), or (3) when the user is static and the update network timer expires.

Further work includes integrating data coming from sensors deployed in stores to study how visitors move and behave inside buildings. This model will also be used to replay agents' behavior in order to understand the city dynamics and lead to more efficient urban designs.

## Acknowledgment

This work has been developed within the framework of collaboration between MIT Media Lab City Science and Fundació ActuaTech.

## References

1. S. Jiang, J. Ferreira, and M. C. González, "Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, 2017.
2. M. Batty, "Cities as complex systems: scaling, interactions, networks, dynamics and urban morphologies," 2008.
3. C. Reas and B. Fry, *Processing: a programming handbook for visual designers and artists*. No. 6812, Mit Press, 2007.
4. A. Grignard, P. Taillandier, B. Gaudou, D. A. Vo, N. Q. Huynh, and A. Drogoul, "Gama 1.6: Advancing the art of complex agent-based modeling and simulation," in *International Conference on Principles and Practice of Multi-Agent Systems*, pp. 117–131, Springer, 2013.
5. R. J. Allan, "Survey of agent based modelling and simulation tools," tech. rep., 2009.
6. A. Grignard and A. Drogoul, "Agent-based visualization: A real-time visualization tool applied both to data and simulation outputs," 2017.