Horizon Europe MSCA Project

(Grant Agreement Number 101086219)

Intelligent and Proactive Optimisation for Service-centric Wireless Networks

IPOSEE

D1.1

Datasets curated for wireless traffic/mobility forecasting

UPPSALA UNIVERSITET

Authors(s)	Jiming Chen, Kan Lin
Author(s) Affiliation	Ranplan Wireless Network Design Ltd.
Editor(s):	Jiming Chen, Xiaoli Chu, Yiouli Kritikou, Kan Lin, Weijie Qi, Di Yuan
Status-Version:	Final
Project Number:	101086219
Project Title:	Intelligent and Proactive Optimisation for Service-centric Wireless Networks
Project Acronym:	IPOSEE
Work Package Number	1

Table of Contents

List of Figures

List of the Table

Executive summary

Wireless networks have become indispensable to citizens, enterprises and industries, such as transport (including autonomous vehicles and drones), logistics, utilities and manufacturing. As a result, wireless networks are facing increasingly diverse, stringent service requirements, which causes insufficient network management. So how to characterize and forecast high-dimensional complex traffic patterns for individual (or aggregated) services become open challenges. In this report, we will take advantage of Artificial Intelligence (AI) to forecast spatial-temporal patterns of traffic and mobility for individual (or aggregated) services, identify the representative subset(s) of various services, and then optimise the typical RAN deployment based on diverse traffic patterns for different scenarios in Ranplan Professional.

1. Introduction

Future wireless networks are the Internet of Everything (IoE) network, which have become indispensable to citizens, enterprises and industries, such as transport (including autonomous vehicles and drones), logistics, utilities and manufacturing, therefore, they have become the cornerstone of the global digital economy and a key component in our daily lives. Each year, the mobile cellular network industry contributes 3.6% to the global GDP ($2.4 trillion), supports 10.5 million jobs, and has contributed to $336 billion of public funding. The wireless network relies on widely distributed communication nodes, such as terrestrial base stations, to enable service for diverse scenarios. Generally, cell densification results in a growth of capacity and coverage.

In dense urban areas, large crowds can form, disperse, migrate, and demand wireless services in spontaneous and unpredictable ways. The diverse, stringent service requirements of wireless networks indicate that existing reactive network management will be insufficient, while intelligent and proactive control of service-centric networks becomes essential. However, mobile users and their data demands are not uniformly distributed in space or over time, thus it is challenging to characterize and forecast high-dimensional complex traffic patterns for individual (or aggregated) services.

Network planning and optimisation request to configure the traffic pattern, and then optimise the network deployment. This report will develop probabilistic deep learning (DL) to recognize the environment and forecast spatial-temporal patterns of traffic and mobility for individual (or aggregated) services and quantify the associated service-specific prediction uncertainties, identify the representative subset(s) of various services, and then optimis the network deployment and radio access network (RAN) parameters based on the diverse services for the different scenarios.

1.1 Purpose of this document

The objective of this report is to present part of the work and activities of the IPOSEE consortium during this project duration for work package one. The general aim is to use dataset to analyze and forecast the traffic pattern, and optimise network deployment to meet the requirement of services, i.e.

- Use alternative dataset to identify the environment for the hotspot;
- Run unsupervised learning algorithm to separate the data to clusters and predict the traffic trend;
- Run the automatic cell optimization (ACO) to optimise the network deployment and RAN parameters to meet the requirements of Quality of Service (QoS) and capacity;

1.2 Document structure

This report will present the main implementation in Section 3that is split into three sub-sections:

- Section 3.1 Determine the traffic hotspots by using alternative dataset.
- Section 3.2 Run unsupervised learning algorithm to separate the data into clusters and identify the spatial-temporal traffic trend.
- Section 3.3 Optimise the network deployment by comparing iteratively optimises the network deployment to meet the requirement of traffic

2. Wireless traffic forecasting

More wireless traffic is generated at hotspots, such as office, commercial centre, train station and airport, etc., where wireless traffic is highly correlated with the type of the environment, and knowledge of these characteristics enables proactive and dynamic network operation. This section will present how to identify the traffic hotspots area automatically, how to predict the wireless spatial-temporal patterns of traffic and trend, and how to optimise the wireless network based on the traffic requirements.

2.1 Environment identification for traffic hotspot areas determination

This subsection provides a solution to determine the traffic hotspot areas automatically, and then classify these areas into different traffic clusters. Subsection 3.2 provides a prediction algorithm to forecast the traffic trend for different clusters based on the traffic data set.

In order to reach our objective, we need to train a model to be able to apply semantic segmentation on aerial or satellite images. The purpose of this model is to recognize various types of environments, which may have different traffic requirements.

2.1.1 Dataset

The first step is to prepare the dataset. We have used the public dataset ‘Semantic segmentation of aerial imagery from Kaggle (https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery). The dataset consists of aerial imagery of Dubai obtained by MBRSC satellites and annotated with pixel-wise semantic segmentation in various classes (Building, Land, Road, etc.). The total volume of the dataset is 72 images grouped into 6 larger tiles. Since we have mainly focused on the obstacles that may affect signal propagation, we have modified the original classes into more detailed 6 classes, as shown in Table 1:

Class	**Class Name**	**Color Hex #**	**Color RGB**
0	Unlabelled	9B9B9B	(226, 169, 41)
1	Metro and Trains	3C1098	(60, 16, 152)
2	Airport	8429F6	(132, 41, 246)
3	Commercial centres	6EC1E4	(110, 193, 228)
4	Corporate Office	FEDD3A	(254, 221, 58)
6	Stadiums	E2A929	(80, 227, 194)

Table 1. Class definition

Original classes are labelled with hex colour code, and we have transferred the original masks in RGB and encode the output as integers for later segmentation tasks training.

2.1.2 Data augmentation

Next, we have applied data augmentation by the following techniques:

- Random cropping

- Horizontal flipping
- Vertical flipping
- Rotation
- Gaussian noise and filtering operations

2.1.3 Model training

We have used the InceptionResNetV2 model which has been pre-trained on the ImageNet dataset as an encoder network. A decoder network has been extended from the last layer of the pre-trained model, and it is concatenated to the consecutive layers to suit for our class design. The architecture is shown in Figure 1.

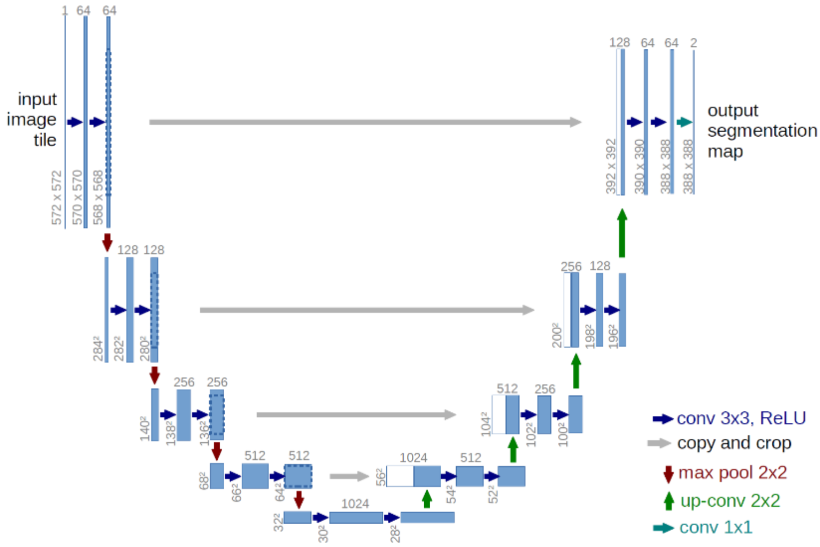


Figure 1. Model training architecture.

During training, we have tried several combinations of hyper-parameters, and the final training parameters are as follows:

- Batch Size = 16.0
- Steps per Epoch = 32.0
- Validation Steps = 4.0
- Input Shape = (512, 512, 3)
- Initial Learning Rate = 0.0001
- Number of Epochs = 60

Based on the model training algorithm, we can determine the spatial traffic hotspot areas, as shown in Figure 2, where the environment and buildings are identified.

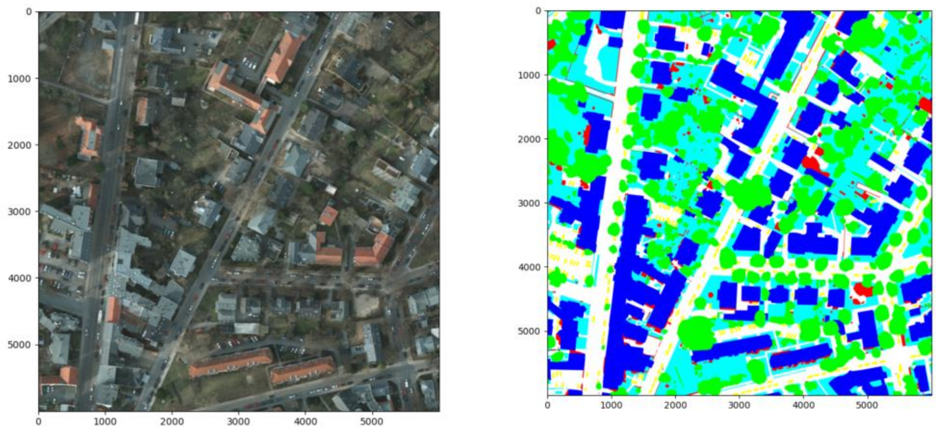


Figure 2. Environment type identification.

2.2 Spatial-temporal traffic forecasting

This section provides a prediction algorithm to forecast the traffic trend for different clusters based on the traffic data set. We use network deployment of France to illustrate the algorithms, and the network deployment across France is shown in Figure 3, where thousands of base stations are installed at more than 1000 locations.

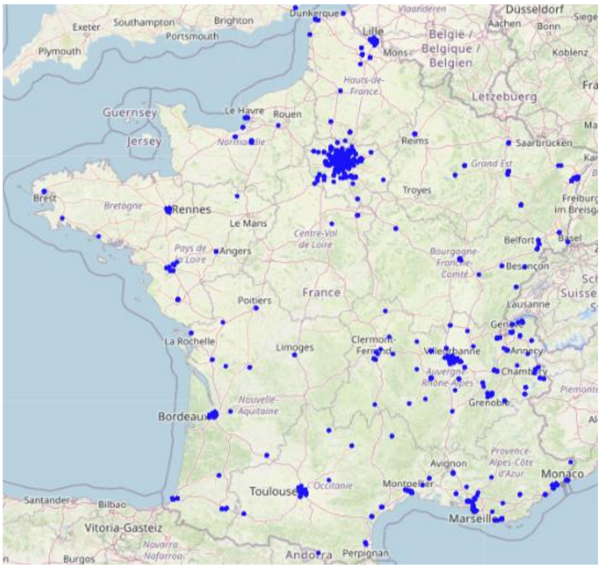


Figure 3. Network deployment across France.

2.2.1 Spatial traffic classification

Downlink and uplink internet traffic load of base stations in the wireless network and mobile application utilization, at an aggregated level, allow us to distinguish different environment types, e.g., metros, corporate offices, expo centers etc. The data set include metros, train stations, stadiums, airport, highway and tunnel, commercial centers, corporate offices, etc. Based on these traffic data, the spatial traffic can be predicted, where a hierarchical clustering algorithm is used to identify common trends and split the data to clusters, as shown in Figure 4 and Figure 5.

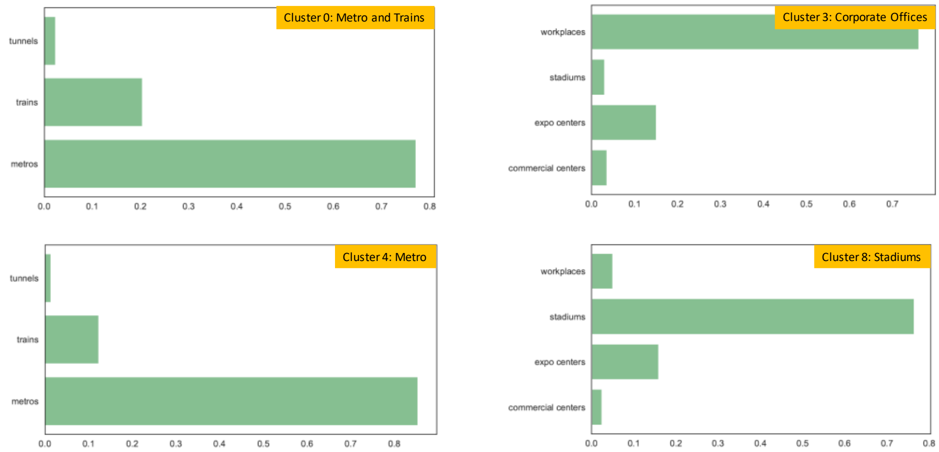


Figure 4. Spatial traffic classification.

Figure 5. Traffic clustering.

2.2.2 Temporal traffic

Based on the traffic clusters, the traffic with different time can be analyzed, as shown in Figure 6, where traffic distribution in one week is curved, which can be trained to predict the traffic pattern.

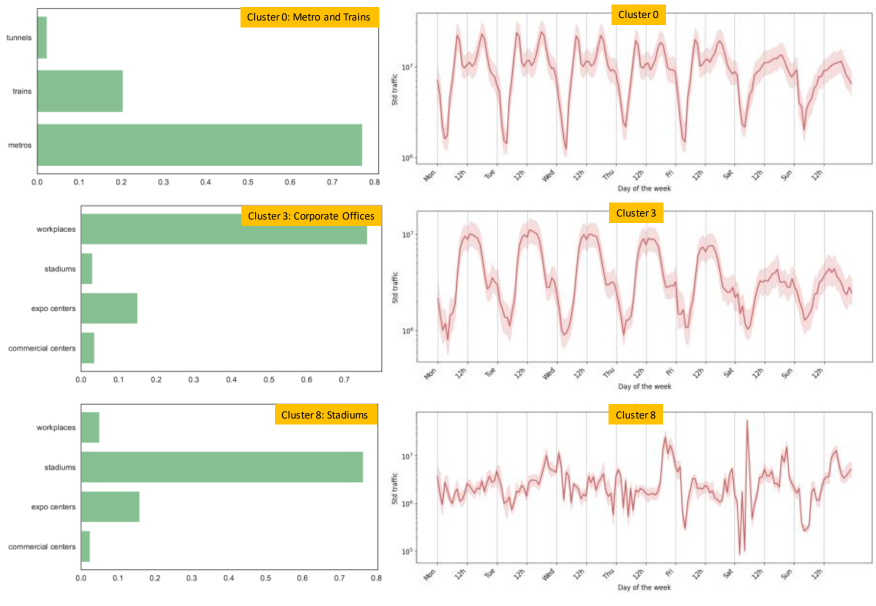


Figure 6. Temporal traffic analysis.

2.2.3 Spatial-temporal traffic predictions

We use explainable artificial intelligence to interpret the results of clustering analysis, as shown in Figure 7. To probe the dynamics of the hotspots' mobile traffic demands and characteristics across the nationwide network infrastructure, we pursue an unsupervised learning approach. In particular, we consider the aggregated sum of the downlink and uplink traffic recorded for each mobile service over the target two-month period as a distinct feature, and we form a matrix, $T^{N\times M}$, comprising the overall traffic in megabytes (MB) for N = 4, 762 indoor antennas and M = 73 mobile services included in the data set. To capture the impact of each mobile service in the data generated at each antenna, we use the revealed symmetric comparative advantage (RSCA) as a metric, defined as:

$$RSCA_{i,j} = (RCA_{i,j} - 1)/(\ RCA_{i,j} + 1),$$

Where RCA means revealed comparative advantage, $RCA_{i,j} = (T_{i,j}/T_i)/(T_j/T_{tot})$ and $T_{i,j}$ stands for the traffic recorded for the j-th service at the i-th antenna, T_i refers to the total traffic generated at the i-th antenna for all the services, T_j depicts the summed traffic over all antennas for the j-th service, and T_{tot} is the total traffic channeled through the network during the entire period in question.

To cluster the traffics according to these values, we used agglomerative clustering, which is a state-of-the-art hierarchical clustering algorithm, following a bottom-up approach. Initially, it is assumed that each data point forms its own cluster, and thereafter the clusters are merged greedily according to a specified criterion. In particular, we use Ward's criterion which aims at minimizing the total intra-cluster variance, measured as the squared distance between the cluster centers, when merging two clusters. Hence, the clustering algorithm starts from N distinct clusters and repeatedly merges the clusters, reducing the total number in a way that the new cluster yields a reduced intra-cluster variance. The number of clusters is set to k = 9, according to the Silhouette Score and the Dunn index.

Then, to uncover the patterns that the hotspots' traffic exhibit over time, the data from the different clusters indicated from the previous analysis are aggregated and a statistical analysis is performed. The following heatmaps show the evolution of the normalized median traffic per hour across all the antennas found in workspaces (cluster 3) and metro/train station (cluster 0). Evidently, it can be pointed out that there is a strong correlation between the temporal patterns and the indoor environment type; cluster 4 demonstrates a traffic peak during the common weekly commuting hours, while cluster 3 consisting primarily of workspaces that remains idle during weekends and after working hours, i.e., 5.30 pm, and all traffic is concentrated between working hours.

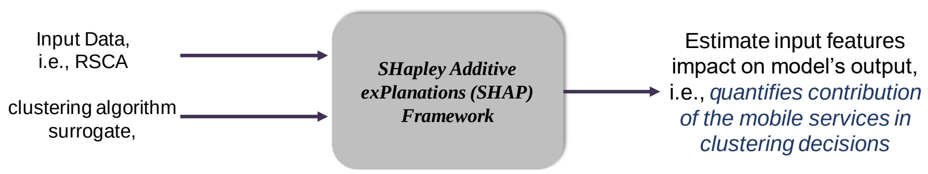


Figure 7. Traffic predictions.

Figure 8 presents the spatial-temporal traffic patterns after predicting the spatial traffic clusters and temporal traffic, where a high-dimensional complex traffic patterns for individual (or aggregated) services is analyzed. The spatial-temporal traffic patterns can be inputted into Automatic Cell Optimization (ACO) module in Ranplan professional.

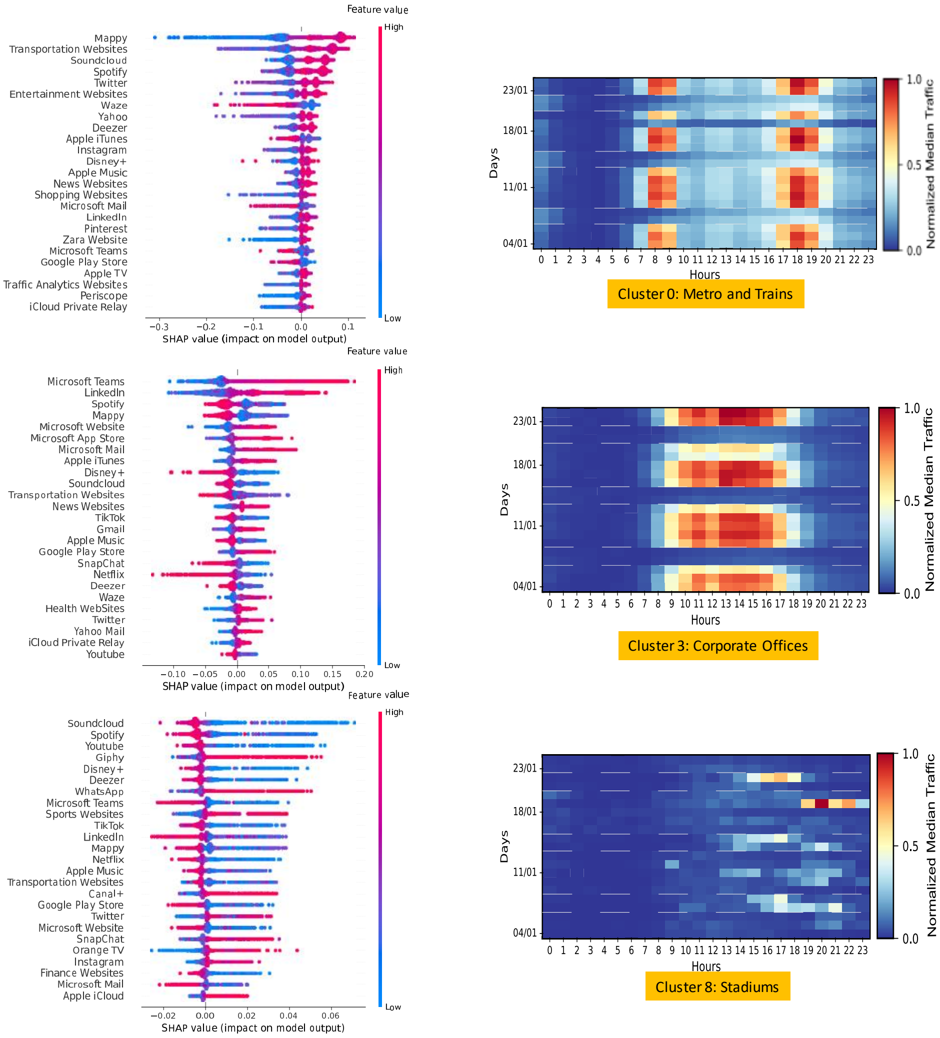


Figure 8. Spatial-temporal traffic patterns.

2.3 Automatic cell optimisation in Ranplan Professional

Accordingly, this section takes advantage of Ranplan Professional to model the buildings and wireless environments according to the traffic requirements. This simulation will accurately and quantitatively deploy the network to meet the traffic requirement of hotspots by the ACO (Automate Cell Optimiser). We will analyze an indoor office scenario.

2.3.1 Outdoor environment and indoor building modeling in Ranplan Professional

Considering that the outdoor network will impact the coverage and capacity of hotspot significantly, the outdoor scenarios should be modeled close to reality. In this outdoor case, the outdoor scenario is modelled via importing the GIS data, including building, foliage, terrain, clutter data, that can be directly imported to Ranplan

Professional using the following function within the 'Project explorer', as Figure 9 shows.

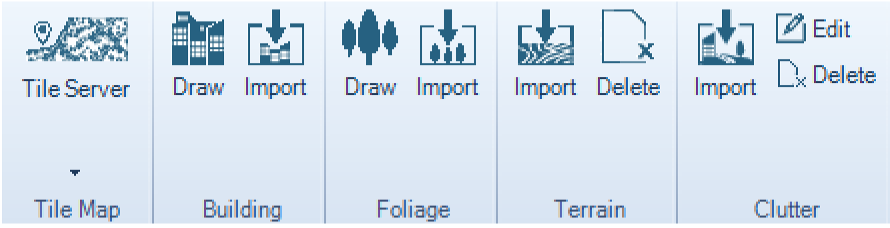


Figure 9. GIS data import function in Ranplan Professional.

The software will automatically generate the outdoor scenario based on the imported data. The generated outdoor scenario can be checked in either 2D or 3D. Two views of the outdoor scenario are shown in Figures 10, where the corresponding indoor building is modelling in this tool, as shown in Figure 11.

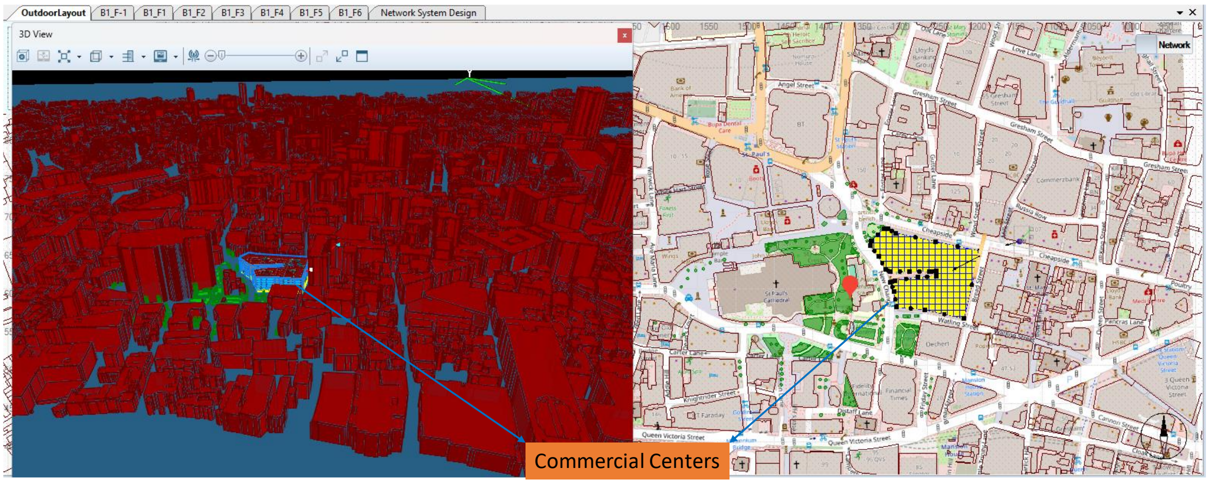


Figure 10. 3D and 2D view of the imported GIS data.

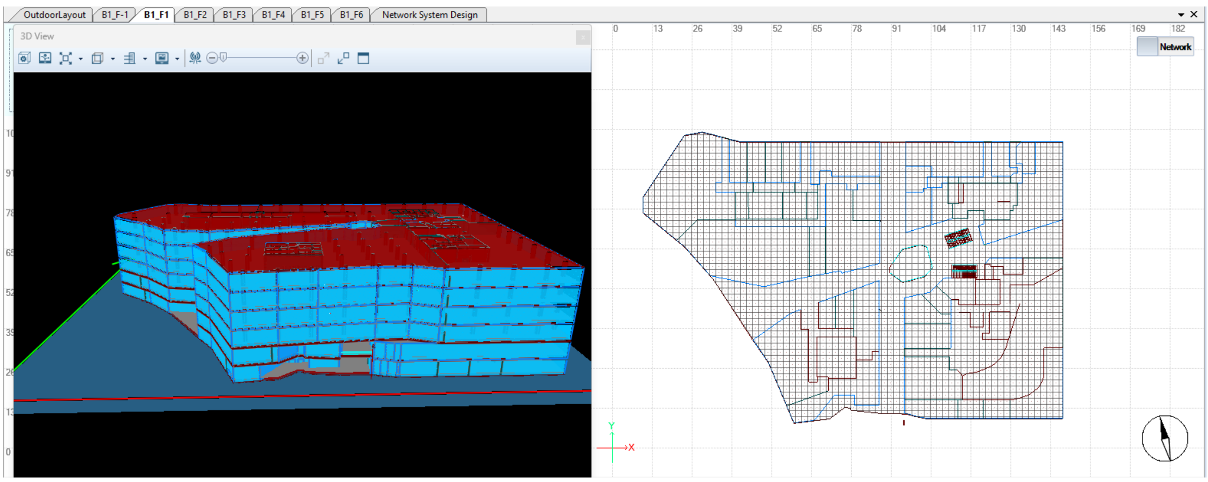


Figure 11. 3D and 2D indoor building.

Based on the solution of Section 3.1, this building is identified as a commercial center, and the corresponding spatial-temporal traffic patterns are forecasted.

2.3.2 Traffic configuration in Ranplan Professional

Based on the predicted traffic pattern in Section 2, the corresponding traffic maps can be created to optimise the network deployments and RAN parameters. Figure 12 and Figure 13 show the service types and traffic maps, which can be automatically configured from the traffic patterns.

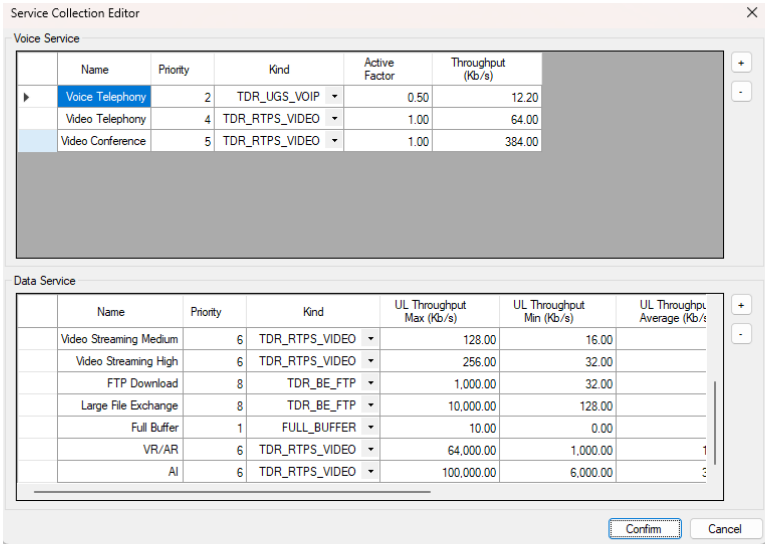


Figure 12. Service types in Ranplan Professional.

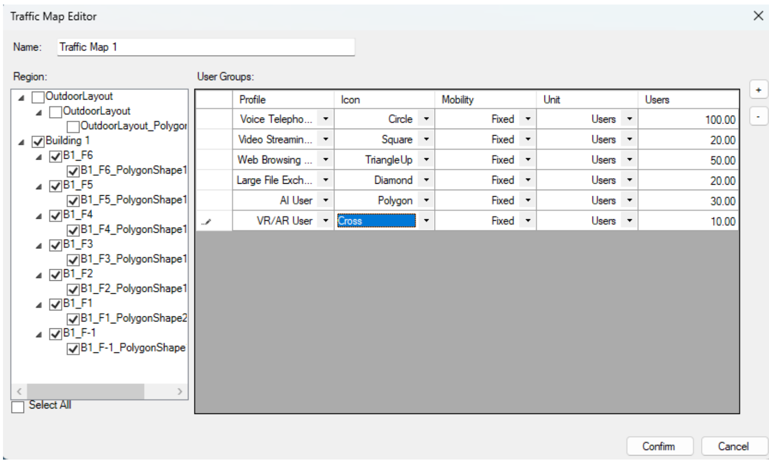


Figure 13. Traffic maps.

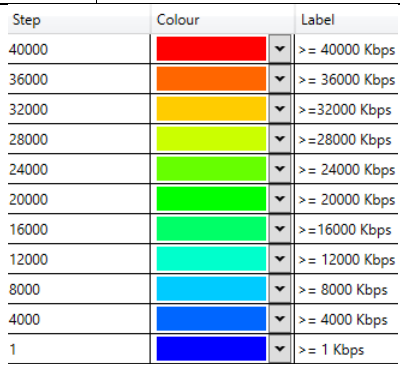
2.3.3 Simulation parameters configuration

Table 2 shows the system configuration in the optimisation case study, and legend of capacity is shown in Figure 3.

Table 2. Simulation parameters

Parameters	Configuration
Wireless System	FDD LTE

Carrier Frequency	2.6GHz
Band Width	20 MHz
Cell Transmission Power	15 dBm
Channel Model	Ranplan Maxwell Propagation Engine (MPE)
Antenna Resolution	0.5 Meters (indoor)
Antenna Coverage Radius	100 Meters (indoor)



Step	Colour	Label
40000		>= 40000 Kbps
36000		>= 36000 Kbps
32000		>=32000 Kbps
28000		>=28000 Kbps
24000		>= 24000 Kbps
20000		>= 20000 Kbps
16000		>=16000 Kbps
12000		>= 12000 Kbps
8000		>= 8000 Kbps
4000		>= 4000 Kbps
1		>= 1 Kbps

Figure 14. Legend of capacity.

2.3.4 Optimisation for traffic requirements

Ranplan Professional provides the automatic cell optimisation (ACO) tool to optimise the network deployment to reduce blackspots and meet the traffic requirement of hotspots. Deployment optimisation traditionally considers several traffic-independent factors, such as propagation loss and interference. Optimisation, therefore, naturally assumes either a uniform user demand distribution or uses long-term statistics (i.e., census data) to spatially weight the optimisation. Moving beyond stationary radio planning is important as urban areas become more dynamically changing, more complex (tourists, commuters, changing urban landscape), and demand becomes more stochastic. In this section, ACO is used to optimise the network.

Ranplan Professional can create multiple key performance indicators (KPIs) with different weights to optimise the network deployment, such as reference signal received power (RSRP) for coverage, throughput for traffic, as shown in Figure 15, where we can set high weight to optimise for important KPI.

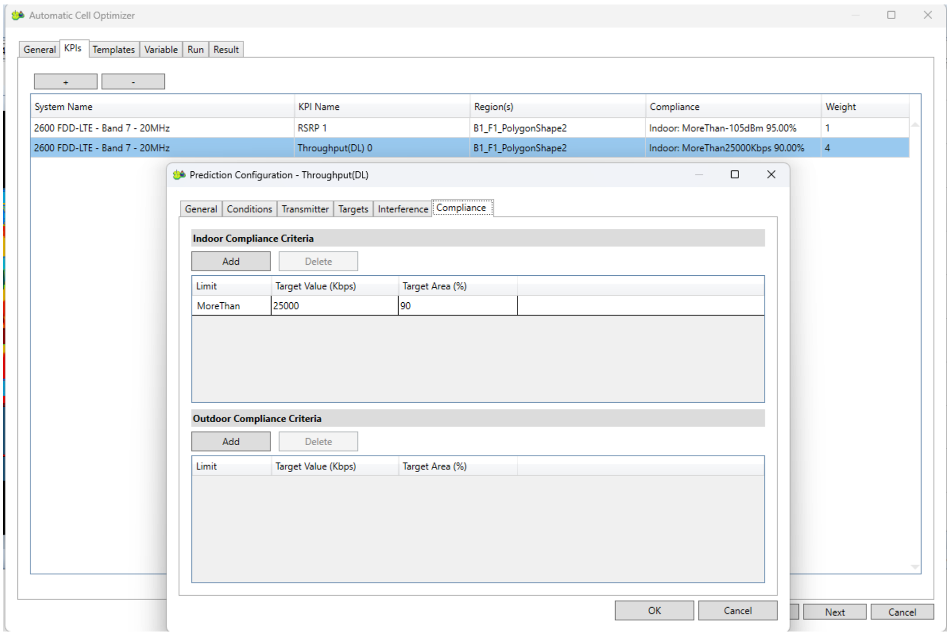


Figure 15. KPI configuration for different regions-.

In cellular communications, an outdoor base station is stationary installed on the top of buildings or towers. They are responsible for utilizing the radio waves as information carriers. Adequate transmission power should be set to enable that the users could receive the information correctly. A typical macro cell base station's transmission power is between 43 dBm and 48 dBm. Two or three antennas divide the covered region into diverse sectors. Each antenna has a particular azimuth. Generally, the configurations (power, downtilt, azimuth, and antenna's height, location) will influence the coverage. In this part, these configurations can be optimised in the ACO tool, as shown in Figure 16.

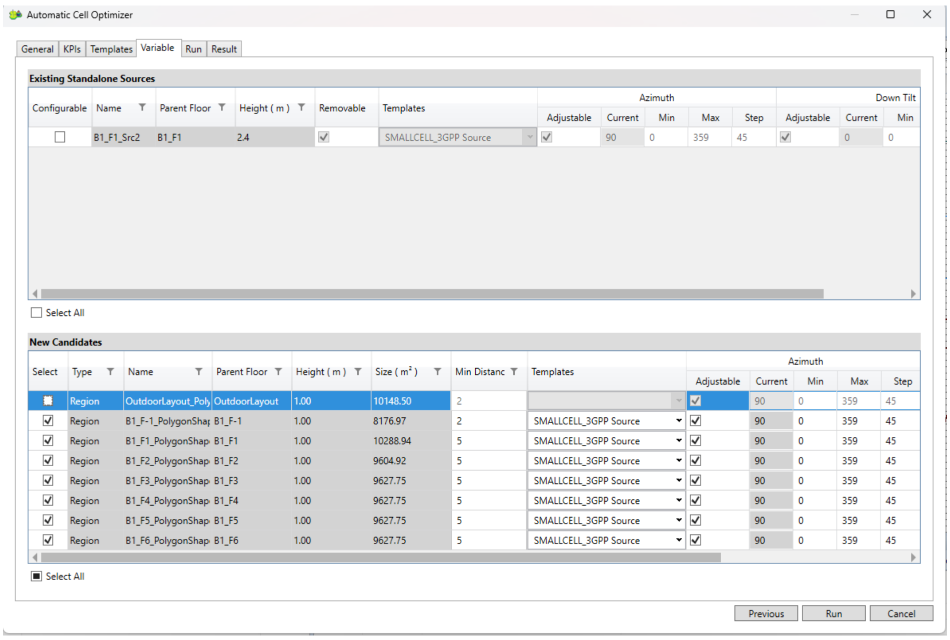


Figure 16. Variable optimisation in the ACO tool.

When running the optimisation module, ACO can optimise the network deployment with respect to different regions, different KPIs compliances, multiple variables, and different weights.

2.3.5 Performance comparison of capacity

Capacity can also quantify how the hotspot area is improved to meet the high traffic requirement. The capacity indicates the amount of data that can be transferred in a time unit, so it is an important indicator of the quality of service. Higher capacity can provide better user experience.

Figure 17 shows the results of base stations based on the traffic prediction, where only one snapshot traffic requirement is used to optimise the network. But the traffic pattern can be inputted into ACO module to optimise the network parameters automatically.

The following two figures, Figure 17 and Figure 18, show the capacity distribution. These results show the BS deployment can meet the high traffic requirement at the hotspot areas.

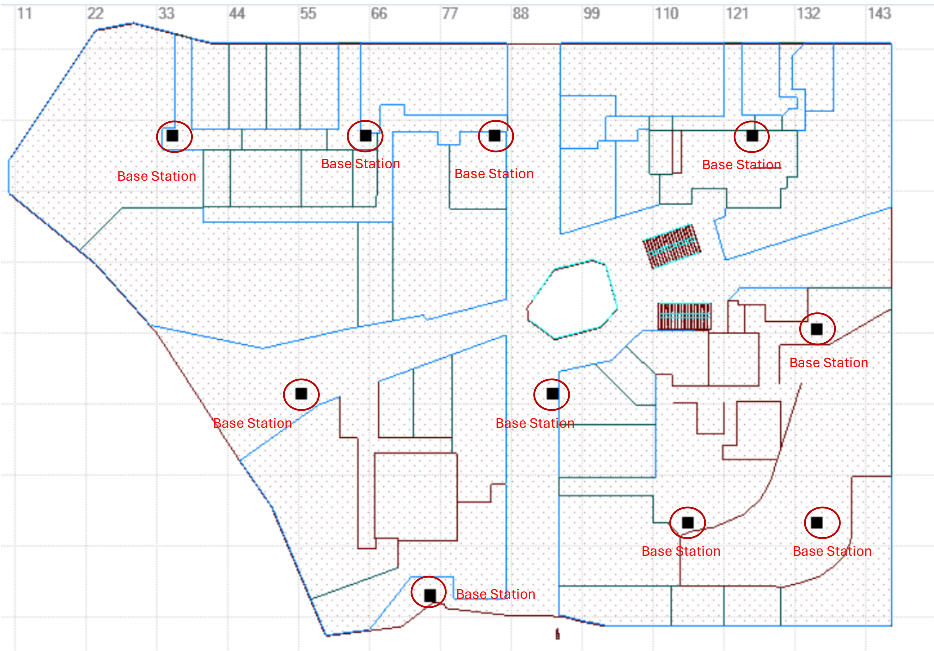


Figure 17. Optimised base station

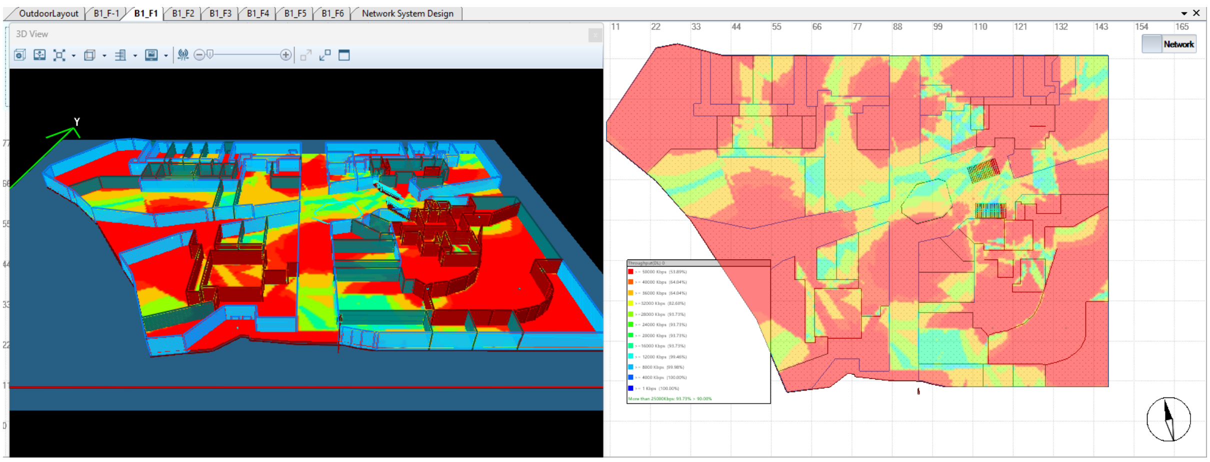
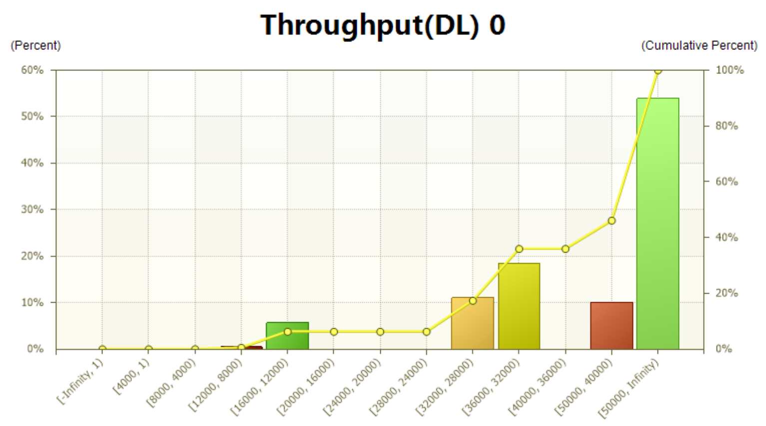


Figure 18. Capacity results.



[-Infinity, 1)	0.00%
[4000, 1)	0.00%
[8000, 4000)	0.02%
[12000, 8000)	0.52%
[16000, 12000)	5.73%
[20000, 16000)	0.00%
[24000, 20000)	0.00%
[28000, 24000)	0.00%
[32000, 28000)	11.13%
[36000, 32000)	18.56%
[40000, 36000)	0.00%
[50000, 40000)	10.15%
[50000, Infinity)	53.89%

Figure 19. Capacity distribution.

From the simulation results, based on the traffic pattern, Ranplan optimisation module can optimise the network deployment and RAN parameter to meet the capacity requirement.

3. Conclusions

This report uses artificial intelligent algorithm to identify the wireless spatial-temporal traffic of hotspots. Methods are developed to combine heterogeneous data with current state-of-the-art 3D urban indoor-outdoor propagation modeling (using Ranplan Professional), cost-benefit metrics ACO, heterogeneous network (HetNet) with load-driven interference modeling. The focus is on integrating and developing low complexity deployment optimisation algorithms based on automatic spatial-temporal traffic identification methods, and subsequent optimisation for improvement. Based on the spatial-temporal traffic pattern, wireless networks can adaptively adjust the network configuration and RAN parameters, and reduce the cost and improve the energy efficiency.