

# Evolutionary reinforcement learning multi-agents system for intelligent traffic light control: new approach and case of study

Mohamed Amine Basmassi<sup>1</sup>, Sidina Boudaakat<sup>2</sup>, Jihane Alami Chentoufi<sup>1</sup>, Lamia Benameur<sup>3</sup>,  
Ahmed Rebbani<sup>2</sup>, Omar Bouattane<sup>2</sup>

<sup>1</sup>Informatics, Systems and Optimization Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco

<sup>2</sup>Signals Distributed Systems and Artificial Intelligence Laboratory, Ecole Normale Supérieure de l'Enseignement Technique  
Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

<sup>3</sup>Information Technology and Systems Modelization Laboratory, Faculty of Sciences, Abdelmalik Essaadi University, Tétouan, Morocco

## Article Info

### Article history:

Received Jun 29, 2021

Revised May 28, 2022

Accepted Jun 25, 2022

### Keywords:

Adaptive traffic signal control

Fuzzy modeling

Genetic algorithm

Q-learning

Reinforcement learning

Traffic simulation

## ABSTRACT

Due to the rapid growth of urban vehicles, traffic congestion has become more serious. The signalized intersections are used all over the world and still established in the new construction. This paper proposes a self-adapted approach, called evolutionary reinforcement learning multi-agents system (ERL-MA), which combines computational intelligence and machine learning. The concept of this work is to build an intelligent agent capable of developing senior skills to manage the traffic light control system at any type of junction, using two powerful tools: learning from the confronted experience and the assumption using the randomization concept. The ERL-MA is an independent multi-agents system composed of two layers: the modeling and the decision layers. The modeling layer uses the intersection modeling using generalized fuzzy graph technique. The decision layer uses two methods: the novel greedy genetic algorithm (NGGA), and the Q-learning. In the Q-learning method, a multi Q-tables strategy and a new reward formula are proposed. The experiments used in this work relied on a real case of study with a simulation of one-hour scenario at Pasubio area in Italy. The obtained results show that the ERL-MA system succeeds to achieve competitive results comparing to urban traffic optimization by integrated automation (UTOPIA) system using different metrics.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Mohamed Amine Basmassi

Laboratory ISO, Faculty of Sciences, Ibn Tofail University, University Campus

Kenitra, BP 133, Morocco

Email: basmassi.med.amine@gmail.com

## 1. INTRODUCTION

The traffic congestion is one of the biggest problems related to big cities in the world. This problem can be frequented for several reasons such as peak periods and big events or at tourist destinations [1] which not only impacts people's travel but also limits the development of the urban economy. Traffic control is one of the important tools to contain the congestion, restrain traffic flow and reduce emissions. Many researchers have proposed adaptive traffic signal control (ATSC) to solve the traffic congestion problem using a variety of optimization techniques, such as heuristics [2]–[4], evolutionary algorithms [5], [6] and self-organization strategy [7].

Several researchers have studied this problem using different concepts. The ATSC has the potential to efficiently adjust signal timing in real time conforming to travel demand, weather or seasonal traffic fluctuations. Systems using this technology have outperformed the actuated control and pre-timed methods

[8]. The two most famous ATSC are the split cycle offset optimization technique (SCOOT) [9] and Sydney coordinated adaptive traffic system (SCATS). The optimization process of those systems is based on real time data and predictions to assign a signal timing plan. Other control system like optimization policies for adaptive control (OPAC) [10] and real time hierarchical optimized distributed effective system (RHODES) [11] using similar principle where they intervene into the fluctuation of the time variation by dynamically adjusting the signal timing parameters. Also, the urban traffic optimization by integrated automation (UTOPIA) [12] is a hierarchical decentralized traffic signal control strategy used in many countries. It focuses on optimizing the traffic flows and gives selective priority to public transport with taking into consideration the travel times of the private traffic [13].

Recently, a new generation of traffic control system starts the implementation of machine learning methods, where multiple optimization and prediction techniques have been beneficial to create more powerful system with the ability to solve complex traffic condition issues [14]. Reinforcement learning (RL) is a third machine learning paradigm that is used for self-learning traffic signal control in the stochastic road traffic environment [15], [16]. Frequent studies have proved the capabilities to solve the ATSC problem using the Q-learning which is a RL method [15].

In this perspective, this paper presents a new approach called evolutionary reinforcement learning multi-agents system (ERL-MA) which is an independent multi-agents system composed of two layers. The first one is the modeling layer that uses the information provided by the intersection modeling using generalized fuzzy graph (IMGFG) [17] to understand the junction constraint and complexity. The second one is the decision layer which is composed of two activities: activity per cycle and activity per phase. In fact, the novel greedy genetic algorithm (NGGA) [18] is used in the activity per cycle to compose the optimal sequence of phases, and the accumulated knowledge of the Q-learning is applied to generate flexible signal traffic light timing in the activity per phase.

A real-world scenario from Bologna city, built in the project of iTERTRIS (an integrated wireless and traffic platform for real-time road traffic management solutions), was prepared and described by Bieker *et al.* [19]. This real-world scenario is used as a case of study, which gives the opportunity for the best evaluation and appropriate demonstration of the proposed approach. This paper is organized. In the next section, the background of this study is presented. In section 3, the proposed approach is detailed. The experimental results are shown in section 4 where a real case of study is simulated. Finally, a discussion and conclusion are given in section 5.

## 2. BACKGROUND

This section will cover the background of the material presented in the paper. The first step in optimizing the signal timing for traffic congestion is finding a powerful model; this model should present the junction's influencing parameters and display their correlation. The traffic congestion problem in signalized junction (or intersection) depends on traffic fluctuations, green timing, cycle length, phasing. Therefore, an efficient system should have a powerful model to present the influencing parameters of the junction. Modeling traffic behavior has been for years interesting issue research, traditionally accomplished using a variety of methods. As queuing theory [20], Dotoli and Fanti [21], cell transmission model [22] and intersection graph [17], [23], [24]. Boudaakat *et al.* [17] have introduced the intersection modeling IMGFG, where a generalized fuzzy graph concept with fuzzy vertices and fuzzy edges, through the modeling of an isolated intersection in two stages, steady elements in the first stage which leads to fuzzy graph with crisp vertices and a situational status which leads to a generalized fuzzy graph. In this work, the IMGFG is adopted as the modeling technique because this approach has the potential to present precisely the congestion points and conflicts level.

Reinforcement learning has been proved to be an effective method for developing adaptive traffic signal controllers by adjusting signal timing in response to traffic fluctuations [25], [26]. The objectives of the Q-learning which is a technique of model-free reinforcement learning algorithm are gain experience to make better decisions. However, regarding all the advantages of the existing ATSC, the safety impact of applying these methods is still unclear. Some studies showed that implementing ATSC algorithms leads to lower traffic safety and extends traffic conflicts significantly [27] or a minor reduction on traffic collisions [28]. The mobility optimization does not necessarily assure the safety optimization, so discounting the traffic safety as a main objective in the existing ATSC is probably responsible of the inconsistency in the safety impact [29]. Therefore, the proposed ATSC is a new approach combining computational intelligence and reinforcement learning; it adjusts optimal signal timing in real-time and ensures the appropriate phasing with maximum traffic safety.

**3. ERL-MA APPROACH**

The proposed approach is an independent multi-agents system, where each agent supervises a traffic light system (composed of mono-junction or multi-junction). The agent is designed to solve the traffic light control problem in such a way that fits with all junction formats (with multi-lanes or mono-lanes edges, directed and undirected lanes) and adapt traffic lights to the instantaneous flow changes. As depicted in Figure 1, the agent, in the proposed approach, consists of two main layers. The IMGFG modeling layer is responsible for modeling and transforming the junction to an adjacency matrix and provides the necessary parameters. The second layer is a decision part divided into two activities: activity per cycle and activity per phase. In the activity per cycle, the NGGA is applied to construct homogeneous clusters with the maximum traffic safety and fluidity depending on the congestion level and the traffic demand density. In the activity per phase, the Q-learning algorithm is employed to associate the adequate signal timing. The agent takes into account the occurred changes at the intersection in every cycle and phase to manage the intersection.

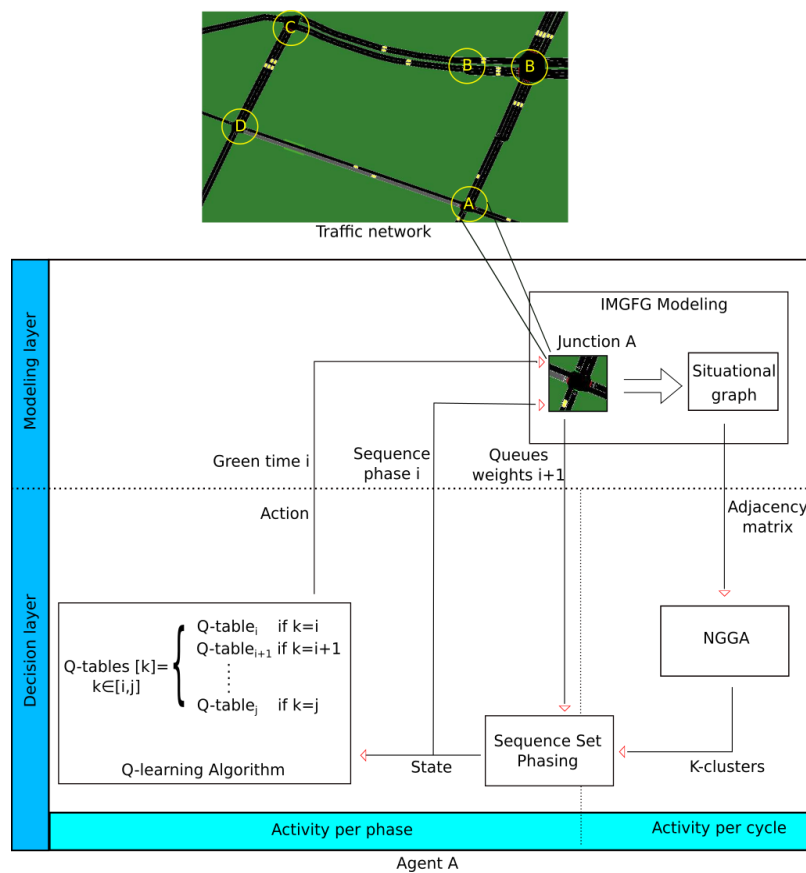


Figure 1. The ERL-MA system

**3.1. The modeling layer**

The modeling layer uses the IMGFG technique [17] where a generalized fuzzy graph coloring approach is used to model any signalized intersection. All possible conflicts between outgoing lanes and the congestion level in lanes at the intersection are presented in this generalized fuzzy graph called the situational graph. In this generalized graph, the vertices represent lanes, the vertices weight are the level of congestion, the edges express the existing conflicts between those lanes, and the edges weight are the kinds of this conflicts. The adjacency matrix represents the situational graph that will be sent to next layer. The decision layer is divided into two activities: activity per cycle and activity per phase.

**3.2. The decision layer**

**3.2.1. The activity per cycle**

In this activity, the NGGA combines genetic and greedy algorithm. The genetic algorithm is selected as the framework that explores the search space. Also, the rate at which the algorithm can explore a space of possibilities is satisfying and adequate to combinatorial problems. The greedy algorithm is added as

a function in the genetic algorithm to accelerate the search and improve chromosomes fitness [18]. In fact, the NGGA was implemented on fuzzy graphs and has proved its efficiency to this kind of problem [30]. In the proposed approach, the algorithm takes as an input the adjacency matrix generated from the IMGFG, and provides at the output an optimal combination of lanes (k-clusters) that ensures a high flux with the maximum safety. These clusters present phases that the agent will apply at the intersection during this cycle. After that, the agent sorts these phases referring to the average queue length and the waiting time of each phase to produce a sequence phasing set for the next step.

### 3.2.2. The activity per phase

In this activity, the agent uses the Q-learning algorithm to decide the green timing period for every upcoming phase sequence. The Q-learning strategy adopted in this work uses a set of Q-tables, where every Q-table concerns a cycle format. A Q-table is a matrix [31], where each row represents a specific state (intersection situation) and each column represents a specific action (green time). When the agent decides to apply a cycle with k-phases, the  $Q\text{-table}_k$  is called to give the adequate green timing (action) for the associated phasing sequence (state). The agent compares all the Q-values of the requested state and the action with the highest Q-value is chosen. The characteristics of this activity are:

**Search space:** In the signalized traffic control problem, the space of states and actions can be very large depending on the parameters used to present them. The search space grows exponentially when the states are numerous which impacts the Q-learning knowledge space (Q-learning matrices). Bigger the search space is less parts of the Q-matrix are explored. In the proposed approach, the Q-tables number varies from an intersection to another where the maximum and minimum numbers of phases are the first parameters to define in order to know how many Q-tables exist. In order to reduce the number of states describing all possible situations generated by the NGGA, states are represented with normalized forma where one state can describe a number of similar situations in every Q-table. Also, actions selected at every Q-table are limited to the maximum phasing time.

**State definition:** The general format of a state at the search space is [K, CP, AQW table], where K represents the number of phases in the cycle, CP the current phase index and AQW table the average of the queue weight at every phase. The states existing in every  $Q\text{-table}_k$  are represented with [CP, AQW table], where the average queue weight values are limited at  $\emptyset$ , low (L), medium (M) and high (H) and the average queue weight table size depends on the number of phases used at the sequence phasing, see Figure 2.

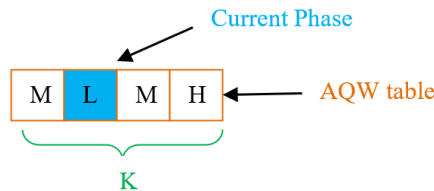


Figure 2. Illustration of a state in cycle with four phases

**Action definition:** Action presents green time phasing. For every state in the knowledge matrices (Q-tables) different timing [min phase time, max phase time] are subject of test in the learning stage. The max green time per phase differs from a  $Q\text{-table}_k$  to another and depends on the max cycle time and the number of phases (k) in the cycle. The best-rated actions are actions with the max benefit for the current phase and the less negative impact on the intersection.

**Reward definition:** Any action applied on a phase has a positive impact on the related lanes and a negative impact on the other lanes. In order to get the action with the max benefit to the intersection, this discrimination is necessary to balance action/effect. The reward is calculated in (1) based on the effect of the selected action in all lanes of the intersection, where the degree of passed vehicles and the queue weight change are used as a metric parameter for lanes belonging to the current phase; for the lanes belonging to other phases, the average of the cumulative waiting time and the occurring queue weight are used.

$$R_i = PV_i * (1 + (QWB_i - QWA_i)) - \frac{\sum_{j=1, j \neq i}^{NbrPhases} QWA_j * (WTA_j - WTB_j)}{(NbrPhases - 1) * A_i} \quad (1)$$

Where,  $PV_i$  is the degree of passed vehicles at the current phase.  $QWB_i$  is average queue weight of the current phase before the action.  $QWA_i$  is average queue weight of the current phase after the action.  $QWA_j$  is average

queue weight of the  $j$  phase after the action.  $WTA_j$  is cumulative waiting time of the  $j$  phase after the action.  $WTB_j$  is cumulative waiting time of the  $j$  phase before the action.  $A_i$  is current action.

Q-values and coefficient: The Q-table in (2) defines the knowledge space where every Q-table $_k$  presents a sub-space with a common format;  $k$  is the number of phases presenting the state.

$$Q\text{-table}(k) = \begin{cases} Q\text{-table}_i, & k = i \\ Q\text{-table}_{i+1}, & k = i + 1 \\ \vdots \\ Q\text{-table}_j, & k = j \end{cases} \quad (2)$$

The Q-learning uses the experience of each state transition to update one element of a Q-table $_k$ . A Q-table $_k$  is a matrix in which each row represents a specific state and each column represents a specific action. Each cell in this matrix represents a Q-value for a specific state-action pair  $Q_k(s, a)$  [32]. The Q-value, in general, is used to compare various actions at a specific state. The Q-learning algorithm improves its policy by updating the Q-table $_k$  in (3) according to Bellman's equation [33].

$$Q_k^{t+1}(s^t, a^t) = Q_k^t(s^t, a^t) + \alpha^{t+1} [r^{t+1} + \gamma \max_{a'} Q_k^t(s^{t+1}, a^{t+1}) - Q_k^t(s^t, a^t)] \quad (3)$$

Where  $s^t, a^t$  is the current state and the selected action at the current state.  $Q^{t+1}, Q^t$  is the updated and the old Q-value.  $r^{t+1}$  is the reward of applying action at state  $s^t$ .  $s^{t+1}, a^{t+1}$  is the new state and the best action at the new state.  $\alpha^{t+1}$  is the learning rate.  $\gamma$  is the discount rate.

Training stage: Before testing the ERL-MA system, a learning phase is mandatory. Every agent undergoes a training stage to acquire the knowledge of behaving with the intersection, and be able to assign the adequate action to a current situation taking into consideration the structure of the intersection and impact of the action on it. The policy matrices are set to zeros; then, the agent passes to learning stage where it is exposed to a random generated scenario combining low, medium and high flows in order to access at the maximum state-actions component of the Q-tables. Also, the action selection is subject to a noise coefficient to stimulate the exploration. In result, the agent becomes ready for the evaluation with optimal policy matrices. The agent does not stop upgrading the optimal policy matrix in this stage, but is able to become greater with the future experience.

## 4. EXPERIMENTAL RESULTS

### 4.1. Simulation environment

In this paper, simulation of urban mobility (SUMO) [34] is used as the simulation environment. It is a free and open-source microscopic traffic simulation. It has been available since 2001 and allows modeling of intermodal traffic systems including road vehicles, pedestrians and public transport; it is a tool widely used for traffic research.

### 4.2. Testbed network

Traffic light signal plans are rarely open to the public and are often not available in digital format. To replicate a part of a real road network, gathering, converting, and adapting all the data is time-consuming. Also, the correction and the validation of the responsible municipality of the studied area are hardly possible and mandatory to allow performing real-world evaluations and fair comparisons. Real-world scenarios from Bologna built in the project of iTERRIS (co-funded by the European Commission and contributed by the municipality of Bologna as a project partner) where all the previous conditions are respected, prepared and available to the public within the SUMO package by Bieker *et al.* [19]. The proposed approach is applied on the network Pasubio area in Bologna in Italy as shown in Figure 3.

The Pasubio scenario extends the scenarios by the area around the hospital and includes also common routes to the football stadium as shown in Figure 4. Due to the situation and the traffic problems in Bologna, the municipality of Bologna delivered a large set of data and simulation scenarios. The given data included representations of the areas around the Pasubio roads, as input files for the commercial microscopic traffic simulation. The scenarios modeled the peak hour in Bologna (8:00–9:00 am) [19].

The congestion level in Bologna in 2019 is 25% and 205 ranked in the world, the congestion level by road type in highways is 17% and in non-highways is 31%. The real-world scenarios from Bologna were prepared and are described under project iTETRIS which is an integrated wireless and traffic platform for real-time road traffic management solutions to help estimate road traffic engineering. The scenario of Bologna traffic was built to illustrate the traffic congestion in both areas Pasubio and Andrea Costa.

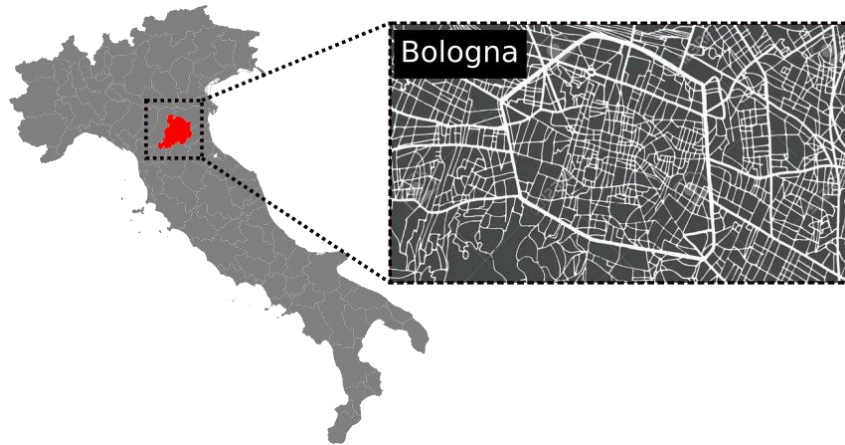


Figure 3. Location of Bologna in Italy

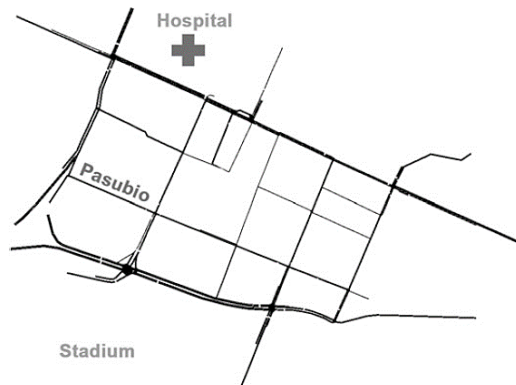


Figure 4. Pasubio traffic network in SUMO

The Pasubio scenario of Bologna city in Italy [19] presented in Figure 4 was chosen to test the proposed approach. The selected area represents 2.45 km<sup>2</sup> of a real city with a total of eight traffic lights, sixty-five nodes and one hundred eleven edges. The scenario displays the real traffic at the morning peak hour in 8:00-9:00 am. The city of Bologna uses the UTOPIA system for traffic light control. UTOPIA [35] optimizes traffic light schedules and sorts the traffic light phases to satisfy traffic demand.

#### 4.3. Simulation

The tested area is composed of 8 traffic lights control system. Some of them use multi junctions (intersections) with one traffic lights control system. Every traffic light control system (TLC) is supervised by an agent as shown in Table 1. Every agent undergoes the training stage separately until it acquires the necessary skills. As shown in Figure 5, the Junctions 4 and 14 are controlled by the TLC 230 and the agent 1 supervise it. The accumulated knowledge and the scenarios exposed during the learning stage for agent 1 are presented in Figures 6 and 7. All the others agents of the network are exposed to the same scenarios.

Table 1. Junctions and controlling agents

| Agent | TLC Id | Junctions Ids |
|-------|--------|---------------|
| 1     | 230    | 14, 4         |
| 2     | 231    | 9, 10, 12     |
| 3     | 232    | 29, 27        |
| 4     | 233    | 15            |
| 5     | 220    | 36            |
| 6     | 219    | 1, 32         |
| 7     | 282    | 18            |
| 8     | 218    | 0, m0         |

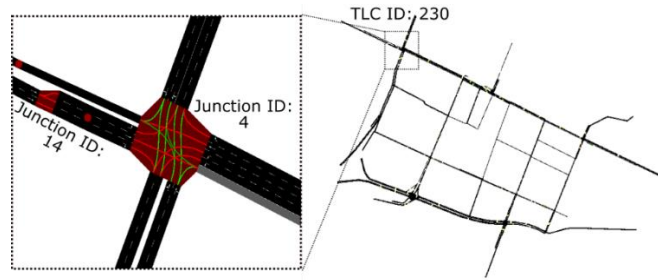


Figure 5. Illustration of junctions with the associated TLC

Agent one is exposed to a random generated scenario combining low, medium and high flows in order to access at the maximum state-actions component of the Q-tables, Figure 6 show the quantities of the new knowledge acquired at the end of every scenario; in the beginning of the learning phases the quantity of the new knowledge is massive because the agent is uninformed, with time the new knowledge accumulated per scenario get lower until it become insignificant, in this point the optimal policy is obtained and the agent is ready for evaluation. Figure 7 show the associated running time to all the random scenarios applied to the agent 1 at the learning stage.

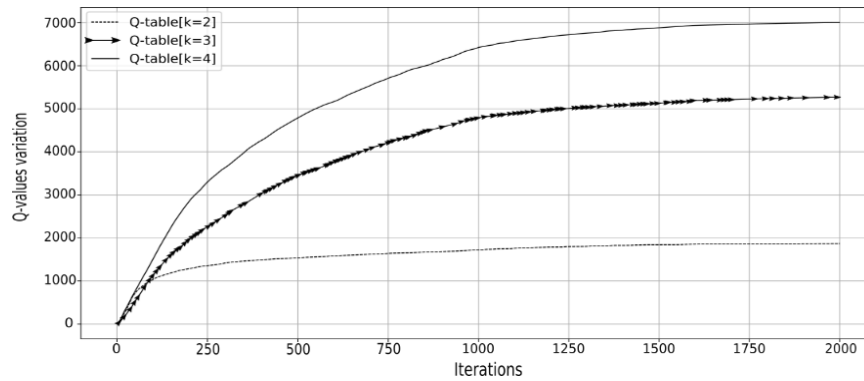


Figure 6. The acquired knowledge variation in the learning phase

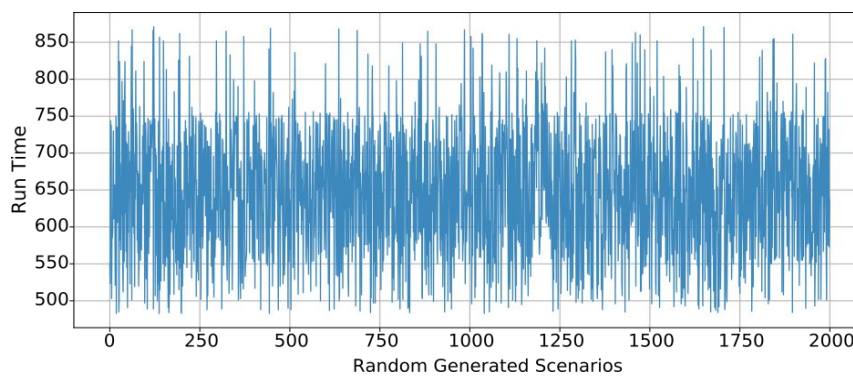


Figure 7. The running time of random generated scenarios in the learning phase

#### 4.4. Approach evaluation

A real-world scenario will evaluate the proposed approach. The network area is 2.45 m<sup>2</sup> with eight traffic lights, 8,776 loaded vehicles, and different types of signalized intersections; the proposed system is compared to the UTOPIA system implemented at the Pasubio network. After, applying the learning phase to all agents controlling the TLC in the Pasubio area, the evaluation of the same scenario provided as data test

were run in 50 simulations to evaluate the performance of the proposed approach comparing the UTOPIO system. Figure 8 shows that the proposed approach exceeds the UTOPIO system in 100% with a performance of [6.60%, 12.44%] less time. The UTOPIO system finishes the simulation in 4,605 sec and the proposed approach running time varies between 4,032 and 4,301 sec; this non-stability is due to the characteristic of the NGGA which is a stochastic algorithm. The effect of the randomness characteristic can be seeing at the first look as a weakness, but with a deep look it should be seeing as a powerful tool that distinguishes this approach, having this characteristic mean that agent is always able to benefice from new experiences and develop new skills to manage the traffic light control system.

The waiting time graphs results show that the average waiting time was controlled by the ERL-MA perfectly compared to UTOPIO. Agent one at junction four and agent eight at junction zero found some difficulty controlling the waiting time. However, generally, they succeeded in handling it compared to UTOPIO, as shown in Figure 9. Likewise, the simulation results presented in Figure 10 show that ERL-MA agents have improved the queue length of all junctions comparing to the UTOPIO system.

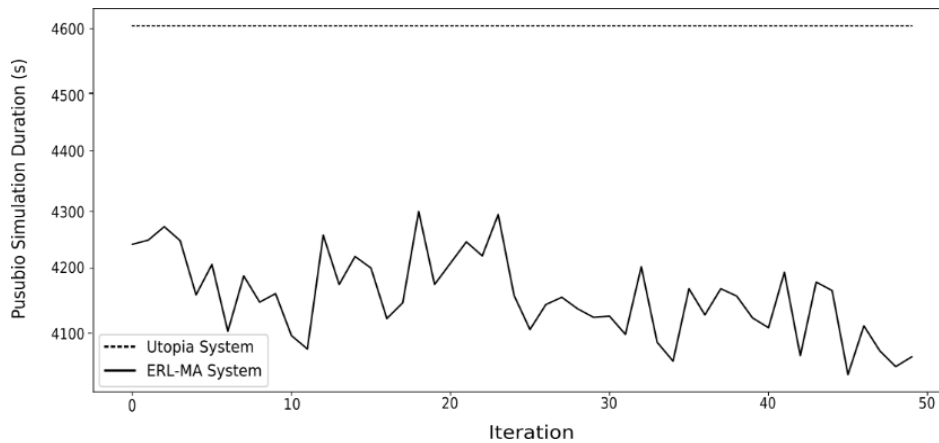


Figure 8. Illustration of simulation duration

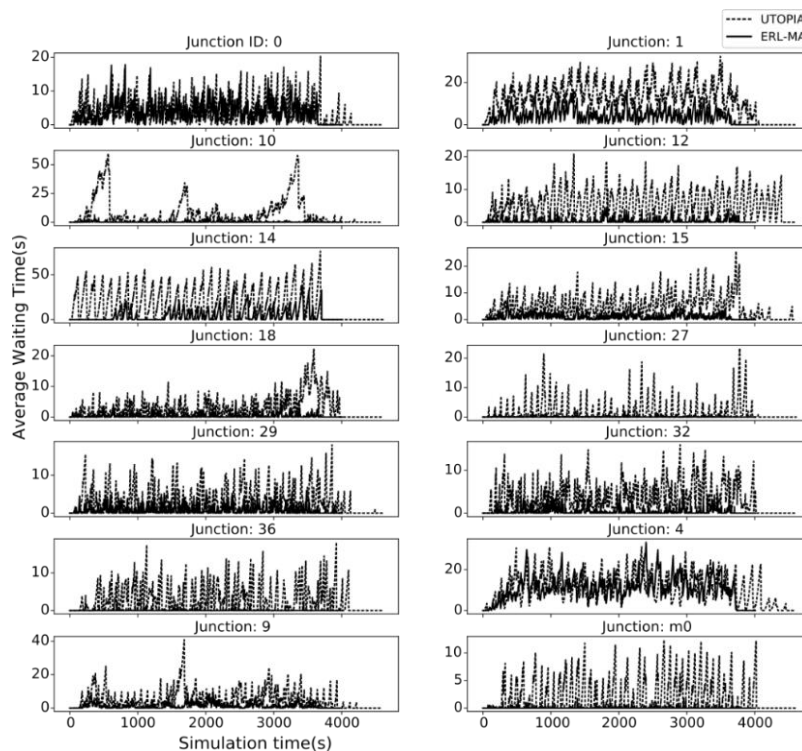


Figure 9. Illustration of the average waiting time at the Pasubio junctions



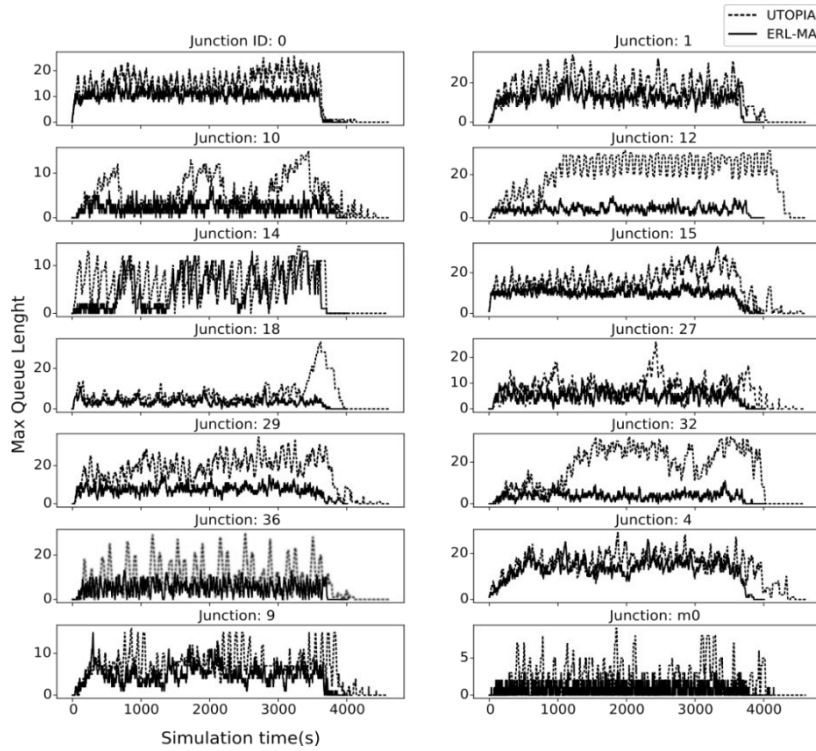


Figure 10. Illustration of the max queue length at the Pasubio junctions

As a result of the above graphs, the CO<sub>2</sub> and CO have been significantly reduced. Due to the shorted waiting time and accumulated queue length of vehicles. Also, the lower simulation time have contributed at CO<sub>2</sub> and CO reducing. The ERL-MA approach reduced CO<sub>2</sub> and CO emissions produced by vehicles, as depicted in Figure 11.

The experiments results determined that ERL-MA system can generate a flexible traffic signal timing to traffic variation for the Pasubio scenario and different intersection forma. The proposed approach was tested using a real-world scenario. The approach was compared with the implemented system at the Pasubio network. The ERL-MA agent use the information provided by the IMGFG to understand the intersection constraint, the NGGA to compose the optimal sequence phasing, and the accumulated knowledge of the Q-learning to generate a flexible signal traffic light timing. However, the model can certainly be extended to consider different aspects such as the communication between agents or between lights and vehicles. The ERL-MA has the ability to obtain excellent solution for traffic light control problem.

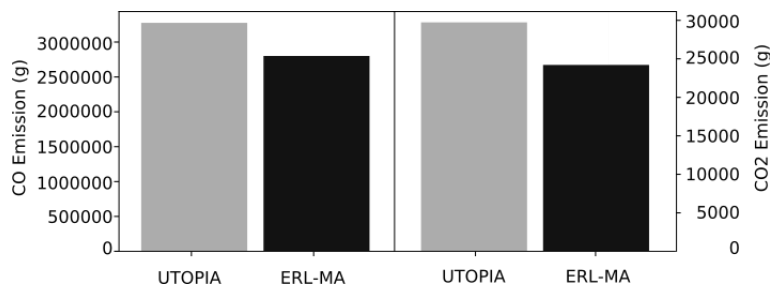


Figure 11. Illustration of CO and CO<sub>2</sub> emission of vehicles

### 5. CONCLUSION

The proposed approach uses two powerful tools of human beings which are learning and assumption to build an intelligent agent capable of managing the traffic light control system. Agents are designed to fit with junction format and have the potential to adjust signal timing conforming to the instantaneous changes. The proposed approach does not request expensive changes or massive conditions demanding. Furthermore,

it does not impose changes at the infrastructure of the intersection or requires direct communication with vehicles. This approach was designed to fit and self-adapt with intersection infrastructure and TLC. Computational intelligence, machine learning approach, and fuzzy graph modeling are used to build this approach. The proposed system ERL-MA was evaluated with a real-world scenario and compared to the existing system. The obtained results showed that the proposed approach succeeded to achieve competitive results using different metrics.





## REFERENCES

- [1] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting traffic congestion using ARIMA modeling," in *2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Jun. 2019, pp. 1227–1232, doi: 10.1109/IWCMC.2019.8766698.
- [2] T. Wongpiromsarn, T. Uthacharoenpong, Y. Wang, E. Frazzoli, and D. Wang, "Distributed traffic signal control for maximum network throughput," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, Sep. 2012, pp. 588–595, doi: 10.1109/ITSC.2012.6338817.
- [3] P. Varaiya, "The max-pressure controller for arbitrary networks of signalized intersections," in *Complex Networks and Dynamic Systems*, Springer New York, 2013, pp. 27–66.
- [4] J. Gregoire, X. Qian, E. Frazzoli, A. de La Fortelle, and T. Wongpiromsarn, "Capacity-aware backpressure traffic signal control," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 2, pp. 164–173, Jun. 2015, doi: 10.1109/TCNS.2014.2378871.
- [5] I. G. P. S. Wijaya, K. Uchimura, and G. Koutaki, "Traffic light signal parameters optimization using modification of multielement genetic algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 1, pp. 246–253, Feb. 2018, doi: 10.11591/ijece.v8i1.pp246-253.
- [6] A. Louati, S. Darmoul, S. Elkosantini, and L. ben Said, "An artificial immune network to control interrupted flow at a signalized intersection," *Information Sciences*, vol. 433–434, pp. 70–95, Apr. 2018, doi: 10.1016/j.ins.2017.12.033.
- [7] R. N. Rithesh, R. Vignesh, and M. R. Anala, "Autonomous traffic signal control using decision tree," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 3, pp. 1522–1529, Jun. 2018, doi: 10.11591/ijece.v8i3.pp1522-1529.
- [8] W. R. McShane, R. P. Roess, and E. S. Prassas, *Traffic engineering*. Prentice Hall, 1998.
- [9] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton, "SCOOT—a traffic responsive method of coordinating signals," United Kingdom, 1981.
- [10] N. H. Gartner, P. J. Tarnoff, and C. M. Andrews, "Evaluation of optimized policies for adaptive control strategy," *Transportation Research Record 1324*, pp. 105–114, 1991.
- [11] P. Mirchandani and L. Head, "A real-time traffic signal control system: architecture, algorithms, and analysis," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 6, pp. 415–432, Dec. 2001, doi: 10.1016/S0968-090X(00)00047-4.
- [12] V. Mauro and C. Di Taranto, "UTOPIA," *IFAC Proceedings Volumes*, vol. 23, no. 2, pp. 245–252, Sep. 1990, doi: 10.1016/S1474-6670(17)52678-6.
- [13] D. Pavleski, D. Koltovska-Nechoska, and E. Ivanjko, "Evaluation of adaptive traffic control system UTOPIA using microscopic simulation," in *2017 International Symposium ELMAR*, Sep. 2017, pp. 17–20, doi: 10.23919/ELMAR.2017.8124425.
- [14] S. Araghi, A. Khosravi, and D. Creighton, "A review on computational intelligence methods for controlling traffic signal timing," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1538–1550, Feb. 2015, doi: 10.1016/j.eswa.2014.09.003.
- [15] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Design of reinforcement learning parameters for seamless application of adaptive traffic signal control," *Journal of Intelligent Transportation Systems*, vol. 18, no. 3, pp. 227–245, Jul. 2014, doi: 10.1080/15472450.2013.810991.
- [16] S. M. A. Shabestary and B. Abdulhai, "Deep learning vs. discrete reinforcement learning for adaptive traffic signal control," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 286–293, doi: 10.1109/ITSC.2018.8569549.
- [17] S. Boudaakat, M. A. Basmassi, A. Rebbani, J. Alami Chentoufi, L. Benameur, and O. Bouattane, "Intersection modeling using generalized fuzzy graph coloring," in *Innovations in Smart Cities Applications Volume 4*, 2021, pp. 1479–1489.
- [18] M. A. Basmassi, L. Benameur, and J. A. Chentoufi, "A novel greedy genetic algorithm to solve combinatorial optimization problem," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 117–120, Nov. 2020, doi: 10.5194/isprs-archives-XLIV-4-W3-2020-117-2020.
- [19] L. Bieker, D. Krajzewicz, A. Morra, C. Michelacci, and F. Cartolano, "Traffic simulation for all: a real world traffic scenario from the City of Bologna," in *Modeling Mobility with Open Data*, Springer International Publishing, 2015, pp. 47–60.
- [20] T. S. Babicheva, "The use of queuing theory at research and optimization of traffic on the signal-controlled road intersections," *Procedia Computer Science*, vol. 55, pp. 469–478, 2015, doi: 10.1016/j.procs.2015.07.016.
- [21] M. Dotoli and M. P. Fanti, "An urban traffic network model via coloured timed Petri nets," *Control Engineering Practice*, vol. 14, no. 10, pp. 1213–1229, Oct. 2006, doi: 10.1016/j.conengprac.2006.02.005.
- [22] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, Aug. 1994, doi: 10.1016/0191-2615(94)90002-7.
- [23] I. Rosyida, Widodo, C. R. Indrati, and D. Indriati, "Some properties on fuzzy chromatic number of union of fuzzy graphs through  $\alpha$ -cut graphs coloring," *Journal of Physics: Conference Series*, vol. 1306, no. 1, Aug. 2019, doi: 10.1088/1742-6596/1306/1/012042.
- [24] S. Munoz, M. Teresaortuno, J. Ramirez, and J. Yanez, "Coloring fuzzy graphs," *Omega*, vol. 33, no. 3, pp. 211–221, Jun. 2005, doi: 10.1016/j.omega.2004.04.006.
- [25] L. Prashanth and S. Bhatnagar, "Reinforcement learning with function approximation for traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 412–421, Jun. 2011, doi: 10.1109/TITS.2010.2091408.
- [26] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (MARLIN-ATSC): methodology and large-scale application on downtown Toronto," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1140–1150, Sep. 2013, doi: 10.1109/TITS.2013.2255286.
- [27] A. Tageldin, T. Sayed, M. H. Zaki, and M. Azab, "A safety evaluation of an adaptive traffic signal control system using computer vision," *Advances in Transportation Studies*, vol. 2, pp. 83–96, 2014.





- [28] Z. H. Khattak, M. D. Fontaine, B. L. Smith, and J. Ma, "Crash severity effects of adaptive signal control technology: An empirical assessment with insights from Pennsylvania and Virginia," *Accident Analysis and Prevention*, vol. 124, pp. 151–162, Mar. 2019, doi: 10.1016/j.aap.2019.01.008.
- [29] H. Tang, V. V. Gayah, and E. T. Donnell, "Crash modification factors for adaptive traffic signal control: an empirical Bayes before-after study," *Accident Analysis and Prevention*, vol. 144, Sep. 2020, doi: 10.1016/j.aap.2020.105672.
- [30] M. A. Basmassi, S. Boudaakat, L. Benameur, O. Bouattane, A. Rebbani, and J. A. Chentoufi, "Hybrid genetic approach for solving fuzzy graph coloring problem," in *Sensor Network Methodologies for Smart Applications*, IGI Global, 2020, pp. 54–64.
- [31] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3/4, pp. 279–292, 1992, doi: 10.1023/a:1022676722315.
- [32] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning*. Cambridge: MIT press, 1998.
- [33] M. Essa and T. Sayed, "Self-learning adaptive traffic signal control for real-time safety optimization," *Accident Analysis and Prevention*, vol. 146, Oct. 2020, doi: 10.1016/j.aap.2020.105713.
- [34] P. A. Lopez *et al.*, "Microscopic traffic simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 2575–2582, doi: 10.1109/ITSC.2018.8569938.
- [35] V. Mauro and C. Di Taranto, "UTOPIA," in *Control, Computers, Communications in Transportation*, Elsevier, 1990, pp. 245–252.

## BIOGRAPHIES OF AUTHORS







**Mohamed Amine Basmassi**     is currently Ph.D. student in the informatics, systems and optimization laboratory in Ibn-Tofail University, Kenitra, Morocco. In 2015, he received the M.S degree in Computer Engineering from the Abdelmalek Essaadi University, Tetouan, Morocco. In 2012, he had the B.S. degree in fundamental mathematics and computer science from Cadi Ayyad University, Marrakech-Safi, Morocco. His research interests fall in machine learning, heuristic techniques, computational intelligence and multi-agent system to solve problems related to transport and traffic management, combinatorial optimization and decision-making problems. He can be contacted at email: basmassi.med.amine@gmail.com.







**Sidina Boudaakat**     received the B.S. degree from Mohamed V University of Rabat, Morocco in 2013, and M.S degree in Abdelmalek Essaadi University, Tetouan, Morocco, in 2015. Since 2016 he has been working toward the Ph.D. degrees on Computational Intelligence in Urban Traffic to solved problems related to the management of road traffic and road safety at Laboratory SSDIA in Normal School of Technical Education (NSTE), Mohammadia, Morocco. He can be contacted at email: boudaakat.sidina@gmail.com.






**Jihane Alami Chentoufi**     is a Professor and Ph.D. supervisor in the Research on Computer Science Laboratory in Ibn Tofail University, Faculty of Sciences-Kenitra. She received her M.S. and Ph.D. degrees in Computer Sciences from Mohammed V University, Faculty of Sciences-Rabat, Morocco in 2003 and 2007, respectively. Her main research interests include metaheuristics, combinatorial optimization and decision-making problems. She can be contacted at: j.alami@uit.ac.ma.






**Lamia Benameur**     is a Professor and Ph.D. supervisor in the Research on Information Technology and systems modelization Laboratory in Abdelmalik Essaadi University, Faculty of Sciences-Tetouan. She received her M.S. and Ph.D. degrees in Computer Sciences from Mohammed V University, Faculty of Sciences-Rabat, Morocco in 2003 and 2010, respectively. Her main research interests include metaheuristics, combinatorial optimization and decision-making problems. She can be contacted at: l.benameur@uae.ac.ma.



**Ahmed Rebbani**    he received the B.S. degree in Electronics in 1988, the M.S. degree in Applied Electronics in 1992 from the ENSET Institute, Mohammedia, Morocco. He received the DEA diploma in information processing in 1997 from the faculty of sciences Ben Msik, Casablanca, Morocco. has his Ph.D. degree in 2014 In electrical energy storage from the faculty of sciences and techniques Mohammedia, Morocco. He is now a teacher and researcher at the University Hassan II, ENSET Institute. His research is focused on Internet of things and renewable energy. He can be contacted at email: a.rebbani@gmail.com.



**Omar Bouattane**    he received his PhD from the University Hassan II of Casablanca, MOROCCO in 2001 in Parallel Computing and Image processing. He currently serves as a full Professor in the Department of Electrical Engineering at “*Ecole Normale Supérieure de l’enseignement technique*” ENSET of Mohammedia. He has more than 150 scientific publications in various domains of Computational Intelligence, high performance computing, image processing and renewable energy. He has registered 6 national and PCT international Patents regarding to the networking technology and signal synthesis. He was awarded as the owner of the best PCT patent in Morocco on 2011. Since 2012, He was the head of the laboratory of Signals Distributed Systems and Artificial Intelligence. He involved his laboratory in several partnership activities and developed many funded projects in Morocco and in his university. Overall, Prof. Bouattane work has received more than 300 citations. Prof. Bouattane has been the principal investigator and leader in 4 academic projects, funded either publicly or privately, in the USA, Canada and France. He was the supervisor from Morocco of a partnership program entitled “Linkage for entrepreneurship achievement program” funded by the USAID and HED of USA from December 2012 to December 2014. He supervised beside his US partners a scholarship program named “study abroad, Moroccan culture” December 2016 to December 2018. All his academic and research activities are in the ResearchGate portal at: [https://www.researchgate.net/profile/Omar\\_Bouattane/contributions](https://www.researchgate.net/profile/Omar_Bouattane/contributions). He can be contacted at email: o.bouattane@gmail.com.