Reinforcement Learning (Q-LEARNING) traffic light controller within intersection traffic system

Saif Islam Bouderba
bouderba.s@ucd.ac.ma
LAROSERI, Department of Computer Science,Faculty of Science, University of Chouaib Doukkali
EL Jadida, Morocco

Najem Moussa
moussa.n@ucd.ac.ma
LAROSERI, Department of Computer Science,Faculty of Science, University of Chouaib Doukkali
EL Jadida, Morocco

ABSTRACT

In this paper we study the effect of signalized traffic intersection control strategies in a cellular automaton model for transportation in urban networks. Starting with a simple synchronized strategy, then a green wave which gave a surprising result. Finally, a reinforcement learning approach (Q-LEARNING) is presented to learn the traffic light controller how to interact with drivers with different situation. By keeping a belief the level of cooperation drivers, we improved the performance of Q-LEARNING algorithms. We show that our traffic light controller successfully learns how to manage the intersection with less deadlocks than without learning.

KEYWORDS

cellular automaton,reinforcement learning,Q-LEARNING, green wave

ACM Reference Format:
Saif Islam Bouderba and Najem Moussa. 2019. Reinforcement Learning (Q-LEARNING) traffic light controller within intersection traffic system . In *The 4th International Conference On Big Data and Internet of Things (BDIoT'19), October 23–24, 2019, Rabat, Morocco.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3372938.3372999

1 INTRODUCTION

With the core status in the urban traffic system, intersection management has taken a big part of research attention. In the literature traffic lights are the most existing, which has aided to enhance the traffic condition at crossroads. the intersections may include traffic circle, roundabout, T-intersection, signalized intersection and unsignalized intersections, etc [4–7, 10]. However, city traffic signal control is the result of vehicle modernization: in order to distinct the traffic flows that could result in traffic conflict, it is obligatory to guide and schedule it efficiently by using traffic signals. The problem of city traffic is more and more serious, and several researchers are trying hard to solve it. On one hand, researchers are ceaselessly presenting new theories and new methods, and on the other hand, many areas coordinated traffic control systems based on computers are developed. Generally, traffic control strategies include static-time control method, area static control and area dynamic control method, green wave control method. Thus, approaches have been used trying to decrease waiting times of drivers and to avoid traffic jams. The most ordinary consists of discovering the right phases and periods of traffic lights to quantitatively optimize traffic flow. This results in "green waves" that flow through the main avenues of a city, ideally enabling vehicles to drive through them without confronting a red light, as the velocity of the green wave matches the preferred velocity for the road. However, this approach does not consider the current state of the traffic. If there is a high traffic density, vehicles entering a green wave will be stopped by vehicles ahead of them or vehicles that turned into the road, and once a vehicle misses the green wave, it will have to wait the whole period of the red light to go in the next green wave. On the other hand, for low densities, vehicles might reach too quickly at the next crossroad, having to stop at each crossing. This method is certainly surpassing than having no synchronization at all, however, it can be significantly enhanced.

Reinforcement learning has revealed its advantages in the ability to explore the environment to exploit the most appropriate actions in the dynamic situations. Reinforcement learning is involving by researches in the traffic signal management systems, it's shown a important results. Q-Learning as one of reinforcement learning algorithm is applied in the optimization of the traffic flow in single traffic intersection [3]. Besides that, Q-learning algorithm has also been extremely valued in the researches of traffic control system as multi-agents systems [2], [1]. In this study, we aimed to studied the signalized traffic intersections by applied Q-learning algorithm.

In this paper, we investigate a smart optimization method, based on Reinforcement learning , to enhance automatically the global traffic flow in the city. We compute the global traffic flow for different light control strategies, in particular synchronized traffic lights and green wave and we compare these results with Q-Learning applied in traffic lights.

The paper is organized as follows. In the next section, we give the definition of the model and drivers strategy-update dynamics. In section 3, we present the green wave strategy .In section 4 we present Q-learning algorithm and the performance metrics used to study the model. In Section 5, we investigate the characteristics of the system by using our simulation program and where some interesting observations are analysed. In Section 6, we will bring our conclusion.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BDIoT'19, October 23–24, 2019, Rabat, Morocco

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7240-4/19/10…$15.00
https://doi.org/10.1145/3372938.3372999

2 TRAFFIC MODEL

2.1 Model structure and moving rules

The model studied in this paper consists of an underlying lattice where the number of intersections is set to N × N = 4. Each roads are two-ways, with one lane in each direction. Each road consists of L cells of identical size and the time t is discrete. All streets cross each others at the intersection sites (see Fig 1). Open boundary conditions are applied to all roads, with four entry(α)/exit(β) points. We define road 1 as the road which goes from East to West, road 2 from West to East, road 3 from South to North and road 4 from North to South. Each position can be also empty or occupied by a vehicle with the speed $v = 0, v_{max}$, where v_{max} is the maximum speed. The dynamics of vehicles in each street is controlled by the NaSch moving rules [8] are as follows:

- Step 1
  Acceleration: $v = \min(v + 1, v_{max})$.
- Step 2
  Slow down:
  - Case 1
    The traffic light is red in front of the n-th vehicle:
    $v = \min(v_n, d_{n-1}, s_{n-1})$.
  - Case 2
    The traffic light is green in front of the n-th vehicle: If the next two cells directly behind the crossroad are occupied
    $v = \min(v_n, d_{n-1}, s_{n-1})$.
    else
    $v = \min(v_n, d_{n-1})$.
- Step 3
  Randomization: $v = \max(v - 1, 0)$ with probability p_r.
- Step 4
  Movement: the vehicle moves v sites forward.

The boundary condition is defined as follows: at each time step and for each road i, a vehicle k is injected with probability α and then the velocity $v_k = v_{max}$ is assigned to the vehicle. Here x_n denotes the position of the n-th car and$d_n = x_{n+1} - x_n$is the distance to the next car ahead (see Fig 1). The distance to the next traffic light ahead is given by s_n. The length of a single cell is set to 7.5 m in accordance to the NaSch model. In the initial state of the system, N_v vehicles are distributed among the roads. Here we only consider the case where the number of vehicles on east-bound roads is equal to the one on north-bound roads. The global density then is defined by $\rho = \frac{Nv}{N^2 \times 2(d-1)}$ since in the initial state the N^2 crossroads are left empty (see ref. [1] for more details).

3 GREEN WAVE STRATEGY

Next we will introduce a simple “green wave” strategy that can improve the overall network. A green wave occurs when a series of traffic lights are synchronized to allow maximum traffic flow over many intersections in one direction. Any car travelling along with the green wave will see a progressive cascade of green lights, and not have to stop at intersections. To implement the green wave strategy in a traffic city with traffic lights, Brockfeld et al attributed an individual offset parameter $\triangle T$ to every crossroad. This offset parameter is used to implement a certain time delay T between the traffic light phases of two successive crossroads. For a crossroad of indices (i,j) in the network, we assign the offset parameter $\triangle T i, j = 0, ..., 2T$. Notice, that a higher $\triangle T$ has no effect because 2T corresponds to one complete cycle of a traffic light. We pick out the crossroad at the bottom left corner of the network as the starting point with no time delay$\triangle T = 0$. Then the offset parameter of the crossroads is chosen with respect to the following formulas:

$$\triangle T_{i,j} = ((i + j)T_{delay}) mod(2T), (i, j = 0, 1, .., N) \qquad (1)$$

4 Q-LEARNING ALGORITHM

Reinforcement learning is usually presented as Q-learning algorithm which reward not just the actions taken but as well the states caused by the actions. Decisions made from the past is evaluated and stored as experiences data which will provide valuable help in the future.

The most metaphor used to explain the Q-learning algorithm is the relationship between a trainer and trainee, for example, a coach and student, or parents and their child. By taking the training of a student by a coach as example, the coach will evaluate each actions of the student after the commands are given. When a command is given to the student, the student will respond to it, and then the coach will observe and evaluate the performance of the student. If the student’s action is within the outlooks, a reward will be given to the student; and no reward will be given if the student did not act accordingly. In this way, the student will make a relationship between the orders and the actions; it will realize that only the correct order and action pairs will be rewarded. All these experiences motivate the student to respond according to the order of the coach.

4.1 Structure of Q-Learning

As indicated in the previous section, there are several ways to describe about Q-learning algorithm’s learning process. In Q-learning algorithm, the task of trainer is played by the environment model, while the Q-learning algorithm itself learns from the environment. Thus, in the development of a Q-learning algorithm, every operation includes in the process must be determined carefully. Q-Learning algorithm is constituted with several phases or operations. The development of Q-learning process is applied through the evaluation

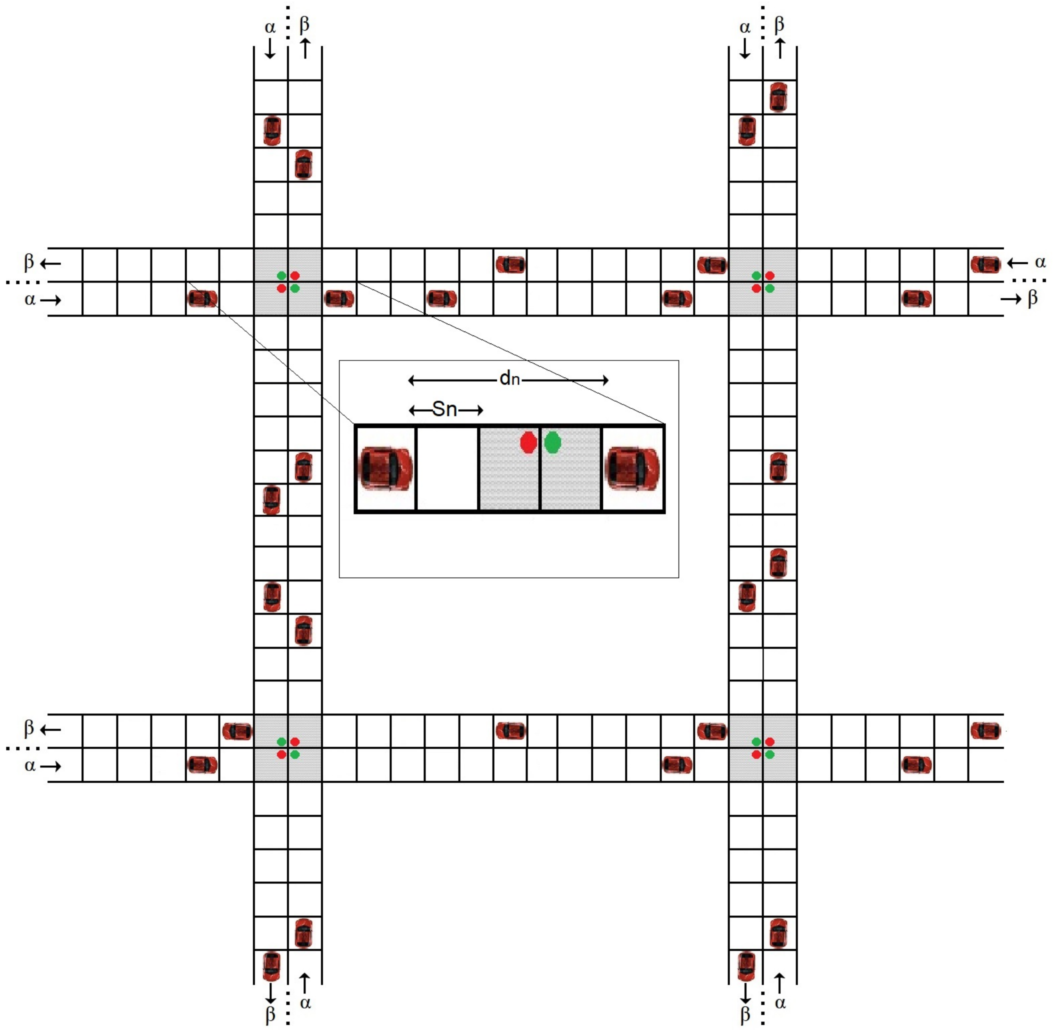


Figure 1: Snapshot of the underlying lattice of the model.

of (1). The flow chart of Q-learning is being illustrated in figure 2. The development of Q-learning begins with the initialization of the states and actions in the Q-table. After that, Q-learning will identify its present state in the environment. An action will be chosen from the action lists available by searching for the maximum possible rewards return by the action. Then, the actions chosen will be executed or evaluated. The rewards obtained from the actions selected will be updated in the Q-table. Q-learning will identify the next states in the environment model after the actions has been executed. Finally, Q-learning will verify the goal accomplishment, the process will start again if the goal did not accomplish.

$$Q(s,a)_i = (1-\alpha)Q(s,a)_{i-1} + \alpha[R(s,a)_i + \gamma_a'^{max} Q(s',a')] \quad (2)$$

where:

s = current state
a = action taken in current state
s'= next state
a'= action taken in next state
i = iteration
α = learning rate
γ = discounting factor

Q-Learning is evaluated from quotation (2), each of evaluated Q value is the rewards gained from the experiences in the exploration process. Q-table stored each single state and action pairs along with their rewards. α , is the factor that will influence the learning rate of the Q-learning algorithm. Q-learning rate Learning is vary between 0 and 1, and responsible for the weight of the recently learnt experience. Discounting factor γ is the variable that decides the importance of the future states. High discounting factor will

Table 1: Q-Learning algorithm parameters

parameter	Value initial
ϵ	0.1
γ	0.9
α	0.7

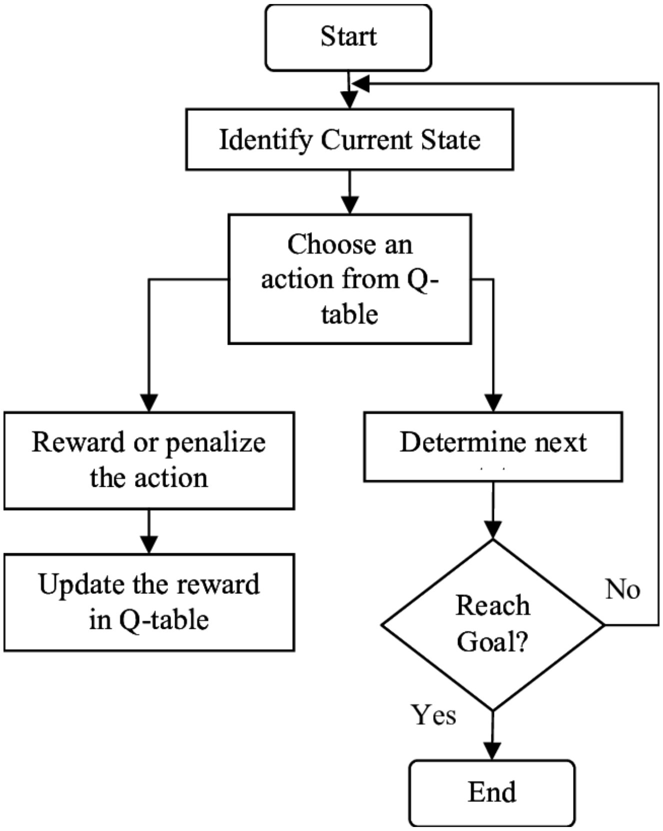

Start
Identify Current State
Choose an action from Q-table
Reward or penalize the action
Update the reward in Q-table
Determine next
Reach Goal?
No
Yes
End

Figure 2: Q-Learning algorithm flow chart

make Q-learning algorithm to be too speculative, where it will concentrate more on the possible future rewards and ignore the weight of the current experience.

4.2 State-Action Pairs

Q-Learning algorithm increases its experience through the exploration in the modelled environment. An accurately defined environment will ease the Q-learning's exploration process. The environment of the Q-learning is formed by the states and actions [9].

In this study of traffic intersection, the states of the designed Q-learning are the level of cars in queue at each intersection. There are four different density of cars in queue defined for this study, where they are classified from no cars in queue to high cars in queue. With each intersection having four traffic phases, the full possible states are 256 states combination from the permutation of four phases and four different density of cars in queue.

The actions list that is presented for the Q-learning algorithm is also a significant parameter, as they act as the navigators for the Q-learning algorithm in the environment. Each action made by the Q- learning will lead it to another state. If the state and action pairs are not set to be correct, then the entire Q-learning will not be able to get the optimum solution in the total process. Therefore, advancing or stopping at the intersection are defined as the actions of the proposed Q-learning algorithm.

4.3 Rewards and Penalties Functions

Although states and actions pairs of the Q-learning algorithm are set, rewards and penalties for each selected action have to be decided for ensuring the Q-learning algorithm is performing well. In Q-learning algorithm, the basic idea is the best actions will be valued with the highest rewards and the worst actions will be assigned with the least rewards. The goal of the proposed Q-learning algorithm is to enhance the traffic flow in the system. Thus, rewards functions are computed carefully for each appropriate traffic light and the actions that yield an accident will be penalized. The proposed Q-learning algorithm will end after it achieve the aim of the system. All of the rewards and penalties returned by the reward and penalties function are stored in the memory of Q-learning as their own experience for their future references.

4.4 Simulation Method and Performance Metrics

In this paper, simulation results for the proposed model, are carried out to investigate traffic characteristics and cooperation levels at traffic intersection. The network size of each road is $L = 100$. In this paper, we restricted our study to the case of equivalent roads, i.e. N_1 is the number of vehicles on the system . The network density is defined as $\rho = N/L$. The following parameters are fixed throughout this paper. The maximum speed of vehicles is set as $v_{max} = 1$ and the random braking probability is $p_r = 0.1$. The initial proportion of cooperators in each road is set to 50 per cent. The numerical results are obtained by averaging over 100 initial random configurations and 15000 time-steps after discarding 10000 initial transient steps.

- **Waiting time :** we define the waiting time of vehicles as the fraction of stopped vehicles on roads during the simulation time divided by the time interval.

$$\textbf{Waiting time} = \frac{1}{N \times T} \sum_{t=1}^{T} \sum_{i=1}^{N} (1 - v_i(t)) \tag{3}$$

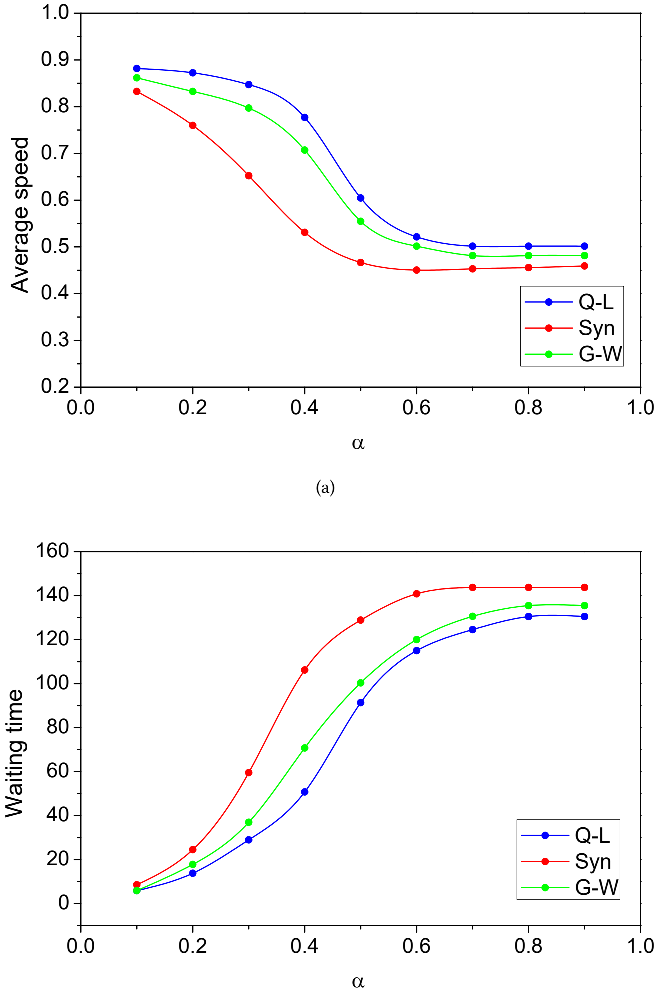
  where T is the time interval and N is the number of vehicles on roads. The variable v_i is equal to 1 if the driver i is not stopped and equal to 0 if the driver i is stopped.

- **Average velocity:** we define the average velocity of vehicles on roads as the sum velocities of all vehicles on roads during the simulation time divided by the time interval.

$$\textbf{Average velocity} = \frac{1}{N \times T} \sum_{t=1}^{T} \sum_{i=1}^{N} v_i(t) \tag{4}$$

5 RESULTS AND DISCUSSION

In this section, we present the simulation results for our indicators of system performance in order to investigate the relationship between the states of our transportation system and the model parameters. In this paper, we are interested to the effect of the Q-learning algorithm.



(a)

(b)

Figure 3: The average velocity (a)and the waiting time (b) as a function of injection rate α for different methods.

Firstly, we show in figure 3,the average velocity as a function of α for different strategies.As we can see after applied the Q-learning, the traffic light make the good action in the intersection, for that , the average velocity with the Q-learning (blue line) give as the best result over the green wave strategy (green line) and the synchronized strategy (red line). In addition, for $\alpha > 0.6$ the system is in the current phase, here, the average velocity is not more dependent on the injection rate α . Moreover ,in figure 3(b), we show the waiting time as a function of α for different strategies.When we increase the value ofα,the density of cars increase and the road becomes crowded.This, results an increase of the waiting time of stopped cars,so we notice in figure 3(b) that with the Q-learning algorithm (blue line) reduce the waiting time of cars.

6 CONCLUSIONS

In this paper, we reconsider different traffic light control strategies within the framework of a cellular automata model for city traffic. For this purpose, we started with the original formulation of the ChSch model where the traffic lights are switched synchronously. Moreover, there exist an important strategy, the green wave strategy. We have introduced a developed Q-learning traffic system, it's performing well throughout the simulations. This shows the abilities and capabilities the of Q-learning in the traffic systems. Indeed, our simulation results show that the green wave strategy implemented reveals an improvement through the synchronized traffic lights over different densities. However, the Q-learning of each traffic light able to work independently and having traffic information sharing with other traffic lights. Q-learning algorithm has exposed its strength in exploration in the traffic environment and also the flexibility towards the fast changes of the environment by efficaciously manages the traffic lights decision at intersection within the traffic networks.

REFERENCES

[1] Baher Abdulhai, Rob Pringle, and Grigoris J Karakoulas. 2003. Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering* 129, 3 (2003), 278–285.

[2] PG Balaji, X German, and Dipti Srinivasan. 2010. Urban traffic signal control using reinforcement learning agents. *IET Intelligent Transport Systems* 4, 3 (2010), 177–188.

[3] Yit Kwong Chin, Nurmin Bolong, Aroland Kiring, Soo Siang Yang, and Kenneth Tze Kin Teo. 2011. Q-learning based traffic optimization in management of signal timing plan. *International Journal of Simulation, Systems, Science & Technology* 12, 3 (2011), 29–35.

[4] M Ebrahim Fouladvand, Zeinab Sadjadi, and M Reza Shaebani. 2004. Characteristics of vehicular traffic flow at a roundabout. *Physical Review E* 70, 4 (2004), 046132.

[5] Ding-wei Huang. 2015. Modeling gridlock at roundabout. *Computer Physics Communications* 189 (2015), 72–76.

[6] Xin-Gang Li, Zi-You Gao, Bin Jia, and Xiao-Mei Zhao. 2009. Cellular automata model for unsignalized T-shaped intersection. *International Journal of Modern Physics C* 20, 04 (2009), 501–512.

[7] R Marzoug, H Ez-Zahraouy, and A Benyoussef. 2015. Simulation study of car accidents at the intersection of two roads in the mixed traffic flow. *International Journal of Modern Physics C* 26, 01 (2015), 1550007.

[8] Kai Nagel and Michael Schreckenberg. 1992. A cellular automaton model for freeway traffic. *Journal de physique I* 2, 12 (1992), 2221–2229.

[9] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3-4 (1992), 279–292.

[10] Dong-Fan Xie, Zi-You Gao, Xiao-Mei Zhao, and Ke-Ping Li. 2009. Characteristics of mixed traffic flow with non-motorized vehicles and motorized vehicles at an unsignalized intersection. *Physica A: Statistical Mechanics and its Applications* 388, 10 (2009), 2041–2050.