UDC 519.237.5, 519.83
MSC 91A25

Dynamic decision-making under uncertainty: Bayesian learning in environmental game theory*

J. Zhou[1], *O. L. Petrosian*[1,2], *H. Gao*[2]

[1] St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation
[2] Qingdao University, 308, Ningxia Road, Qingdao, 266071, China

For citation: Zhou J., Petrosian O. L., Gao H. Dynamic decision-making under uncertainty: Bayesian learning in environmental game theory. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2024, vol. 20, iss. 2, pp. 289–297. https://doi.org/10.21638/spbu10.2024.213

This paper investigates the issue of pollution control dynamic games defined over a finite time horizon, with a particular focus on parameter uncertainty within the ecosystem. We employ a dynamic Bayesian learning method to estimate uncertain parameters in the dynamic equation, differing from traditional single-instance Bayesian learning which does not involve continuous signal reception and belief updating. Our study validates the effectiveness of the dynamic Bayesian learning approach, demonstrating that, over time, the beliefs of the players progressively converge towards the true values of the unknown parameters. Through numerical simulations, we illustrate the convergence process of beliefs and compare optimal control strategies under different scenarios. The findings of this paper offer a new perspective for understanding and addressing the uncertainties in pollution control problems.

Keywords: dynamic Bayesian learning, pollution control games, ecological uncertainty, optimal control strategy.

1. Introduction. The corpus of literature on dynamic game models with unknown parameters, specifically within the realm of pollution control games, has expanded considerably. At the beginning of the nineties, several papers formulated a deterministic growth of pollution stocks [1–4]. The introduction of uncertainty into these models was pioneered by [5–8], which examined a dynamic game of international pollution control under ecological uncertainty and integrated a learning mechanism for the players. This approach allowed for the characterization and comparison of non-cooperative emission strategies in response to incomplete knowledge about the distribution of ecological uncertainty.

Paper [9] delved into learning within the context of a resource extraction and investment game, acknowledging that agents grapple with uncertainties in both the dynamics of the environment and their payoff functions. While most of the literature has focused on scenarios where parties have incomplete knowledge of others' private information, the work [10] considered the strategic implications of a player's partial knowledge of the truth. This body of work collectively enhances our understanding of strategic behavior in the face of uncertainty in dynamic environmental games.

Bayesian updating is a pivotal method for addressing uncertainties regarding unknown parameters within dynamic systems. This method has been previously utilized to investigate decision-making processes involving binary choices. Specifically, the Continuous

* This research was supported by Saint Petersburg State University (ID project: N 94062114).

Opinions and Discrete Actions (CODA) model, introduced by Martins, provides a framework where each choice is linked to a continuous probability, bridging the gap between continuous beliefs and discrete actions [11, 12]. In this model, agents maintain continuous opinions and base their decisions on the discrete actions observed from their peers.

Further expanding on this concept [13–15], explored the dynamics of agents' opinions concerning an unknown parameter. Agents would form and adjust their opinions continuously, while also considering the likelihood of different values, which could follow a distribution that combines Gaussian and uniform characteristics. This dual-distribution approach encapsulates the possibility of agents operating under partial ignorance during interactions.

Building upon the foundational work in Bayesian updating and its applications in dynamic systems, our research explores new dimensions in environmental economics and game theory. We focus on how players form and adjust their beliefs in the face of ecological uncertainties, and how these beliefs, in turn, guide their strategies in a dynamic setting. We extend our inquiry to the long-term effects of belief-driven strategies on the stability of the system and the achievement of optimal outcomes. The research questions whether a belief-centric approach leads to more efficient decision-making or whether it potentially limits adaptability and innovation in response to environmental challenges. Through a blend of theoretical exploration and empirical analysis, our paper aims to deepen the understanding of the role of beliefs in shaping strategies within dynamic, uncertain environments.

2. Description of the problem. We focus on tracking the cumulative net emissions in the environment, represented as a function of time $S(t)$. The dynamics of this stock variable are crucial for understanding the environmental impact and are formulated as follows [5]:

$$S(t+1) = \widetilde{\eta}\left(\sum_{i=1}^{N} u_i(t) + \delta S(t)\right), \quad S(t_0) = S_0, \tag{1}$$

where $0 < \delta < 1$ indicates that $1 - \delta$ is the natural decay rate of pollution. The variable $\widetilde{\eta}$ introduces ecological uncertainty into our model, accounting for various factors that significantly influence the environmental dynamics.

A pivotal aspect of our model is the incorporation of uncertainty, represented by $\widetilde{\eta}$ in the dynamic equation. This uncertainty factor is integral to the environmental model, capturing the unpredictable elements that typically affect ecological systems. The variable $\widetilde{\eta}$ is not just a mathematical construct but a reflection of real-world phenomena like climatic variations, unanticipated ecological responses, and other environmental stochasticities. It serves to model the often erratic and non-linear impact of emissions on the environment, thereby adding a layer of realism to the simulation. Understanding and quantifying this uncertainty is crucial for accurate modeling and effective decision-making in environmental management.

The realization of the random variable $\widetilde{\eta}$, denoted by x, follows a probability distribution $\varphi(x|\theta)$. Here, θ in the parameter space $\Theta \subset \mathbb{R}^l$ defines the vector of sufficient parameters for the probability density function (p.d.f.) φ.

Players form their beliefs about the unknown parameter $\xi(\theta)$, particularly crucial in contexts laden with uncertainty, like environmental issues. This uncertainty may represent factors such as the rate at which nature absorbs pollutants.

The optimization problem for player $i = 1, 2, ..., N$, assuming a behavior of welfare maximization over a finite horizon, is given by

$$\max_{u_i(t)} \quad \sum_{t=0}^{T} [u_i(t)(a_i - u_i(t)) - b_i S(t)], \tag{2}$$

subject to the pollution dynamics described in equation (1). In formula (2), a_i, $i = 1, 2, ..., N$, are constants, and b_i for each, $i = 1, 2, ..., N$, denotes the positive marginal cost of the pollution stock.

3. Dynamic Bayesian learning method. In this chapter, we delve into the strategies players employ for updating their beliefs in the face of uncertainty within the state equation. Such uncertainty might stem from variables like unknown parameters or erratic environmental factors. Grasping the nuances of belief revision in response to new information and observations is pivotal.

We initiate our discussion by outlining the conditional expectation method, which is used to predict the realization of a random variable at time t, denoted as x_t. We will then explore how players incorporate fresh signals into their decision-making, thereby enhancing the accuracy of their estimations.

In this Section, we analyze how the planner employs Bayesian methods to interpret and learn from the signal x_t observed at each time t. To streamline our model, we standardize the initial beliefs of all players. The following assumptions are made for simplicity and clarity: at the beginning of each stage, players collectively adopt a prior belief $\xi_t(\theta)$, which is recognized as common knowledge among them. Players make their decisions without the knowledge of the current period's signal; the signal is observed only post-decision. This information is then utilized to update their beliefs about θ. In each stage, players use mathematical expectation to estimate the unknown parameter θ. The estimated value at stage t, denoted as $\overline{\theta}_t$.

Formally, given the prior belief $\xi_t(\theta)$ and the signal x_t at time t, the posterior belief $\hat{\xi}_t(\theta|x_t)$ is defined by the following equation:

$$\hat{\xi}_t(\theta|x_t) = \frac{\varphi(x_t|\theta)\xi_t(\theta)}{\int_\Theta \varphi(x_t|y)\xi_t(y)\mathrm{d}y} \tag{3}$$

for $\theta \in \Theta$. This formulation, based on Bayes' inference, describes the learning process through the continual updating of beliefs in light of the insights derived from each observed signal x_t. Significantly, equation (3) functions independently of the control vector.

In the context of sequential decision-making, dynamic games of incomplete information are characterized by the uncertainty and evolving nature of the environment in which players operate. Players are required to make strategic decisions based on limited information, aiming to maximize their expected payoffs over time. The fundamental principles governing their behavior and decision-making process can be described as follows:

a) at each time $t = 0, 1, 2, \ldots, T$, players do not know $x_t, x_{t+1}, \ldots, x_T$, when choosing their strategies $u_i(t)$;

b) they aim to maximize their payoff from time t to T, which depends on future states;

c) players use their prior beliefs $\xi_t(\theta)$ about unknown parameters based on information available up to time t, which are based on the information accessible up to time t (i.e., $x_0, \ldots, x_{t-1}$), in order to estimate the unknown parameters. They then apply this estimation as the actual value of $\widetilde{\eta}$ starting from time t;

d) upon receiving the signal x_t, players update their belief to the posterior belief $\hat{\xi}_t(\theta|x_t)$, which is assigned as the prior belief at time $t + 1$.

Proposition. At any stage t, the posterior belief about an unknown parameter θ aligns with the prior distribution for the next stage $t+1$. This proposition underscores the

consistency and progression in belief formation, ensuring a seamless transfer of knowledge from one stage to the next.

At the outset of each stage, players establish a prior belief regarding the unknown parameters. This belief, coupled with the current state, guides them in selecting an optimal strategy. Following this decision, players are presented with signals that provide insights into these parameters. Utilizing the information from these signals, they update their initial belief to a posterior belief. This updated belief then serves as the new prior for the subsequent stage, setting the stage for the next cycle of decision-making. This process repeats itself, creating a dynamic, iterative game, where Bayesian learning plays a central role in strategy adaptation and decision-making.

4. Theoretical analysis. At the start of each stage, players assign prior beliefs to the unknown parameters θ. These beliefs are based on normal-gamma conjugate priors and are represented as follows:

$$\xi_t(\theta) \stackrel{\text{def}}{=} \mathcal{N}\left(\mu \mid \mu_t, (k_t\lambda)^{-1}\right) \mathcal{G}\left(\lambda \mid \alpha_t, \text{ rate } = \beta_t\right),$$

where $\mathcal{G}$ denotes the gamma distribution, $\mathcal{N}$ is the normal distribution.

Using equation (3) of Bayes' inference, the posterior distribution of unknown parameters θ can be updated as

$$\hat{\xi}_t(\theta|x_t) \propto \varphi(x_t|\mu,\lambda)\xi_t(\theta) \propto$$

$$\propto \mathcal{N}\left(\mu \mid \frac{k_t\mu_t + x_t}{k_t+1}, ((k_t+1)\lambda)^{-1}\right) \times \mathcal{G}\left(\lambda \mid \alpha_t + \frac{1}{2}, \beta_t + \frac{k_t\left(x_t-\mu_t\right)^2}{2\left(k_t+1\right)}\right). \tag{4}$$

Applying Proposition and exploit the property that the distribution is conjugate, we get the following formula:

$$\begin{aligned} \mu_{t+1} &= \frac{k_t\mu_t + x_t}{k_t+1}, \\ k_{t+1} &= k_t + 1, \\ \alpha_{t+1} &= \alpha_t + \frac{1}{2}, \\ \beta_{t+1} &= \beta_t + \frac{k_t\left(x_t-\mu_t\right)^2}{2\left(k_t+1\right)}, \end{aligned} \tag{5}$$

here $t \in \{0,1,2,...,T-1\}$, given the initial beliefs μ_0, k_0, α_0, and β_0. The left-hand side of equation (5) represents the players' initial beliefs for stage $t+1$, whereas the right-hand side of equation (5) is linked to equation (4) based on Proposition.

As we delve deeper into the implications of our Bayesian learning model, it becomes essential to establish the theoretical underpinnings that guarantee its effectiveness over time. A key aspect of this is the convergence of the Bayesian estimator, which we address in the following theorem. This theorem demonstrates that the iterative Bayesian learning process leads to a convergence of the estimated mean towards the true mean as the number of stages increases indefinitely.

Theorem. *As the number of stages progresses to infinity, the expected value of the Bayesian estimator for the unknown mean converges to the true value of the parameter. Formally,*

$$\lim_{n\to\infty} E(M_n) = \mu,$$

where M_n is a random variable representing the Bayesian estimate of the unknown mean μ at stage n.

P r o o f. In our model, the sequence $X_0, X_1, \ldots$ represents an infinite series of independent and identically distributed (i.i.d.) Lebesgue integrable random variables. Each variable in this sequence has an expected value denoted as $E(X_i) = \mu$, which is constant across the series.

Considering the Bayesian estimator M_n for the unknown mean at stage n, we have

$$E\left(M_n\right) = \frac{k_0\mu_0}{k_0+n} + \frac{1}{k_0+n}\sum_{i=0}^{n-1} E(X_i) = \frac{k_0\mu_0}{k_0+n} + \frac{n\mu}{k_0+n} = \frac{\frac{k_0\mu_0}{n} + \mu}{\frac{k_0}{n}+1}. \tag{6}$$

Taking the limit of equation (6) as n approaches infinity, we find

$$\lim_{n\to\infty} E\left(M_n\right) = \frac{0+\mu}{0+1} = \mu.$$

This result demonstrates that the expected value of the Bayesian estimator M_n converges to the true mean μ as n becomes infinitely large. Therefore, the theorem is proved. □

The aforementioned theorem has profound implications for our dynamic game model. It validates the approach of using Bayesian learning mechanisms with evolving uncertainties. The convergence of the Bayesian estimator to the true parameter value as the number of stages increases provides a solid foundation for decision-making in uncertain environments. This theoretical assurance of convergence is not just mathematically significant; it also instills confidence in the practical application of our model, especially in complex, real-world scenarios where long-term strategy and adaptability are crucial.

Having estimated the unknown parameters in our dynamic game of pollution control, we now turn our attention to identifying the Nash equilibrium of the model. This equilibrium represents a state, where no player can benefit by changing their strategy while others keep theirs unchanged. To find this equilibrium, we employ the Hamilton — Jacobi — Bellman method, a powerful tool in dynamic optimization.

The Hamilton — Jacobi — Bellman equation for each player in our pollution control model is central to identifying the Nash equilibrium strategies. We denote the value function of player i as $V_i(t, S, \mu_t, k_t, \alpha_t, \beta_t)$, which represents the player's maximum expected utility at a given state. This equation for this value function is expressed as

$$V_i(t, S, \mu_t, k_t, \alpha_t, \beta_t) = \max_{u_i(t)}\{(u_i(t)(a_i - u_i(t)) - b_i S) +$$

$$+ V_i(t+1, \hat{\eta}_t(\sum_{i=1}^{N} u_i(t) + \delta S), \mu_t, k_t, \alpha_t, \beta_t)\},$$

where $\hat{\eta}_t = E(\widetilde{\eta}|\mu_t, \alpha_t/\beta_t)$.

Assuming a linear-state structure for our model, the value function is hypothesized to be linear and is formulated as

$$V_i(t, S, \mu_t, k_t, \alpha_t, \beta_t) = A_i(t, \mu_t, k_t, \alpha_t, \beta_t)S + B_i(t, \mu_t, k_t, \alpha_t, \beta_t),$$
$$i = 1, 2, ..., N.$$

To derive the coefficients of the value function, we equate the coefficients in terms of S:

$$A_i(t, \mu_t, k_t, \alpha_t, \beta_t) = -b_i + A_i(t, \mu_t, k_t, \alpha_t, \beta_t)\hat{\eta}_t\delta,$$

$$A_i(T+1, \mu_t, k_t, \alpha_t, \beta_t) = 0$$

for $t \in \{0, 1, ..., T\}$, $i = 1, 2, ..., N$.

We then define the Nash equilibrium strategies with dynamic Bayesian learning for each player $i = 1, 2, ..., N$ in the subgame:

$$\widetilde{u}_i^*(t, \mu_t, k_t, \alpha_t, \beta_t) = \frac{a_i}{2} - \frac{\hat{\eta}_t}{2(1 - \hat{\eta}_t\delta)} b_i(1 - (\hat{\eta}_t\delta)^{T-t}),$$

where $a = \sum_{i=1}^{N} a_i$; $b = \sum_{i=1}^{N} b_i$; $t \in \{0, 1, ..., T\}$.

5. Numerical simulation and results analysis. We present the results of our numerical simulations designed to demonstrate the effectiveness of our dynamic Bayesian learning approach in the context of ecological uncertainty. The signals we generate follow a normal distribution with a mean of 0.5 and a variance of 1, thus the true value of the unknown parameter regarding the mean of the random variable is 0.5. The parameters of the model are set as $a = 8$, $b = 3$, $N = 5$, and $\delta = 0.6$. The time interval considered is $T = 70$.

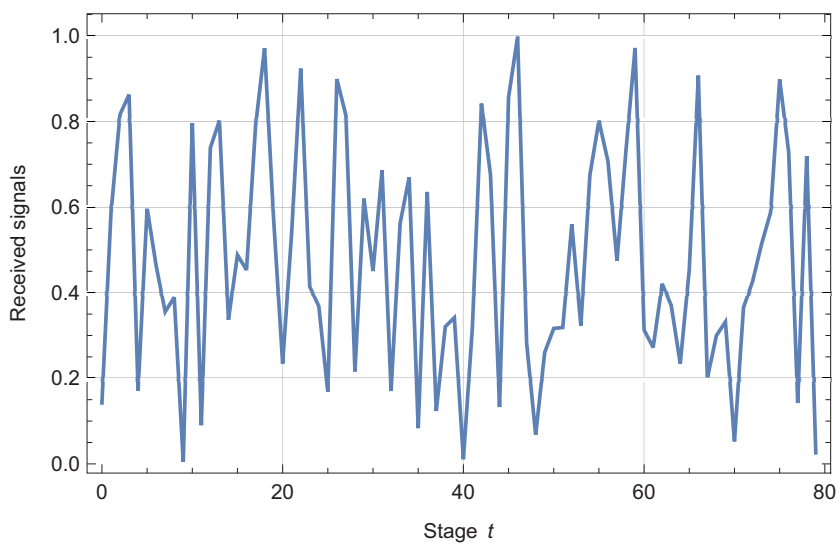

Figure 1. Received signals in estimating ecological uncertainty

Figure 1 illustrates the dynamic nature of signal reception in the context of ecological uncertainty estimation. It highlights how varying data inputs, represented through different signal types, contribute to the evolving understanding of ecological systems.

Upon analyzing Figure 2, we observe a significant trend: as time progresses, the players' estimates of the unknown parameter increasingly converge to the true value, which in this case is 0.5. This convergence is a clear indication of the effectiveness of dynamic Bayesian learning in reducing uncertainty and enhancing the accuracy of parameter estimation. The simulation thus validates our model's capability to adaptively learn and

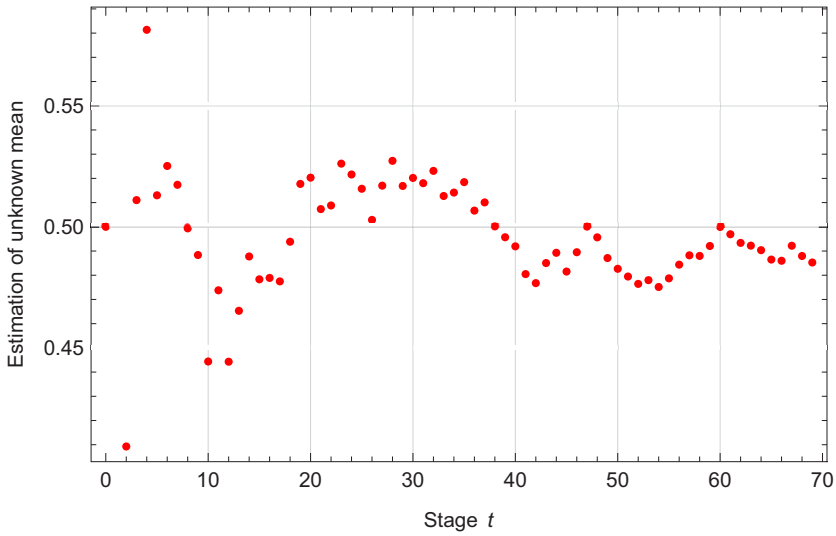

Figure 2. Dynamic Bayesian estimation of unknown parameters over time

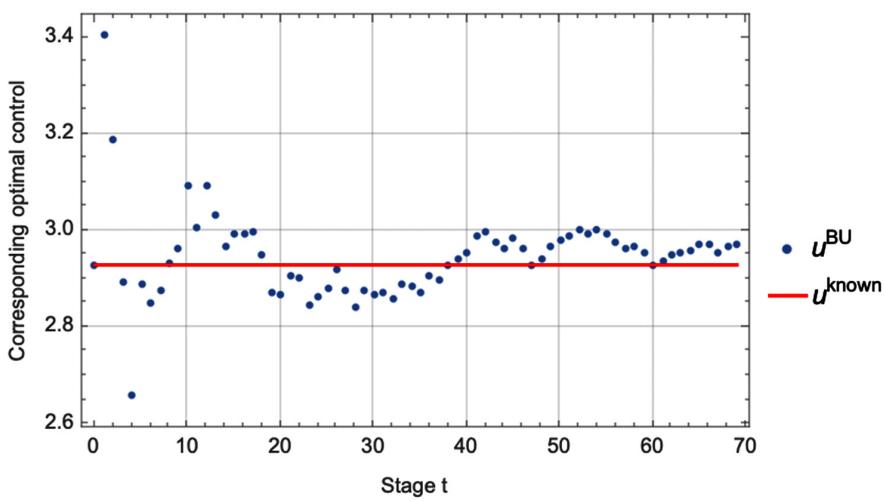

Figure 3. Comparison of optimal control strategies with and without ecological uncertainty

refine estimates, ultimately leading to more informed and precise decision-making in the face of ecological uncertainty.

Figure 3 presents a comparison between the optimal control strategies adopted by players under different conditions of ecological uncertainty. When ecological uncertainty is absent, the control strategy adopted by players is markedly more stable.This line represents the benchmark control action, assuming the players have complete knowledge of the ecological parameters. On the other hand, the blue dots depict the control strategy under ecological uncertainty, where players must continuously adapt their actions based on the signals they receive. The scattered nature of these points highlights the variability and dynamism in the players' control strategies as they attempt to estimate and account for ecological uncertainty.

6. Conclusion. This study has provided a comprehensive examination of dynamic Bayesian learning in the context of ecological uncertainty within pollution control games. Through a series of detailed numerical simulations, we have demonstrated the significant impact of uncertainty on players' decision-making processes. Our results show that as players receive signals over time, their estimates of unknown ecological parameters converge to the true values, allowing for more informed and optimal control strategies.

The simulation analysis reveals that the presence of ecological uncertainty necessitates a dynamic approach to control. Players must adapt their strategies in real-time, based on the evolving signals they receive, which capture the uncertain state of the ecological system. This adaptability is crucial for managing environmental challenges effectively, especially when long-term outcomes are uncertain.

References

1. Van der Ploeg F., De Zeeuw A. A differential game of international pollution control. *Systems & Control Letters*, 1991, vol. 17, iss. 6, pp. 409–414.

2. Long N. V. Pollution control: A differential game approach. *Annals of Operations Research*, 1992, vol. 37, iss. 1, pp. 283–296.

3. Hoel M. Emission taxes in a dynamic international game of CO_2 emissions. *Conflicts and Cooperation in Managing Environmental Resources*. Berlin, Heidelberg, Springer, 1992, pp. 39–70.

4. Veijo K., Matti P., Tahvonen O. Transboundary air pollution and soil acidification: A dynamic analysis of an acid rain game between Finland and the USSR. *Environmental and Resource Economics*, 1992, vol. 2, pp. 161–181.

5. Masoudi N., Santugini M., Zaccour G. A dynamic game of emissions pollution with uncertainty and learning. *Environmental and Resource Economics*, 2016, vol. 64, iss. 3, pp. 349–372.

6. Ulph A., Maddison D. Uncertainty, learning and international environmental policy coordination. *Environmental Resource Economics*, 1997, vol. 9, iss. 4, pp. 451–466.

7. Arrow K. J., Fisher A. C. Environmental preservation, uncertainty, and irreversibility. *Quarterly Journal of Economics*, 1974, vol. 88, iss. 2, pp. 312–319.

8. De Zeeuw A., Zemel A. Regime shifts and uncertainty in pollution control. *Journal of Economic Dynamics and Control*, 2012, vol. 36, iss. 7, pp. 939–950.

9. Mirman L. J., Santugini M. Learning and technological progress in dynamic games. *Dynamic Games and Applications*, 2014, vol. 4, pp. 58–72.

10. Liu Z. Games with incomplete information when players are partially aware of others' signals. *Journal of Mathematical Economics*, 2016, vol. 65, pp. 58–70.

11. Martins A. C. Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 2008, vol. 19, iss. 4, pp. 617–624.

12. Martins A. C. Mobility and social network effects on extremist opinions. *Physical Review E*, 2008, vol. 78, iss. 3, art. no. 036104.

13. Martins A. C. Bayesian updating rules in continuous opinion dynamics models. *Journal of Statistical Mechanics: Theory and Experiment*, 2009, no. 02, art. no. P02017.

14. Lorenz J. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 2007, vol. 18, iss. 12, pp. 1819–1838.

15. Sirbu A., Loreto V., Servedio V. D. P., Tria F. Opinion dynamics: Models, extensions and external effects. *Participatory Sensing, Opinions and Collective Awareness*, 2017, pp. 363–401.

Received: January 21, 2024.
Accepted: March 12, 2024.

Authors' information:

Jiangjing Zhou — Postgraduate Student; 17854230890@163.com

Ovanes L. Petrosian — Dr. Sci. in Physics and Mathematics, Professor; petrosian.ovanes@yandex.ru

Hongwei Gao — Dr. Sci. in Physics and Mathematics, Professor; gaohongwei@qdu.edu.cn

Динамическое принятие решений в условиях неопределенности: байесовское обучение в теории экологических игр*

Ц. Чжоу[1], *О. Л. Петросян*[1,2], *Х. Гао*[2]

[1] Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9
[2] Университет Циндао,
Китайская Народная республика, 266071, Циндао, Дорога в Нинся, 308

Для цитирования: *Zhou J., Petrosian O. L., Gao H.* Dynamic decision-making under uncertainty: Bayesian learning in environmental game theory // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2024. Т. 20. Вып. 2. С. 289–297. https://doi.org/10.21638/spbu10.2024.213

Исследуется проблема динамических игр по борьбе с загрязнением, определенных на конечном временном горизонте, с особым акцентом на неопределенность параметров внутри экосистемы. Используется динамический байесовский метод обучения для оценки неопределенных параметров в динамическом уравнении, отличающийся от традиционного байесовского обучения на единичных примерах, которое не предполагает непрерывного приема сигнала и обновления представлений. Проведенное исследование подтверждает эффективность динамического байесовского подхода к обучению, демонстрируя, что со временем убеждения игроков постепенно приближаются к истинным значениям неизвестных параметров. С помощью численного моделирования иллюстрируется процесс конвергенции убеждений, сравниваются стратегии оптимального управления в различных сценариях и анализируется влияние сигналов на управленческие решения участников. Результаты статьи открывают новую перспективу для понимания и устранения неопределенностей в задачах борьбы с загрязнением.

Ключевые слова: динамическое байесовское обучение, игры по борьбе с загрязнением окружающей среды, экологическая неопределенность, стратегия оптимального управления.

Контактная информация:

Чжоу Цзянцзин — аспирант; 17854230890@163.com

Петросян Ованес Леонович — д-р физ.-мат. наук, проф.; o.petrosian@spbu.ru

Гао Хунвэй — д-р физ.-мат. наук, проф.; gaohongwei@qdu.edu.cn

* Работа выполнена при финансовой поддержке Санкт-Петербургского государственного университета (проект ID: 94062114).