

УДК 004.42

DOI <https://doi.org/10.32782/IT/2025-1-18>

Олексій КУЧЕРЕНКО

аспірант кафедри інженерії програмного забезпечення, Державний університет «Житомирська політехніка», вул. Чуднівська, 103, м. Житомир, Україна, 10005

ORCID: 0009-0000-1488-5645

Тетяна ВАКАЛЮК

доктор педагогічних наук, професор, завідувач кафедри інженерії програмного забезпечення, Державний університет «Житомирська політехніка», вул. Чуднівська, 103, м. Житомир, Україна, 10005

ORCID: 0000-0001-6825-4697

Scopus Author ID: 57211133927

Бібліографічний опис статті: Кучеренко, О., Вакалюк, Т. (2025). Порівняння методів Reinforcement Learning та Model Predictive Control для побудови маршрутів руху БПЛА. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 135–142, doi: <https://doi.org/10.32782/IT/2025-1-18>

ПОРІВНЯННЯ МЕТОДІВ REINFORCEMENT LEARNING ТА MODEL PREDICTIVE CONTROL ДЛЯ ПОБУДОВИ МАРШРУТІВ РУХУ БПЛА

Стаття присвячена огляду та порівнянню двох методів планування маршрутів руху БПЛА: навчання з підкріпленням (Reinforcement Learning) та методу прогнозного керування (Model Predictive Control), що є актуальним питанням у сучасній індустрії навігації БПЛА.

Метою роботи є визначення особливостей використання, обмежень та інших властивостей вищезгаданих методів. У статті наведено принципи роботи вищезгаданих методів.

Методологія. Здійснено порівняння характеристик методів навчання з підкріпленням та прогнозного керування для побудови маршрутів руху БПЛА. Оцінювання проводилось за критеріями: тип середовища, необхідністю точної моделі середовища, адаптивністю у динамічних середовищах, обчислювальною складністю, та вимогами до даних.

Наукова новизна полягає у проведенні комплексного аналізу ефективності методів навчання з підкріпленням та методу прогнозного керування, з метою отримання найбільш затребуваних порівняльних характеристик, необхідних для вибору методу.

Висновки. Для методу навчання з підкріпленням описано роль агента, зворотнього зв'язку та вплив розрахованої за допомогою Q-функції винагороди на процес навчання моделі. Для методу прогнозного керування вказано особливості використання лінійної та нелінійної моделей прогнозування майбутньої поведінки у стратегії відступаючого горизонту та пов'язану з цим високу обчислювальну складність. Наведено ключові переваги та обмеження використання вищезгаданих методів. Наголошено, що вибір відповідного методу планування маршруту залежить не лише від специфіки поставленого завдання, а також від вимог до обчислювальних ресурсів, можливостей реагування розглянутої системи в режимі реального часу. Підкреслюється важливість інтеграції різних підходів для досягнення оптимальних результатів у складних умовах. Подальші дослідження в цій області можуть бути зосереджені на вдосконаленні методів та алгоритмів, які дають змогу безпілотним літальним апаратам в реальному часі адаптуватися до мінливих умов навколишнього середовища, на поєднанні методів навчання з підкріпленням та прогнозного керування з іншими алгоритмами планування маршрутів, наприклад A* або Ant Colony Optimization, з метою використання переваг кожного з них.

Ключові слова: дрон, БПЛА, навчання з підкріпленням, прогнозне керування, планування маршрутів.

Oleksii KUCHERENKO

Postgraduate Student at the Department of Software Engineering, Zhytomyr Polytechnic State University, 103, Chudnivska Str., Zhytomyr, Ukraine, 10005

ORCID: 0009-0000-1488-5645

Tetiana VAKALIUK

Doctor of Pedagogical Sciences, Professor, Head of the Department of Software Engineering, Zhytomyr Polytechnic State University, 103, Chudnivska Str., Zhytomyr, Ukraine, 10005

ORCID: 0000-0001-6825-4697

Scopus Author ID: 57211133927

To cite this article: Kucherenko, O., Vakaliuk, T. (2025). Porivniannia metodiv Reinforcement Learning ta Model Predictive Control dlia pobudovy marshrutiv rukhu BPLA [Comparison of Reinforcement Learning and Model Predictive Control methods for UAV route planning]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 136–142, doi: <https://doi.org/10.32782/IT/2025-1-18>

COMPARISON OF REINFORCEMENT LEARNING AND MODEL PREDICTIVE CONTROL METHODS FOR UAV ROUTE PLANNING

The article is devoted to an overview and comparison of two methods of UAV route planning: Reinforcement Learning and Model Predictive Control, which is a hot topic in the modern UAV navigation industry.

The purpose of this paper is to determine the specifics of the use, limitations and other properties of the above methods. The article presents the principles of the above methods.

Methodology. The paper compares the characteristics of Reinforcement Learning and Model Predictive Control methods for UAV route planning. The evaluation was carried out according to the following criteria: type of environment, the need for an accurate model of the environment, adaptability in dynamic environments, computational complexity, and data requirements.

The scientific novelty is a comprehensive analysis of the efficiency of Reinforcement Learning and Model Predictive Control methods, in order to obtain the most popular comparative characteristics necessary for choosing an method.

Conclusions. For the Reinforcement Learning method, the role of the agent, feedback, and the impact of the reward calculated using the Q-function on the model learning process are described. For the Model Predictive Control method, the features of using linear and nonlinear models for predicting future behaviour in the receding horizon strategy and the associated high computational complexity are outlined. The key advantages and limitations of the aforementioned methods are presented. The choice of a suitable route planning method depends not only on the specifics of the task, but also on the requirements for computing resources and the real-time response capabilities of the system under consideration. The importance of integrating different approaches to achieve optimal results in difficult conditions is emphasised. According to the results of the study, the characteristics of Reinforcement Learning and Model Predictive Control methods for building UAV routes are presented by the type of environment, the need for an accurate environment model, adaptability in dynamic environments, computational complexity, and data requirements. Further research in this area could focus on improving methods and algorithms that allow UAVs to adapt to changing environmental conditions in real time, combining Reinforcement Learning and Model Predictive Control methods with other route planning algorithms, such as A* or Ant Colony Optimisation, to take advantage of each.

Key words: drone, UAV, Reinforcement Learning, Model Predictive Control, route planning.

Постановка проблеми. Побудова маршруту руху безпілотних літальних апаратів (БПЛА) в умовах сьогодення України набуває неабиякого значення. Адже в умовах війни зросла потреба у використанні БПЛА для вирішення широкого кола завдань 2 від розвідки до доставки вантажів у важкодоступні місця. Ефективне планування маршрутів дозволяє оптимізувати використання дронів, збільшити тривалість і дальність польотів, скоротити витрати. Окрім того, зросли вимоги до автономності польотів дронів. Втрата зв'язку з оператором, складний рельєф, погані погодні умови вимагають від БПЛА здатності самостійно будувати й коригувати маршрути з врахуванням великої кількості факторів.

При цьому завдання планування маршрутів руху дронів передбачає врахування різних пов'язаних факторів. До них відносяться

геометричні межі перешкод, тобто форма і розміри, швидкість польоту, прискорення, кути нахилу, обмеження енергії та можливості адаптації до змін у реальному часі, такі як: раптова поява перешкод, зміна погодних умов, тощо. Врахування цих факторів ускладнює задачу побудови маршрутів руху БПЛА, особливо у тривимірному просторі. Крім того, у реальних сценаріях застосування БПЛА потрібна реакція БПЛА у межах короткого часу, що висуває додаткові вимоги до обчислювальної ефективності методу. Отже, постає потреба адаптації методів та алгоритмів, щоб вони були здатні забезпечити побудову безпечного та водночас оптимального маршруту в умовах реального часу.

Аналіз останніх досліджень і публікацій. У роботі (Буй, 2024) науковці оцінили ефективність методу планування руху БПЛА MPC за

допомогою моделювання. Метод MPC порівнювали з планувальником поля штучного потенціалу (APF), орієнтуючись на такі показники, як час руху, тривалість руху та споживання енергії. Зазначається, що метод MPC перевершив планувальника APF за показниками побудованої траєкторії та швидкістю (Буй, 2024). У роботі (Лю, 2024) дослідники розглядають проблему планування руху для декількох БПЛА, яка викликана труднощами в збалансуванні ефективності обчислень з необхідністю дотримуватися фізичних обмежень в програмах реального часу. Для вирішення проблеми автори пропонують новий алгоритм послідовного опуклого програмування на основі розподіленої моделі прогнозного керування (SCP-DMPC). Алгоритм SCP-DMPC поєднує можливості прогнозного управління розподіленою моделлю прогнозного управління (DMPC) з сильними сторонами оптимізаційного послідовного опуклого програмування (SCP) для підвищення ефективності та надійності в плануванні маршрутів руху групи БПЛА (Лю, 2024). У роботі (Гевара, 2024) автори запропонували систему нелінійної моделі Predictive Contouring Control (NMPC) для відстеження запланованої траєкторії руху, мінімізуючи ортогональні та тангенціальні помилки. Такий метод гарантує БПЛА точне слідування за маршрутом з максимізацією своєї швидкості. Автори провели моделювання у середовищах з статичними та динамічними перешкодами, підтвердивши стабільність та безпеку під час виконання завдань (Гевара, 2024).

Останніми роками набули популярності методи, які базуються на навчанні з підкріпленням. Вони дають змогу агенту (БПЛА) навчатись ефективним діям, безпосередньо взаємодіючи з моделлю середовища або реальним оточенням (Полідорос, 2017; Сюе, 2022). У роботі (Сюе, 2022) автори зосереджують свою увагу на проблемі використання методу глибокого навчання з підкріпленням (DRL), щоб дозволити БПЛА орієнтуватися у великих невідомих середовищах, наповнених перешкодами.

Дослідженнями методів планування маршрутів руху MPC та RL займались також інші науковці: Ву Дж., Кім Н. (Ву, 2020), Райхенсдорфер Е., Оденталь Д., Воллгерр Д. (Райхенсдорфер, 2018), Канані К. К.-Дж., Тіварі Б. К., Кім Т.-в. (Канані, 2024), Аліярі М. (Аліярі, 2022), Зурарі А. (Зурарі, 2023) та ін.

Мета дослідження. Метою даного дослідження є порівняльний аналіз двох методів планування маршрутів руху БПЛА: навчання з підкріпленням (RL) та прогнозного керування (MPC); визначення особливостей та обмежень

кожного з цих методів; виявлення середовищ, для яких кожен підхід є оптимальним.

Виклад основного матеріалу. Розглянемо принцип роботи метода RL. Метод навчання з підкріпленням відноситься до обчислювальних підходів до цілеспрямованого навчання та прийняття рішень (Саттон, 2018). У процесі навчання головну роль відіграє взаємодія з навколишнім середовищем за допомогою агента. Агент взаємодіє з навколишнім середовищем протягом дискретних часових кроків (рис. 1).

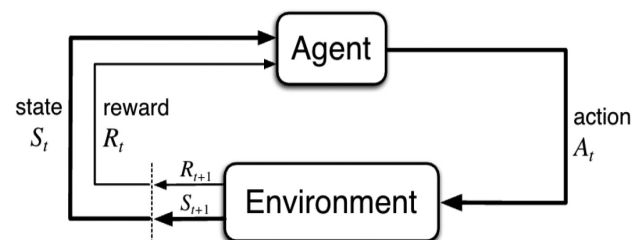


Рис. 1. Схема навчання RL за Саттон Р. С. (Саттон, 2018)

Кожного разу агент приймає рішення на основі поточного стану та розрахованої за допомогою Q-функції винагороди. У складних системах з великою кількістю параметрів у якості Q-функції використовується нейронна мережа (глибоке навчання з підкріпленням). Отже, задачу побудови маршруту руху БПЛА можна визначити як задачу Марковського процесу прийняття рішень. Ключовими компонентами є: S – множина станів середовища, наприклад швидкість, координати, тощо; A – множина можливих дій агента, наприклад зміна швидкості, тяги двигунів, кута нахилу, тощо; P – ймовірність переходу зі стану s до стану s_1 після виконання дії a ; R – функція винагороди за виконану дію a у стані s (значення може бути як позитивним так і негативним); γ – коефіцієнт впливу майбутніх винагород порівняно з поточними. Головним завданням процесу навчання є максимізація сумарно зваженої винагороди

$$G_t = \sum_{k=0}^{\infty} \gamma^k R(S_{t+k}, a_{t+k}),$$

де t – поточний часовий крок (Саттон, 2018).

Особливістю використання методу RL є те, що воно фактично розділено на два етапи. Перший етап полягає у навчанні, або створенні моделі. Це тривалий процес, який залежить від складності середовища та умов застосування дрона. Результатом є навчена модель. Другий етап – безпосередньо використання моделі, тобто визначення дій дрона на основі

попередньо навченої моделі. Цей етап має низьку обчислювальну складність.

Розглянемо принцип роботи метода прогнозного керування. Для прогнозування майбутньої поведінки процесу метод MPC використовує явну модель, яка може бути лінійною або нелінійною. Модель застосовується для аналізу поведінки системи та прогнозування майбутніх результатів на певний часовий горизонт. На кожному кроці MPC обчислює контрольну послідовність для оптимізації індексу продуктивності. Цей процес базується на мінімізації передбачуваної помилки або об'єктивної функції. Метод MPC використовує стратегію відступаючого горизонту, де горизонт прогнозування рухається вперед у часі на кожному кроці контролю (рис. 2). Реалізовується лише перша контрольна дія з оптимізованої послідовності і далі процес повторюється у наступному часовому кроці з оновленою інформацією. У результаті метод MPC отримує переваги у вигляді здатності обробляти обмеження на входи та виходи (Холкар, 2010).

Загалом, MPC потрібно розглядати як концепцію, на основі якої різними науковцями було створено багато конкретних методів. Найбільш поширеними є: динамічно-матричне керування (DMC), модельно-алгоритмічне керування (MAC), прогнозне функціональне керування (PFC), самоадаптивне керування з розширенням прогнозування (EPSAC), адаптивне керування з розширенням горизонтом (EHAC), узагальнене прогнозне керування (GPC) (Холкар, 2010).

Як зазначають науковці у своїй роботі (Хатер, 2023), метод MPC ефективно керує обмеженнями вводу-виводу та добре розв'язує

складні задачі оптимізації. Для прогнозування майбутніх станів метод MPC використовує математичну модель, яка дозволяє ефективно обробляти обмеження та оптимізувати продуктивність в межах певного горизонту (Шехаде, 2024). Це дає можливість застосовувати метод MPC в режимі реального часу в автономних системах (Хатер, 2023).

Розглянемо переваги методу MPC. Метод MPC з використанням лінійної математичної моделі може забезпечувати безперебійну роботу автономних транспортних засобів, у тому числі БПЛА, за рахунок обробки фізичних обмежень транспортного засобу з врахуванням граничних значень контрольних входів. Іншою перевагою методу MPC є можливість використання метода MPC у якості як контролера з одним входом та одним виходом (SISO), так і контролера з кількома входами та кількома виходами (MIMO). Ця властивість є необхідною для систем, що вимагають одночасного управління декількома змінними. Ще однією перевагою методу MPC є його прогнозна природа, яка дає можливість враховувати майбутні прогнозовані стани при прийнятті поточних рішень, як результат система має високу точність та стабільність у динамічних середовищах (Канані, 2024; Аліярі, 2022).

Недоліки лінійного методу MPC пов'язані з його використанням після лінеаризації динамічної моделі навколо маршрутних точок, а саме через похибку відображення поведінки системи у певний час, а також повільною перевіркою стабільності динамічної лінеаризованої системи. Іншим недоліком методу MPC є залежність ефективності від точної моделі системи.

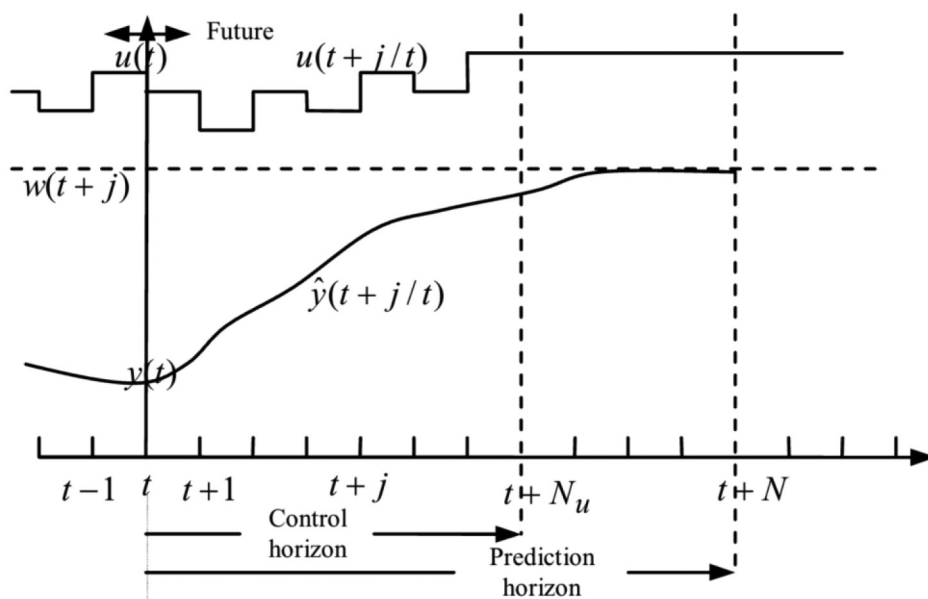


Рис. 2. Стратегія відступаючого горизонту за Холкар К. С. (Холкар, 2010)

Також до недоліків методу MPC відноситься висока обчислювальна складність через необхідність прогнозування майбутньої поведінки системи (Канані, 2024; Аліярі, 2022).

Розглянемо переваги та недоліки методу MPC з використанням нелінійної математичної моделі. До основних переваг відносять високу продуктивність у складних системах у порівнянні з лінійними системами управління, можливість роботи у режимі реального часу, вирішення нелінійних динамічних задач. До недоліків відносяться складність розробки математичної моделі у випадку нелінійної динаміки, високі вимоги до апаратного забезпечення через значні обчислювальні витрати, вища складність застосування у системах реального часу порівняно з лінійними системами управління (Хатер, 2023; Райхенсдорфер, 2018; Канані, 2024).

Розглянемо переваги методу RL. Метод RL дає оптимальні рішення швидко та надійно, може опрацьовувати великі масиви вхідних параметрів, відносно проста адаптація поведінки агента під конкретні цілі за допомогою визначення функції винагород (Сюе, 2022; Саттон, 2018; Шехаде, 2024). Основним недоліком методу RL є необхідність навчання агента, що займає велику кількість часу (Шехаде, 2024). Іншим недоліком є висока чутливість ефективності роботи методу RL до зміни функції винагороди та її параметрів (Саттон, 2018; Зурарі, 2023).

Таким чином, порівнюючи методи планування маршрутів RL та MPC із позиції реального застосування на БПЛА, можна відзначити, що метод RL добре працює у випадках складної та мінливої обстановки, однак вимагає тривалого попереднього навчання й потужних обчислювальних ресурсів (Саттон, 2018; Ву, 2020); метод MPC враховує реалістичну динаміку БПЛА і забезпечує відносно невеликі помилки відхилення від бажаної траєкторії, проте потребує доволі точної математичної моделі та достатньої обчислювальної потужності для розв'язання оптимізаційної задачі в режимі реального часу (Канані, 2024; Аліярі, 2022).

У таблиці 1 наведено основні порівняльні характеристики методів RL та MPC для побудови маршрутів руху БПЛА.

Висновки. У результаті проведеного дослідження, було встановлено, що метод навчання з підкріпленням забезпечує гнучкість та здатність до адаптації у динамічних і невизначених середовищах, проте вимагає тривалої фази навчання, значних обчислювальних ресурсів та великих об'ємів даних. У той же час обчислювальна складність на етапі використання є суттєво нижчою, що дозволяє використовувати метод для роботи у реальному часі. Точна модель середовища не вимагається. Метод прогностичного керування успішно враховує кінематичні та динамічні обмеження БПЛА і дає змогу плавно адаптуватися до зміни умов, але залежить від наявності якісної моделі середовища, може бути складним для розв'язування у реальному часі в складних 3D-сценаріях. Перевагою методу MPC є його прогнозна природа, яка дає можливість враховувати майбутні прогнозовані стани при прийнятті рішень. Також є вибір між лінійною та нелінійною моделями для прогнозування майбутньої поведінки. Слід зазначити, що метод прогнозного керування став основою для розробки численних його модифікацій, зокрема: DMC, MAC, PFC, EPSAC, EHAC та GPC.

У якості перспективи подальших досліджень може бути комбінування методів RL та MPC для побудови маршрутів руху БПЛА з іншими розповсюдженими методами та алгоритмами побудови маршрутів руху. Наприклад, попереднє офлайн планування маршруту за допомогою алгоритмів A* чи ACO, з подальшою корекцією траєкторії з використанням методів MPC або RL у режимі реального часу. Подібні гібридні підходи дозволяють використати сильні сторони кожного методу. Іншою областю подальших досліджень може бути симуляція роботи конкретних реалізацій методів RL та MPC або проведення експериментальних польотів з метою отримання кількісних показників ефективності їх роботи.

Таблиця 1

Порівняльні характеристики методів RL та MPC

Характеристика	RL	MPC
Тип середовища	Динамічне, невизначене	Статичне або частково відоме
Необхідність точної моделі середовища	Не вимагається	Висока залежність від моделі
Адаптивність у динамічних середовищах	Висока	Залежить від точності моделі
Обчислювальна складність	Висока складність навчання; подальше застосування відносно швидке	Висока; оптимізація у реальному часі
Вимоги до даних	Вимагає великих об'ємів даних для навчання	Вимагає точних параметрів моделі

ЛІТЕРАТУРА:

1. Model Predictive Control for Optimal Motion Planning of Unmanned Aerial Vehicles / З.-Н. Буй та ін. In *proceedings of 2024, the 7th International Conference on Control, Robotics and Informatics (ICCRI 2024)*. *arXiv e-prints*, arXiv-2410. 2024. URL: 10.48550/arXiv.2410.09799
2. Motion planning of multiple unmanned vehicles using sequential convex programming-based distributed model predictive control / С. Лю та ін. *International Journal of Robust and Nonlinear Control*. 2025. <https://doi.org/10.1002/rnc.7633>.
3. Model Predictive Contouring Control with Barrier and Lyapunov Functions for Stable Path-Following in UAV systems / Б.С.Гевара та ін. *arXiv e-prints*, arXiv-2411. 2024. URL: <https://arxiv.org/abs/2411.00668>
4. Polydoros A.S., Nalpantidis L. Survey of Model-Based Reinforcement Learning: Applications on Robotics. *Journal of Intelligent & Robotic Systems*. 2017. Т. 86, № 2. С. 153–173. <https://doi.org/10.1007/s10846-017-0468-y>
5. Xue, Y., & Chen, W. UAV Navigation Approach Based on Deep Reinforcement Learning in Large Cluttered 3D Environments. *IEEE Transactions on Vehicular Technology*. 2022. С. 3001–3014. <https://doi.org/10.1109/tvt.2022.3218855>
6. Woo J., Kim N. Collision avoidance for an unmanned surface vehicle using deep reinforcement learning. *Ocean Engineering*. 2020. Т. 199. С. 107001. <https://doi.org/10.1016/j.oceaneng.2020.107001>
7. Reichensdörfer E., Odenthal D., Wollherr D. On the Stability of Nonlinear Wheel-Slip Zero Dynamics in Traction Control Systems. *IEEE Transactions on Control Systems Technology*. 2018. Т. 28, № 2. С. 489–504. <https://doi.org/10.1109/tcst.2018.2880932>
8. Kanani K.K.J., Tiwari B.K., Kim T.W. Linear Parameter Varying Model Predictive-Based Controller for Trajectory Tracking of High-Fidelity Autonomous Vehicle. *2024 9th International Conference on Automation, Control and Robotics Engineering (CACRE)*, м. Jeju Island, Korea, Republic of, 18–20 лип. 2024 р. 2024. С. 148–153. <https://doi.org/10.1109/cacre62362.2024.10635048>
9. Design and implementation of a constrained model predictive control approach for unmanned aerial vehicles / М.Аліярі та ін. *IEEE Access*. 2022. С. 91750–91762. <https://doi.org/10.1109/access.2022.3202020>
10. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. A Bradford Book, 2018. 552 с.
11. Holkar K. S., Waghmare L. M. An overview of model predictive control. *International Journal of control and automation*. 2010. Т. 3, № 4. С. 47–63.
12. Khather S.I., Ibrahim M.A., Abdullah A.I. Review and Performance Analysis of Nonlinear Model Predictive Control–Current Prospects, Challenges and Future Directions. *Journal Européen des Systèmes Automatisés*. 2023. Т. 56, № 4. С. 593–603. <https://doi.org/10.18280/jesa.560409>
13. Should We Use Model-Free or Model-Based Control? A Case Study of Battery Control / М.Ф.Ель-Хадж Шехадє та ін. 2024 56th North American Power Symposium (NAPS), м. El Paso, TX, USA, 13–15 жовт. 2024 р. 2024. С. 1–5. <https://doi.org/10.1109/naps61145.2024.10741791>
14. Reinforcement Q-Learning for Path Planning of Unmanned Aerial Vehicles (UAVs) in Unknown Environments / А.Зурпакі та ін. *International Review of Automatic Control (IREACO)*. 2023. Т. 16, № 5. С. 253–259. <https://doi.org/10.15866/ireaco.v16i5.24078>

REFERENCES:

1. Bui, D.N., Khuat, T.H., Duong Phung, M., Tran, T.H., & Tran, D.L. (2024). Model Predictive Control for Optimal Motion Planning of Unmanned Aerial Vehicles. In *proceedings of 2024, the 7th International Conference on Control, Robotics and Informatics (ICCRI 2024)*. *arXiv e-prints*, arXiv-2410. Retrieved from: 10.48550/arXiv.2410.09799
2. Liu, X., Wu, F., Deng, Y., Wang, M., & Xia, Y. (2025). Motion planning of multiple unmanned vehicles using sequential convex programming-based distributed model predictive control. *International Journal of Robust and Nonlinear Control*, 35(1), 141–167. <https://doi.org/10.1002/rnc.7633>
3. Guevara, B.S., Moya, V., Recalde, L.F., Gandolfo, D.C., & Toibero, J.M. (2024). Model Predictive Contouring Control with Barrier and Lyapunov Functions for Stable Path-Following in UAV systems. *arXiv e-prints*, arXiv-2411. Retrieved from: <https://arxiv.org/abs/2411.00668>
4. Polydoros, A. S., & Nalpantidis, L. (2017). Survey of model-based reinforcement learning: Applications on robotics. *Journal of Intelligent & Robotic Systems*, 86 (2), 153–173. <https://doi.org/10.1007/s10846-017-0468-y>
5. Xue, Y., & Chen, W. (2022). A UAV navigation approach based on deep reinforcement learning in large cluttered 3D environments. *IEEE Transactions on Vehicular Technology*, 72(3), 3001–3014. <https://doi.org/10.1109/tvt.2022.3218855>

6. Woo, J., & Kim, N. (2020). Collision avoidance for an unmanned surface vehicle using deep reinforcement learning. *Ocean Engineering*, 199, 107001. <https://doi.org/10.1016/j.oceaneng.2020.107001>.
7. Reichensdörfer, E., Odenthal, D., & Wollherr, D. (2018). On the stability of nonlinear wheel-slip zero dynamics in traction control systems. *IEEE Transactions on Control Systems Technology*, 28(2), 489–504. <https://doi.org/10.1109/tcst.2018.2880932>
8. Kanani, K. K. J., Tiwari, B. K., & Kim, T. W. (2024, 18–20 July). Linear Parameter Varying Model Predictive-Based Controller for Trajectory Tracking of High-Fidelity Autonomous Vehicle. In *2024 9th International Conference on Automation, Control and Robotics Engineering (CACRE)*, Jeju Island, Korea, Republic of. (pp. 148–153). IEEE. <https://doi.org/10.1109/cacre62362.2024.10635048>
9. Aliyari, M., Wong, W. K., Bouteraa, Y., Najafinia, S., Fekih, A., & Mobayen, S. (2022). Design and implementation of a constrained model predictive control approach for unmanned aerial vehicles. *IEEE Access*, 10, 91750–91762. <https://doi.org/10.1109/access.2022.3202020>
10. Sutton R. S., Barto A. G. (2018). Reinforcement Learning: An Introduction. A Bradford Book. 552 p.
11. Holkar, K. S., & Waghmare, L. M. (2010). An overview of model predictive control. *International Journal of control and automation*, 3(4), 47–63.
12. Khather, S. I., Ibrahim, M. A., & Abdullah, A. I. (2023). Review and Performance Analysis of Nonlinear Model Predictive Control--Current Prospects, Challenges and Future Directions. *Journal Européen des Systèmes Automatisés*, 56(4). 593–603. <https://doi.org/10.18280/jesa.560409>
13. Chehade, M. F. E. H., Cho, Y. H., Chinchali, S., & Zhu, H. (2024, October). Should We Use Model-Free or Model-Based Control? A Case Study of Battery Control. In *2024 56th North American Power Symposium (NAPS)* (pp. 1–5). IEEE. <https://doi.org/10.1109/naps61145.2024.10741791>
14. Zourari, A., Youssefi, M., Ben Youssef, Y., Dakir, R., Reinforcement Q-Learning for Path Planning of Unmanned Aerial Vehicles (UAVs) in Unknown Environments, (2023) *International Review of Automatic Control (IREACO)*, 16(5), pp. 253–259. <https://doi.org/10.15866/ireaco.v16i5.24078>