Reinforcement Learning in Control Theory: A New Approach to Mathematical Problem Solving

Kala Agbo Bidi[*]
Sorbonne University

Jean-Michel Coron[*]
Sorbonne University

Amaury Hayat[*]
Ecole des Ponts Paristech

Nathan Lichtlé[*]
Ecole des Ponts Paristech & UC Berkeley

Abstract

One of the central questions in control theory is achieving stability through feedback control. This paper introduces a novel approach that combines Reinforcement Learning (RL) with mathematical analysis to address this challenge, with a specific focus on the Sterile Insect Technique (SIT) system. The objective is to find a feedback control that stabilizes the mosquito population model. Despite the mathematical complexities and the absence of known solutions for this specific problem, our RL approach identifies a candidate solution for an explicit stabilizing control. This study underscores the synergy between AI and mathematics, opening new avenues for tackling intricate mathematical problems.[1]

1 Introduction

AI for mathematics often refers to automatic theorem proving, either in formal language [39, 31, 26, 40], or in natural language [27], usually using language models. This article takes a different approach, presenting a Reinforcement Learning (RL) framework to solve a mathematical problem from control theory. The goal is to help mathematicians by finding a candidate solution to the problem in the spirit that checking a solution is often easier than finding it.

Control theory is about asking oneself: "if I can act on a system, what can I make it do?". In this area of mathematics, the system in question is usually a described by a set of differential (or partial differential) equations in which there is a component –called control– that can be chosen. One of the main branches of this field, called stabilization, aims to find a way to make an equilibrium stable by choosing this control as a function of the state of the system. This is called a feedback control. Many mathematical techniques exist to solve this problem [14]. However, in some cases the current mathematical theories are unable to find a successful feedback control. In this article we show that an RL approach combined with a mathematical analysis can help to find new mathematical feedback controls in such complicated cases.

We study a practical case: the SIT system, that models the control of insect pests (in particular mosquito population). SIT stands for *Sterile Insect Technique* which consists in releasing sterilized insects to reduce or eliminate a target population. Initially used in agriculture to control insect pests, it is today employed in the vector-born disease fight against mosquitoes that carry illnesses such as malaria and arboviruses [2, 4] and there is a great interest both in research and in practice to understand which control to use [9, 8, 3]. A more detailed overview of the literature is given in Appendix C. From a mathematical point of view, without any control, this system of differential

---

[*]The authors contributed equally.
[1]The code and trained models are released at https://github.com/nathanlct/rl-control.

equations has a globally stable undesired equilibrium with high population of insect pests, and an unstable equilibrium with no insect pests, the *zero equilibrium*. The mathematical goal is to find a feedback control such that the zero equilibrium is globally stable instead of the unwanted equilibrium.

The difficulties come from three reasons: showing global stability of dynamical systems (as opposed to local stability) is a very challenging mathematical problem for which there are only few mathematical tools; the system is not continuous, which is known in mathematics to bring some difficulties; we only have a partial measurement of the state of the system (see Section 4.2 for more details). Because of these three difficulties, finding a feedback control for this system is a mathematically open question. With our approach we are able to derive an explicit mathematical feedback control to achieve the global stability. While there is still no mathematical proof that this feedback control is a solution to the problem, the numerical simulations strongly suggest that it is. We believe that this approach could be generalized to other open problems in control theory and give a new impulse to their resolution.

2 Mathematical framework

The dynamical system we consider is the SIT model for pests, here mosquitoes, given by

$$\dot{E} = \beta_E F \left(1 - \frac{E}{K}\right) - (\nu_E + \delta_E)E, \tag{2.1}$$

$$\dot{M} = (1 - \nu)\nu_E E - \delta_M M, \tag{2.2}$$

$$\dot{F} = \nu\nu_E E \frac{M}{M + M_s} - \delta_F F, \tag{2.3}$$

$$\dot{M}_s = u - \delta_s M_s, \tag{2.4}$$

where $E(t) \geq 0$ represents the mosquito density in the aquatic phase, $M(t) \geq 0$ the wild adult male density, $F(t) \geq 0$ the density of adult feconded females, $M_s(t) \geq 0$ the sterilized adult male density, and $u(t) \geq 0$, the control, is the density of sterilized males released at time t. We also denote the number of unfertilized females by $F_s(t) = F(t)M_s(t)/M(t)$. All SIT model notation is defined in Appendix A as well as the parameters we used.

When $u(t) = M_s(t) = 0$ for any $t \geq 0$, the system (2.1)–(2.3) has a unique globally asymptotically stable equilibrium $(E(t), M(t), F(t)) \equiv (E^*, M^*, F^*)$ where E^*, M^* and F^* are large constant values. This corresponds to the situation where mosquitoes reproduce freely. The state $(E(t), M(t), F(t)) \equiv (0, 0, 0)$ is also an equilibrium, albeit an unstable one. The mathematical problem is to find $u(t)$ of the form

$$u(t) = f(M(t) + M_s(t), F(t) + F_s(t)), \tag{2.5}$$

where $f \in L^\infty(\mathbb{R}^2)$ such that the zero equilibrium $(0, 0, 0)$ is globally asymptotically stable and M_s is asymptotically small, meaning there exists $c \in \mathbb{R}_+$ such that

$$\lim_{t \to +\infty} \|u(t)\| = c < U^* := \frac{K\beta_E\nu(1-\nu)\nu_E^2\delta_s}{4(\delta_E + \nu_E)\delta_F\delta_M} \left(1 - \frac{\delta_F(\nu_E + \delta_E)}{\beta_E\nu\nu_E}\right)^2, \tag{2.6}$$

and the equilibrium $(0, 0, 0, c/\delta_s)$ of system (2.1)–(2.4) is globally asymptotically stable (see Definition 2.1 below).

Definition 2.1. The equilibrium $(0, 0, 0, c/\delta_s)$ of the system (2.1)–(2.4) is *globally asymptotically stable* if, for any initial condition $(E_0, M_0, F_0, M_{s,0})$ there exists a unique solution (E, M, F, M_s) on $[0, +\infty)$ to the system (2.1)–(2.4) and for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|(E_0, M_0, F_0, M_{s,0} - c/\delta_s)\| \leq \delta \implies \|(E(t), M(t), F(t), M_s(t) - c/\delta_s)\| \leq \varepsilon, \ \forall t \in [0, +\infty), \tag{2.7}$$

$$\lim_{t \to +\infty} \|(E(t), M(t), F(t), M_s(t) - c/\delta_s)\| = 0, \tag{2.8}$$

The form constraint (2.5) corresponds to a practical limitation: $M + M_s$ and $F + F_s$ are the total number of males and females which are typically what can be measured in practice (see [2]).

Remark 2.1 (Constant control). The point of the constraint (2.6) is to avoid a constant control. Indeed, for a constant control $u(t) \equiv \bar{U}$, if $\bar{U} > U^*$ then the equilibrium $(0, 0, 0, U^*/\delta_s)$ is globally asymptotically stable (see [3]). In practice, one would like c to be as small as possible in (2.6). For the control we find in Section 5 with our approach, the value of c is much smaller than U^*. In fact, we even show a simplified version of the control where c can be chosen arbitrarily small.

Remark 2.2 (Optimal decay rate). Assume that $E(0) \leq K$. Then, whatever is the control $u(t) \geq 0$, for every time $t \geq 0$, $E(t) \leq K$ and

$$E(t) \geq \tilde{E}(t), \ M(t) \geq \tilde{M}(t), \ F(t) \geq \tilde{F}(t), \tag{2.9}$$

where $(\tilde{E}, \tilde{M}, \tilde{F})$ is the solution to the Cauchy problem

$$\dot{\tilde{E}} = \beta_E \tilde{F} \left(1 - \frac{\tilde{E}}{K}\right) - \left(\nu_E + \delta_E\right)\tilde{E}, \ \dot{\tilde{M}} = (1 - \nu)\nu_E \tilde{E} - \delta_M \tilde{M}, \ \dot{\tilde{F}} = -\delta_F \tilde{F}, \tag{2.10}$$

$$(\tilde{E}(0), \tilde{M}(0), \tilde{F}(0)) = (E(0), M(0), F(0)). \tag{2.11}$$

It would be interesting to see if one can get with suitable output feedback laws (vanishing or small at the origin) a decay rate close to the one imposed by (2.9), i.e.

$$E(t) \simeq \tilde{E}(t), \ M(t) \simeq \tilde{M}(t), \ F(t) \simeq \tilde{F}(t). \tag{2.12}$$

(Note that it is possible to get (2.12) by taking u constant and large, depending on $(E(0), M(0), F(0))$.) This would be particularly useful in the case where the insect under study reproduces both sexually and asexually: indeed, this would give the best way to reduce the sexual reproduction part as much as possible by output feedback laws (vanishing or small at the origin).

3   Related works

From a control theory perspective, several mathematical approaches have already been used in the literature to treat this problem either for the complete system (2.1)–(2.4) or for reduced models, using classical tools in control theory (control Lyapunov functions, LaSalle invariance principle, maximum principle, monotone dynamical systems, etc. [3, 1, 7, 5, 6, 15]). See Appendix C for more details. In particular, a feedback control was found in [1], however this control depends on the four variables (E, M, F, M_s) and not only on the observable quantities $M + M_s$ and $F + F_s$. As of now, there is no known control depending only on $M + M_s$ and $F + F_s$.

Over the past few years, RL has emerged as a powerful approach for control, in a wide range of domains and applications (see Appendix C). However, RL techniques, while powerful for decision-making, inherently provide control mechanisms that are discrete and numerical in nature. From a more rigorous mathematical point of view, these mechanisms often do not translate directly into analytical feedback control formulas. Aiming to address this limitation, in our work, we blend RL methodologies with mathematical analysis to extract an explicit mathematical control.

Using AI tools to help mathematicians by giving them an insight or a candidate solution was considered in [17] using a different framework. Other approaches aimed to teach a model to guess mathematical solutions to a problem [25, 13]. However, in these approaches the solution of the mathematical problems involved are already known.

4   Method

4.1   Our approach

Our proposed approach works in four steps, summarized in Figure 1:

(Step 1) **Discretize the equations** in a numerical scheme and use those dynamics to create a training environment by implementing the observations, actions, and rewards described in Section 4.2.

(Step 2) **Train an RL model** that learns to maximize the objective function we assign it through many simulations and obtain a numerical control feedback based on this numerical scheme.
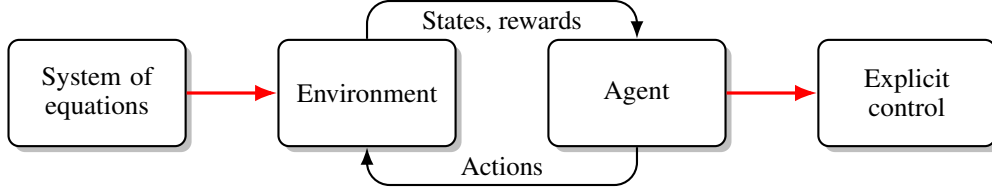
Figure 1: Diagram representing the procedure by which we simulate our model in an environment that is used to train an RL agent, whose policy we then convert into an explicit control.

(Step 3) **Recover an explicit mathematical control** from the numerical control feedback.

(Step 4) **Perform several tests** using different numerical schemes and discretizations to ensure that the explicit control is efficient.

4.2 Reinforcement-Learning Framework

Reinforcement Learning (RL) trains agents to optimize long-term rewards in various environments by maximizing the expected cumulative sum of rewards, denoted as $J(\pi_\theta)$, where π_θ is the policy, often guided by neural network weights θ. Similar to minimizing a cost function $-J(\pi_\theta)$ in control theory, this can be modeled using a partially-observable Markov decision process (POMDP). The observation space has two states: total males $M + M_s$ and females $F + F_s$, as the individual states E, M, F and M_s aren't independently measurable in the real world. Observations are normalized to $[0, 1]$. For better convergence, $M + M_s$ and $F + F_s$, which can typically range from 0 to $100K$, are inputted to the neural network at varied scales and normalized. The single action a_t ranges from $[-1,1]$, remapped to $u(t) \in [0, 10K]$ for model equations (2.1)–(2.4). To simplify training by artificially reducing the horizon, each action is repeated for multiple simulation steps.

Finally, our optimization criterion takes the following form at time step t:

$$r_t = c_1 \|E(t), M(t), F(t)\|_2 + c_2(t)\|M_s(t)\|_2 \qquad \text{with} \qquad c_2(t) = \begin{cases} c_3 & \text{if } t < 0.9T, \\ c_3 + c_4 & \text{otherwise.} \end{cases}$$

Near the horizon's end, specifically when $0.9T \le t \le T$, we introduce a positive weight c_4 to the penalty on $\|M_s(t)\|_2$. This aids steer the RL training toward the desired asymptotic convergence of the state. Parameters and more experiment details can be found in Appendix B.

5 Main Results

The trained RL model converges after around 10M steps to a numerical control that we represent in Figure 2 as a function of $M + M_s$ and $F + F_s$. Interestingly, we see that the plot of the control in linear scale is not really informative (see Figure 2 left), suggesting it would be complicated to have a good regression directly as a function of $M + M_s$ and $F + F_s$. However, in logarithmic scale the expression of the control seems much clearer (see Figure 2 right) and clearly has two identified parts. In each of them the control seems to be close to a bang-bang control with a thin transition. With a simple regression (see Figure 3) we approximate this numerical control with the explicit mathematical control

$$u_{\text{reg}}(M + M_s, F + F_s) = \begin{cases} u_{\text{reg}}^{\text{left}}(M + M_s, F + F_s) & \text{if } M + M_s < M^*, \\ u_{\text{reg}}^{\text{right}}(M + M_s, F + F_s) & \text{otherwise,} \end{cases} \tag{5.1}$$

where u is defined on $(0, +\infty)^2$ and

where

$$u_{\text{reg}}^{\text{left}} = \begin{cases} u_{\min} & \text{if } I_1(F + F_s) > \alpha_2, \\ u_{\max}(\alpha_2 - I_1) & \text{if } I_1 \in (\alpha_1, \alpha_2], \\ u_{\max} & \text{otherwise,} \end{cases} \qquad u_{\text{reg}}^{\text{right}} = \begin{cases} u_{\min} & \text{if } I_2 > \alpha_2, \\ u_{\max}(\alpha_2 - I_2) & \text{if } I_2 \in (\alpha_1, \alpha_2], \\ u_{\max} & \text{otherwise.} \end{cases}$$

$I_1(x) = \frac{\log(M^*)}{\log(x)}$ and $I_2(x, y) = \frac{\log(x)}{\log(y)}$, $M^* = 200$, $\alpha_1 = 3$, $\alpha_2 = 4$, $u_{\max} = 3 \cdot 10^5$ is imposed by physical constraints and $u_{\min}$ can be chosen.
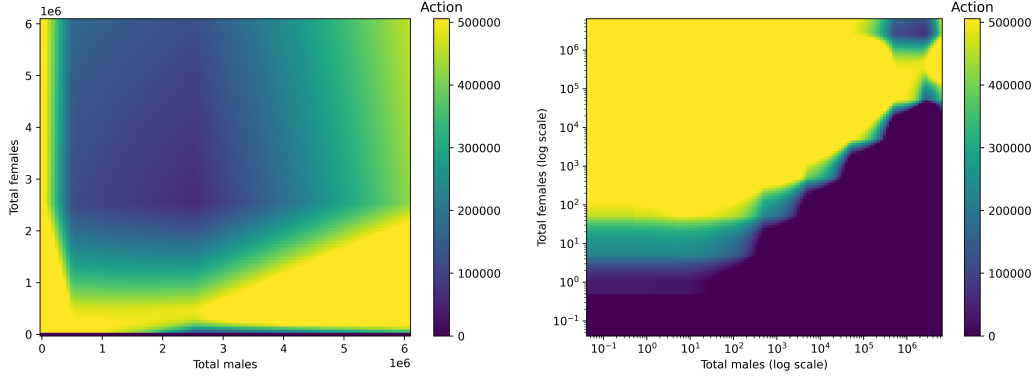
4

Figure 2: Heatmap of the model's action u as a function of $M + M_s$ and $F + F_s$, in linear scale (left) and logarithmic scale (right).

During training and evaluation the numerical feedback control includes a slight noise to enhance robustness and exploration. Surprisingly, when tested with $u_{\min} = 0$, the mathematical control (5.1) with a small additional noise $\eta(t)$ exhibits asymptotic stability, whereas noise-free control does not, displaying cyclic-like behavior (see Figure 3). The paradox arises from $u_{\min} = 0$. Introducing a small positive value ε for this parameter is enough to obtain asymptotic stability. In the noisy control, because of the condition $u \geq 0$, $u_{\min} + \eta(t)$ is positive in average, explaining the stabilization. More details are given in Appendix D.1.

The efficiency of the feedback control (5.1) is illustrated in Appendix D on many numerical simulations for a large array of initial conditions with different discretizations, suggesting that this nonlinear control is indeed a solution to the mathematical problem considered. We recall that there is no known solution to date for this mathematical problem. With this control, the system converges globally to the equilibrium $(E^*, M^*, F^*, M_s^*) = (0, 0, 0, \lambda)$ where $\lambda = \varepsilon/\delta_s$ and ε can be chosen much smaller than the U^* given in (2.6). A more detailed analysis of the result is given in Appendix D.

Using the candidate solution provided by the method, we simplified (5.1) and obtained a second candidate feedback control given by

$$v_{\text{reg}}(M + M_s, F + F_s) = \begin{cases} u_{\min} & \text{if } \frac{\log(M+M_s)}{\log(F+F_s)} > \alpha_2, \\ u_{\max} & \text{otherwise,} \end{cases} \tag{5.2}$$

where $\alpha_2 = 4$ with the parameters of Table 1 and $u_{\max} = 3 \cdot 10^5$ is still imposed. Interestingly, not only is this control faster to converge (compare Tables 2 and 3) but additionally one can choose $u_{\min} = \varepsilon > 0$ arbitrarily small. In Figure 5 we represent $\|E(t), M(t), F(t)\|$ and $u(t)$ as a function of time for $\varepsilon = 10^{-2}$, 1 and 5. We see that the curves of $\|E(t), M(t), F(t)\|$ are very similar, the main difference being that $u(t)$ takes more time to converges to 0 as $u_{\min}$ grows larger. Interestingly, taking $\varepsilon = 0$ does not lead to the converges of the equilibrium, suggesting that there is a mathematical bifurcation. Note that being able to take $\varepsilon > 0$ arbitrarily small is a much more powerful property than the one given by (2.6).

6   Discussion and conclusion

We presented an RL framework to solve a type of mathematical problem in control theory, and we used it to find an explicit candidate solution for the stabilization of the SIT system. In the future, it would be interesting to use a regression to explicit the reward J as a function of the initial state in order to obtain an explicit Lyapunov function which would show the asymptotic stability of the system with the explicit control. Going further, this approach could be likely generalized to other systems. This is an incentive to use more AI techniques to solve mathematical problems, especially in control theory.

5

References

[1] Kala Agbo Bidi, Luis Almeida, and Jean-Michel Coron. Global stabilization of sterile insect technique model by feedback laws. *arXiv, 2307.00846*, 2023.

[2] Luis Almeida, Michel Duprez, Yannick Privat, and Nicolas Vauchelet. Mosquito population control strategies for fighting against arboviruses. *Mathematical Biosciences and Engineering*, 16(6):6274–6297, 2019.

[3] Luís Almeida, Michel Duprez, Yannick Privat, and Nicolas Vauchelet. Optimal control strategies for the sterile mosquitoes technique. *Journal of Differential Equations*, 311:229–266, 2022.

[4] Nina Alphey, Luke Alphey, and Michael B Bonsall. A model framework to estimate impact and cost of genetics-based sterile insect methods for dengue vector control. *PLoS One*, 6(10):e25384, 2011.

[5] Roumen Anguelov, Yves Dumont, and Jean Lubuma. Mathematical modeling of sterile insect technology for control of anopheles mosquito. *Computers & Mathematics with Applications*, 64(3):374–389, 2012.

[6] Roumen Anguelov, Yves Dumont, and Ivric Valaire Yatat Djeumen. Sustainable vector/pest control using the permanent sterile insect technique. *Mathematical Methods in the Applied Sciences*, 43(18):10391–10412, 2020.

[7] H Barclay and M Mackauer. The sterile insect release method for pest control: a density-dependent model. *Environmental Entomology*, 9(6):810–817, 1980.

[8] Pierre-Alexandre Bliman. Feedback control principles for biological control of dengue vectors. In *2019 18th European Control Conference (ECC)*, pages 1659–1664. IEEE, 2019.

[9] Pierre-Alexandre Bliman, M Soledad Aronna, Flávio C Coelho, and Moacyr AHB da Silva. Ensuring successful introduction of wolbachia in natural populations of aedes aegypti by means of feedback control. *Journal of mathematical biology*, 76:1269–1300, 2018.

[10] Pierre-Alexandre Bliman, Daiver Cardona-Salgado, Yves Dumont, and Olga Vasilieva. Implementation of control strategies for sterile insect techniques. *Math. Biosci.*, 314:43–60, 2019.

[11] Pierre-Alexandre Bliman and Yves Dumont. Robust control strategy by the Sterile Insect Technique for reducing epidemiological risk in presence of vector migration. *Math. Biosci.*, 350:Paper No. 108856, 23, 2022.

[12] Eric W Chambers, Limb Hapairai, Bethany A Peel, Herve Bossin, and Stephen L Dobson. Male mating competitiveness of a wolbachia-introgressed aedes polynesiensis strain under semi-field conditions. *PLoS neglected tropical diseases*, 5(8):e1271, 2011.

[13] Francois Charton, Amaury Hayat, and Guillaume Lample. Learning advanced mathematical computations from examples. In *International Conference on Learning Representations*, 2020.

[14] Jean-Michel Coron. *Control and nonlinearity*. American Mathematical Soc., 2007.

[15] Andrea Cristofaro and Luca Rossi. Backstepping control for the sterile mosquitoes technique: stabilization of extinction equilibrium. working paper or preprint, 2023.

[16] Andrea Cristofaro and Luca Rossi. Backstepping control for the sterile mosquitoes technique: stabilization of extinction equilibrium. *Preprint*, 2023.

[17] Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with ai. *Nature*, 600(7887):70–74, 2021.

[18] Amir-massoud Farahmand, Saleh Nabi, and Daniel N. Nikovski. Deep reinforcement learning for partial differential equation control. In *2017 American Control Conference (ACC)*, pages 3120–3127, 2017.

[19] Shixiang Gu, Ethan Holly, Timothy Lillicrap, and Sergey Levine. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3389–3396, 2017.

[20] L Hapairai. *Studies on Aedes polynesiensis introgression and ecology to facilitate lymphatic filariasis control*. PhD thesis, Oxford University, UK, 2013.

[21] Limb K Hapairai, Jérôme Marie, Steven P Sinkins, and Hervé C Bossin. Effect of temperature and larval density on aedes polynesiensis (diptera: Culicidae) laboratory rearing productivity and male characteristics. *Acta tropica*, 132:S108–S115, 2014.

[22] Limb K Hapairai, Michel A Cheong Sang, Steven P Sinkins, and Hervé C Bossin. Population studies of the filarial vector aedes polynesiensis (diptera: Culicidae) in two island settings of french polynesia. *Journal of medical entomology*, 50(5):965–976, 2013.

[23] Harriet Hughes and Nicholas F Britton. Modelling the use of wolbachia to control dengue fever transmission. *Bulletin of mathematical biology*, 75:796–818, 2013.

[24] Bahare Kiumarsi, Kyriakos G Vamvoudakis, Hamidreza Modares, and Frank L Lewis. Optimal and autonomous control using reinforcement learning: A survey. *IEEE transactions on neural networks and learning systems*, 29(6):2042–2062, 2017.

[25] Guillaume Lample and François Charton. Deep learning for symbolic mathematics. In *International Conference on Learning Representations*, 2019.

[26] Guillaume Lample, Marie-Anne Lachaux, Thibaut Lavril, Xavier Martinet, Amaury Hayat, Gabriel Ebner, Aurélien Rodriguez, and Timothée Lacroix. HyperTree Proof Search for Neural Theorem Proving. *Advances in neural information processing systems*, 2022.

[27] Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. Solving quantitative reasoning problems with language models. *Advances in Neural Information Processing Systems*, 35:3843–3857, 2022.

[28] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[29] A. Rupam Mahmood, Dmytro Korenkevych, Gautham Vasan, William Ma, and James Bergstra. Benchmarking reinforcement learning algorithms on real-world robots, 2018.

[30] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[31] Stanislas Polu, Jesse Michael Han, Kunhao Zheng, Mantas Baksys, Igor Babuschkin, and Ilya Sutskever. Formal mathematics statement curriculum learning. In *11th International Conference on Learning Representations*, 2022.

[32] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

[33] François Rivière. Ecologie de aedes (stegomyia) polynesiensis, marks, 1951 et transmission de la filariose de bancroft en polynesie. *Paris (France): Université Paris-Sud*, 1988.

[34] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[35] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[36] Martin Strugarek, Hervé Bossin, and Yves Dumont. On the use of the sterile insect release technique to reduce or eliminate mosquito populations. *Applied Mathematical Modelling*, 68:443–470, 2019.

[37] Takeshi Suzuki and Fola Sone. Breeding habits of vector mosquitoes of filariasis and dengue fever in western samoa. *Medical Entomology and Zoology*, 29(4):279–286, 1978.

[38] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

[39] Yuhuai Wu, Albert Qiaochu Jiang, Jimmy Ba, and Roger Grosse. Int: An inequality benchmark for evaluating generalization in theorem proving. *arXiv preprint arXiv:2007.02924*, 2020.

[40] Yuhuai Wu, Albert Qiaochu Jiang, Wenda Li, Markus Rabe, Charles Staats, Mateja Jamnik, and Christian Szegedy. Autoformalization with large language models. *Advances in Neural Information Processing Systems*, 35:32353–32368, 2022.

A Mathematical system interpretation and parameters

In system (2.1)–(2.4), we assumed that all females are immediately fertilized when they emerge from the pupal stage. The equation on F makes sense when we add the sterile male in which case only a fraction of the females will be fertilized. The interpretation of the parameters are given below [3]:

- $\beta_E > 0$ is the oviposition rate,
- $\delta_E, \delta_M, \delta_F > 0$ are the death rates for eggs, wild adult males and fertilized females respectively,
- $\nu_E > 0$ is the hatching rate for eggs,
- $\nu \in (0, 1)$ the probability that a pupa gives rise to a female (and $(1 - \nu)$ is, therefore, the probability to give rise to a male),
- $\delta_s > 0$ is the death rate of sterilized adult,
- $K > 0$ is the environmental capacity for eggs. It can be interpreted as the maximum density of eggs that females can lay in breeding sites. Since here the larval and pupal compartments are not present, it is as if E represents all the aquatic compartments in which case in this term K represents a logistic law's carrying capacity for the aquatic phase that also includes the effects of competition between larvae.

Besides, we also assume that $\delta_s \geq \delta_M$, which is usually considered as a biologically relevant assumption [3] Typical values for these parameters can be found in a population of Aedes polynesiensis in French Polynesia[36, 23, 33, 37, 12, 21, 22, 20] and are given in Table 1.

Parameter	Name	Value Interval	Chosen Value	Unity
β_E	Effective fecundity	[7.46, 14.85]	8	Day^{-1}
ν_E	Hatching parameter	[0.005, 0.25]	0.25	Day^{-1}
δ_E	Aquatic phase death rate	[0.023, 0.046]	0.03	Day^{-1}
δ_F	Female death rate	[0.033, 0.046]	0.04	Day^{-1}
δ_M	Males death rate	[0.077, 0.139]	0.1	Day^{-1}
δ_s	Sterilized male death rate	-	0.12	Day^{-1}
ν	Probability of emergence	-	0.49	
K	Environmental capacity for eggs	-	50000	

Table 1: Parameters for the system (2.1)–(2.4).

B Experiment details

We train our RL policies using proximal policy optimization (PPO) [34], a state-of-the-art policy gradient algorithm. We use the implementation of PPO provided in Stable Baselines 3 [32] (version 1.6.2, Python 3.8), a popular RL library that provides a collection of state-of-the-art algorithm implementations, as well as various tools for RL research.

The models are trained for 10 million environment timesteps (or 7 billion simulation timesteps) on 12 CPUs, which takes about 7 hours. During each iteration, we collect 12288 (1024 per CPU) environment steps, then run 5 epochs of optimization with a batch size of 1024. The agent's policy is a fully-connected neural network with 2 hidden layers of 256 neurons each, with $\tanh$ non-linearities between each layer, outputting the mean and standard deviation of a normal distribution that is then used to sample the action. More formally, for a given observation vector, the neural network policy outputs a mean μ_t and standard deviation σ_t and the action is sampled as $a_t \sim \mathcal{N}(\mu_t, \sigma_t)$. We train with a learning rate of 3×10^{-4}, gamma factor $\gamma = 0.99$, and all other hyperparameters are left to their default values.

We run each simulation for $T = 1001$ days (or 143 weeks), with a timestep $dt = 0.01$ days, and each action is repeated 700 times, meaning that the environment horizon is 143 steps and a new action is taken each week. For each simulation, the initial condition is uniformly sampled between 0 and $10K$: $E(0), M(0), F(0), M_s(0) \sim \mathcal{U}(0, 10K)$. For our reward function, we use coefficients $c_1 = 0.1, c_3 = 0.001$ and $c_4 = 0.01$.

C   Related works

From a mathematical point of view, several mathematical techniques have been used, either for this model or reduced models. In particular, two reduced model have been considered: A two dimensional model (2D-model) obtained by assuming that the dynamics of males and eggs are fast so that these two populations can be assumed to be at equilibrium (see [3, $(\mathcal{S}_1)$, page 231-232] or [16, (2)]) and a three dimensional model (3D-model) obtained by overlooking the non-adult stages (see[10, (7a)-(7b)-(7c)]). These mathematical approaches have led to the following stabilizing feedback controls:

- Stabilization using impulsive feedback controls for the 3D-model: [10, Theorem 6], and [10, Theorem 7] for the case of sparse measurements. The case of vector migration is also considered in [11].
- Stabilization using optimal feedback controls for the 2D-model: [3, Remark 4].
- Stabilization using the backstepping method: see [16] for the 2D-model and [1, Section 3.1] for the complete model.
- Stabilization using simple linear feedback laws for which the stabilization for the complete model is conjectured and proved for positively invariant subsets [1, Sections 3.2 and 3.3].

Our approach differs by using deep reinforcement learning to construct control feedback laws. In the past few years, RL has emerged as a powerful approach for control, leveraging its ability to learn near-optimal decision making strategies through interactions with an environment, and has demonstrated remarkable successes across a wide range of domains and applications. In robotics, RL has enabled machines to learn complex control tasks such as locomotion, manipulation, and dexterous object handling [19, 29, 28]. In the realm of games, RL algorithms have achieved superhuman performance in challenging domains like Go, chess, and StarCraft [35, 38]. Moreover, RL has excelled in playing classic Atari games, surpassing human-level performance by learning directly from pixel inputs [30]. RL can also excel in controlling ODE/PDE problems directly: [24] applies RL to optimal control problems, [18] applies RL to a flow control problem modeled by PDEs. However our approach differs from these by fitting an explicit mathematical control law using the learned neural network. These remarkable achievements highlight the versatility and potential of RL as a general-purpose approach for solving complex practical control problems in diverse domains.

D   Results

D.1   Effect of the noise

During training and testing the numerical control feedback law includes by default a small noise. This ensures some robustness of the control and a good exploration. We tested the mathematical control we derived (given in (5.1)) with and without noise. To our surprise, the control with a small noise does seem to ensure the asymptotic stability, whereas the control without any noise does not seem to. Indeed, without noise, the control seems to have a cyclic behavior and never converges (see Figure 3 (left)). When adding a small noise, however, the stability is restored (see Figure 3 (right)). The explication to this apparent paradox is that having exactly $u_{\min} = 0$ in the one of the branch of the control given in (5.1) is apparently too strong to allow the model to converge completely. Replacing the value with $u_{\min} > 0$ for a small $u_{\min}$ (typically $u_{\min} = 10$) allows to stabilize the system without noise (see Figure 3 (right)). In the system with noise, because the control u has to be positive, the noise increases in average the effective value of $u_{\min}$ of the control (5.1), which explains the apparent stabilization. Note that $(0, 0, 0, u_{\min}/\delta_s)$ is an equilibrium of the system stabilized, which solves the problem described in Section 5 provided that $u_{\min} < U^*$. Here, with the values of Table 1,

$$U^* \approx 1.6 \ 10^5 \text{ while } u_{\min} = 10,$$

which means that this control is a very good solution to the problem.

D.2   Effectiveness of the mathematical control

We test the candidate mathematical control (5.1) with random initial condition $(E_0, M_0, F_0, M_{s,0})$ in $[0, 10K]$ to check the stability of the equilibrium $(0, 0, 0, u_{\min}/\delta_s)$. This is represented in Figure 4. We also show the numerical values obtained and their variance in Table 2.
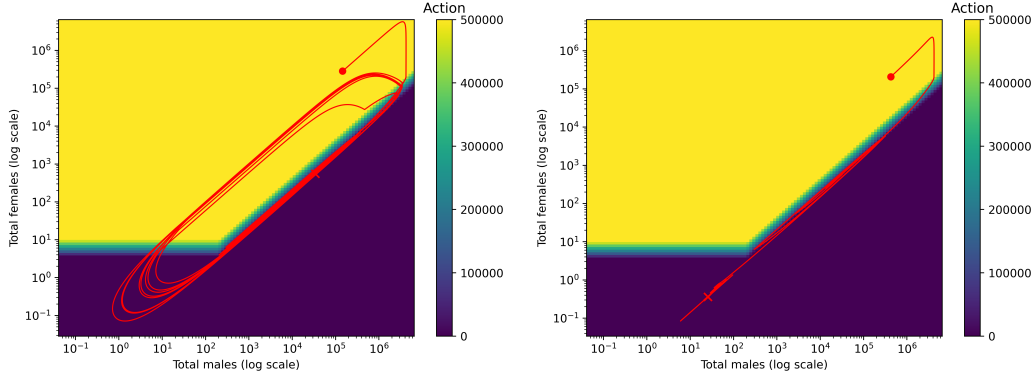
Figure 3: Heatmap of the regression model's action $u(M + M_s, F + F_s)$ as a function of total males and total females. A state-space trajectory is plotted in red, with the dot indicating initial state and the cross final state, for the heatmap only (left) and when a small noise $\mu \sim \mathcal{N}(0, 5)$ is added on top of the action (right).

	200 days	400 days	600 days	800 days
average $|E| + |M| + |F|$	50,801.64	8,020.47	96.25	0.60
variance $|E| + |M| + |F|$	45,422,159	4,012,119	1,394	0.02
maximum $|E| + |M| + |F|$	59,026.78	10,455.31	149.03	0.80
average $|M_s|$	2,207,795.88	693,675.84	16,743.12	50.27
variance $|M_s|$	53,101,242,728	13,102,504,436	41,355,608	81.24
maximum $|M_s|$	2,473,954.23	822,154.59	25,783.57	70.59

Table 2: Statistics over 100 simulations with random initial conditions in $[0, 10K]^4$ using control u_{reg} (see (5.1)) with $u_{\min} = 5$ and $u_{\max} = 300000$ over a duration of 600 days.

	200 days	400 days	600 days	800 days
average $|E| + |M| + |F|$	48,806.91	688.68	2.47	0.002
variance $|E| + |M| + |F|$	75,826,146	78,547.67	1.31	1.61×10^{-6}
maximum $|E| + |M| + |F|$	59,079.04	1,130.92	4.37	0.006
average $|M_s|$	2,500,000	129,308.46	2,204.39	41.67
variance $|M_s|$	3.07×10^{-11}	2,949,520,902	5,197,430.54	2.17×10^{-7}
maximum $|M_s|$	2,500,000	248,387.02	10,757.19	41.67

Table 3: Statistics over 100 simulations with random initial conditions in $[0, 10K]^4$ using control v_{reg} (see (5.2)) with $u_{\min} = 5$ and $u_{\max} = 300000$ over a duration of 600 days.
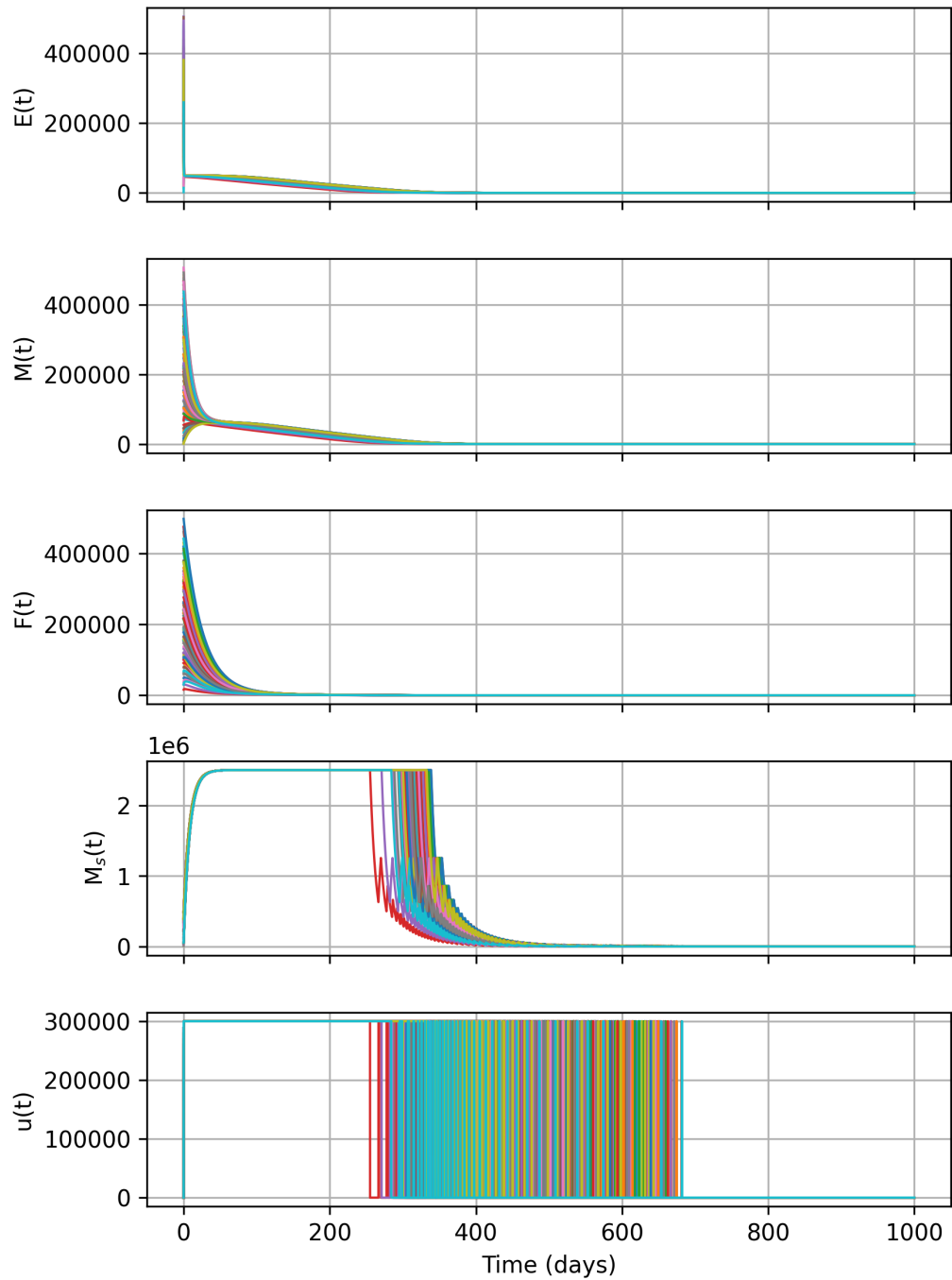
Figure 4: States and control v_{reg} with $u_{min} = 5$ and $u_{max} = 300000$ over a duration of 1000 days for 100 simulations with random initial conditions in $[0, 10K]^4$. Each color correspond to a simulation.
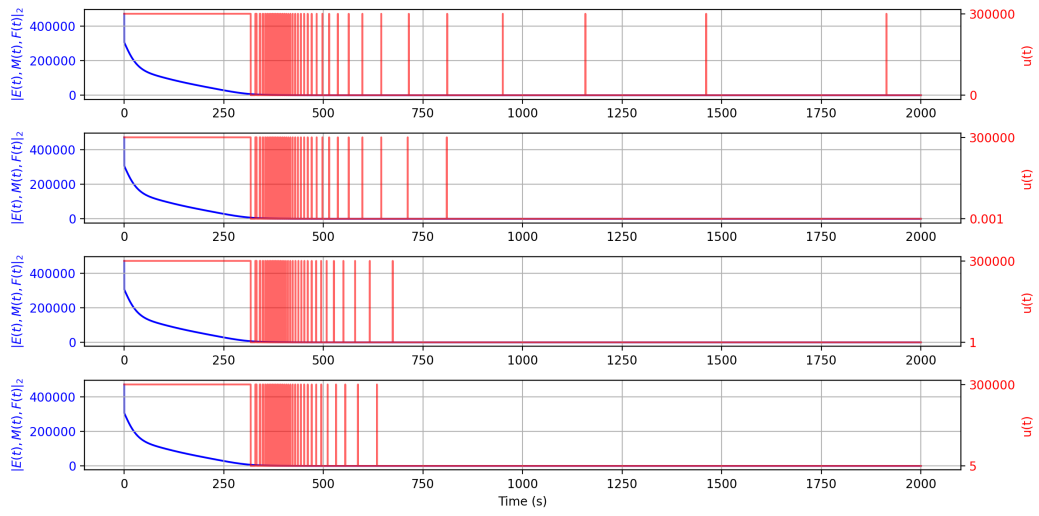
Figure 5: Norm of the states $\|E(t), M(t), F(t)\|_2$ (blue) and control $u(t)$ (red) as a function of time for different values of $u_{\min}$ (0, 0.001, 1, and 5 respectively from top to bottom) and $u_{\max} = 300,000$, over 2000 days and with the same initial condition.