

Step 2 Evaluation Report

CONFIDENTIAL

Call reference	ERC-2023-COG
Activity	HORIZON ERC Grants
Funding scheme	ERC CONSOLIDATOR GRANTS
Panel name	PE6
Proposal No.	101126292
Acronym	Starlux
Applicant Name	Odalric-Ambrym MAILLARD
Title	Statistical Reinforcement Learning to Upscale Experimental Sciences

EVALUATION CRITERIA

Criterion 1 - RESEARCH PROJECT

Ground-breaking nature and potential impact of the research project

To what extent does the proposed research address important challenges?

To what extent are the objectives ambitious and beyond the state of the art (e.g. novel concepts and approaches or development between or across disciplines)?

To what extent is the proposed research high risk-high gain (i.e. if successful the payoffs will be very significant, but there is a high risk that the research project does not entirely fulfil its aims)?

Scientific Approach

To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk-high gain?

To what extent are the proposed research methodology and working arrangements appropriate to achieve the goals of the project?

To what extent does the proposal involve the development of novel methodology?

To what extent are the proposed timescales, resources and PI commitment adequate and properly justified?

Criterion 2 - PRINCIPAL INVESTIGATOR

Intellectual capacity and creativity

The questions below can have one of the following five responses: Exceptional/Excellent/Very Good/Good/Non-competitive

To what extent has the PI demonstrated the ability to conduct ground-breaking research?

To what extent does the PI provide evidence of creative independent thinking?

To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?

PANEL SCORE AND RANKING RANGE

Final panel score: B (meets some but not all elements of the ERC's excellence criterion and will not be funded)	Ranking range*: 66%-100% For your information, only the top 41% of the proposals evaluated in panel PE6 in Step 2 were funded.
--	--

* Ranking range of your proposal out of the proposals evaluated by the panel in Step 2, in percent, from 1% for the highest ranked proposals to 100% for the lowest ranked.

PANEL COMMENT

This evaluation report contains the final recommendations and score awarded by the ERC review panel during the second step of the ERC Consolidator Grant review and the ranking range. The discussion of the panel was conducted within the context of prior reviews submitted by ERC panel members and external referees and the interview with the applicant.

The panel closely examined all the individual review reports and, while not necessarily subscribing to each and every opinion expressed, found that they provide a fair overall assessment. The comments of the individual reviewers are included in this report.

The presentation given by the applicant during the interview and the answers to the questions that were addressed greatly contributed to build the panel's view about the proposal's strengths and weaknesses.

Both the individual reviews and the interview were the basis for the discussion and the final recommendation of the panel.

The panel agreed that the PI has a strong scientific background in areas relevant for the proposed project, especially in multi-armed bandits and reinforcement learning (RL).

The panel noted that the proposed project aims at addressing important objectives related to the practical applicability of RL. However, the panel came to the conclusion that the project is overambitious in that it wants to address 10 key RL challenges. Despite the information provided by the PI during the interview, the panel thought that there is a substantial lack of technical detail regarding the outlined scientific approach and its scalability and practicability in real-world agronomy. The panel was not fully convinced of the coherence of the proposed work packages and thought that the choice of RL is not sufficiently motivated. The panel remained unconvinced about how relevant domain knowledge comes into play and how interaction with domain experts would work out.

The panel therefore does not recommend the proposal for funding.

REVIEWER COMMENTS

The following individual reviews have been carried out independently prior to the panel meeting and do not necessarily reflect the panel's final opinion

Reviewer 1

Research Project
Ground-breaking nature and potential impact of the research project The project addresses very important challenges for reinforcement learning in experimental sciences. It will address challenges related to sample efficiency, scalability, reliability and explainability. The objectives are ambitious and aim to advance significantly the state of the art. The proposed research is high-risk high-gain with significant breakthroughs in reinforcement learning theory, and applications in experimental sciences (agro-ecology and medicare) expected if successful. Scientific Approach The outlined scientific research is mostly feasible, and the PI has the necessary links to experimental sciences and networks. Application-related risks and limitations are less well elaborated in the proposal. This is a minor shortcoming, as the PI has an established network of collaborators involved in the project. The working arrangements are appropriate to achieve the goals of the project. The proposed methodology is adequately elaborated, and relevant. The proposed timescales, resources and PI commitment are adequate and properly justified.

Principal Investigator	
To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Exceptional
To what extent does the PI provide evidence of creative independent thinking?	Exceptional
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Exceptional

Comments (Optional for reviewers)
The PI has an outstanding track record for their seniority and stands apart their peers. They have published several influential papers in highly respected and selective venues. They are capable of inter-disciplinary research as demonstrated from the projects they acquired as a PI, while also achieving theoretical contributions, as demonstrated by numerous awards and highly influential publications. The candidate has provided with solid evidence of their mentoring and leadership qualities.

Reviewer 2

Research Project

Ground-breaking nature and potential impact of the research project

The proposed research addresses several important challenges in the area of stochastic reinforcement learning (sample efficiency, scalability). However, the proposal does not clearly describe the underlying challenges, and the proposed research does not seem to go sufficiently beyond the state of the art. Rather than introducing novel concepts and approaches, it reiterates on and combines known concepts. The description of the objectives and approaches toward achieving them is vague and does not allow the assessment of what exactly the applicant has in mind (examples: the paragraph titled "A deadlock?" does not describe any deadlock at all, sentences like "With a strong emphasis on statistics, the grand challenge of StaRLux is to switch the trajectory of reinforcement learning to assist experimental sciences, rather than games, by developing a new generation of decision companions that are sample efficient, scalable and applicable to real-world agroecology conditions." are extremely general and vague, regarding O2: the whole paragraph is vague, and the link between experiment design and RL is not explained at all, etc.). The proposed research is not high-risk - the applicant does not identify any risks related to the proposal. Section a.5 only mentions one 'external' risk due to competition with other research groups.

Scientific Approach

The scientific approach and the methodology are only vaguely described - there is no way to infer from this description whether they are appropriate, novel and feasible. Regarding novelty, the proposal appears to aim at applying concepts from bandits (the main expertise of the applicant) to reinforcement learning. In addition, the reinforcement learning approach is not described at all. Reinforcement learning (RL) is a broad and well-established field with many results concerning a.o. stochastic systems (as opposed to the applicant's claims that RL has been mainly developed for deterministic systems. The link between RL and experiment design in agroecology - the main application field the project foresees - is not described at all.

The proposal is carelessly written: the text contains typos, broken links, many vague statements and unsupported claims. Recommendation: rewrite the proposal and apply again, if possible.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Very good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)

The PI's scientific expertise is adequate for this proposal. The PI regularly publishes in the main venues of the field (Neurips, ICML, etc.) and definitely belongs among very good researchers, but, considering the impact (Google Scholar) relative to the PI's academic age, not among the top ones.

Reviewer 3

Research Project

Ground-breaking nature and potential impact of the research project

The proposal focuses on the derivation of variants of RL algorithms and theories that address the challenges of real-world experimental sciences. This is an ambitious and worthwhile goal: RL has tremendous potential but, to date, many applications thereof do not address any meaningful societal challenges. The proposal aims to change that by re-envisioning RL from the point-of-view of experimental design and experimental sciences. This is a GREAT idea and proposal!

The proposal is high-risk in the way it aims to validate the proposed ideas: the evaluation/application of the developed methods will be done in a simulation environment which was proposed by the author. It is unclear that this simulation reflects the needs of real-world agronomy or that the good results achieved within would generalize to other domains. Especially given all the emphasis on real-world that is put in the beginning of the document on real-world data, non-determinism, RL beyond games etc. this limitation is slightly puzzling and could pose a substantial risk to the project. On the flip side, any adoption of RL strategies in experimental sciences and agronomy could have substantial economic impacts as well as impact on the well-being of society.

Scientific Approach

The proposed scientific approach is sound and slightly on the ambitious side. The author aims to attack a variety of hallmark problems of RL (explainability, exploration-exploitation, dimensionality) within a single project. While a smaller scaler would have sufficed here, i.e., would be more realistic, I still believe given the authors background and previous work that a substantial dent in these problems can be made. So I am positive about the possibility for substantial outcomes along many of the raised challenges. I would have liked to see more of a "hint" on how the authors aims to realize these goals. What are the new insights or technical hooks that will be used to make headway along these established challenges.

The project will involve 8 persons, which is a reasonable size in my opinion. This will enable the proposer to create a specialized unit working on this. I am enthusiastic about the general approach and I think overall the project, work packages and milestones are well-motivated.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Exceptional
To what extent does the PI provide evidence of creative independent thinking?	Exceptional
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)

Overall, I am very positive about this proposal and the potential impact it can have on society.

My only (constructive) critiques would be the following: a.) I would ask the proposer to be more mindful of existing work when making certain statements. RL has been successfully applied to many stochastic environments, e.g., real-robots playing table-tennis against a human opponent/controlling a helicopter making acrobatics. There are many other examples of stochastic RL policies. I understand what the author is trying to say, but it is important that statements don't become a misrepresentation of the status-quo.

b.) In the beginning the proposal makes an important point that RL needs to be useful in the real-world with all its messiness. However, the proposal is ultimately hinging precariously on the use of the agronomy simulator. There is a disconnect between what is communicated and what ultimately will happen in the project. Frankly, I am worried that all the great arguments about agronomy, users, real-world, etc. will not be followed through - i.e., the project might still end in a purely mathematical endeavor that forgets about its target clientele.

Reviewer 4

Research Project

Ground-breaking nature and potential impact of the research project

This proposal argues for the use and development of new reinforcement learning (RL) algorithms in the context of experimental sciences, and in particular in the field of agroecology.

The use of RL in the general context of experimental sciences (and agroecology) is an important and meaningful objective that could have significant impact on several fields (potential High-impact).

However, there is a disconnect between this stated objective and the proposed research objectives and WPs.

The main blocker to the use of RL in the context of experimental sciences is not the development of new/better algorithms (WP1 and WP2), but rather, the technological transfer of complex and often difficult-to-use algorithms to non-technical users.

This topic is only minimally discussed in the proposal (in WP3).

The objective of WP1 and WP2 are not particularly novel or ambitious, and would not have significant impact on society.

As a result, there are two ways to evaluate this research project:

On one hand, the algorithmic development proposed in WP1 and WP2, is likely to be successful, but the impact is limited (low gain, low risk).

On the other hand, the overall objective is potentially high gain, but extremely likely to fail (very high risk) using the current scientific approach.

Scientific Approach

To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk-high gain?

- The scientific approach to develop new algorithms is feasible. However, this will not result in the desired goal of the project (switch the trajectory of reinforcement learning (RL) to assist experimental sciences), due to the lack of technological transfer.

To what extent are the proposed research methodology and working arrangements appropriate to achieve the goals of the project?

- The proposed research methodology is not appropriate to achieve the goals of the project, due to the lack of technological transfer.

To what extent does the proposal involve the development of novel methodology?

- The proposal in its current state involve the development of new algorithms, but not of a novel methodology. These algorithms will be more or less novel to varying degrees (e.g., there are already many algorithms that can deal with stochasticity)

To what extent are the proposed timescales, resources and PI commitment adequate and properly justified?

- The proposed resources are adequate. However, the planned PhD hires are skewed towards RL theory and Statistical theory. This might not be optimal to achieve the stated goals.

Note: Many citations are missing in Part B.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Non-competitive
To what extent does the PI provide evidence of creative independent thinking?	Good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Non-competitive



Comments (Optional for reviewers)

The PI has a good track record in RL.
However, so do have many other researchers working in the field.
Overall, the PI does not currently have a track records that makes him uniquely suited to tackle the ambitious research project proposed.
In particular, while the PI has a good track record in theory of RL, they do not currently seem to have a strong track record of technological transfer and/or applied research.
A related prior project (SR4SG - Sequential Recommendation For Sustainable Gardening) led by the PI and funded by INRIA for 4 years seems to have led to minor impact (as far as I can find).
As such, the chances of successfully and meaningfully engage with the experimental science communities mentioned in the proposal -- and achieve real-world impact -- are low.

Reviewer 5

Research Project

Ground-breaking nature and potential impact of the research project

The proposed research envisions addressing decision making challenges for agroecology. In particular, the proposed project with a focus on the specific challenges of agroecology is concerned with the issues that hinder the application of reinforcement learning methods to such a domain, that is deemed crucial for modern societies. The applicant outlines some key challenges for decision making systems, i.e., Reinforcement Learning (RL) methods, like their ability to address stochasticity of real-world problems, sample efficiency, delayed rewards, exploration and exploitation, as well as explainability of solutions to practitioners. The proposed approach considers studying and integrating ideas from experimental sciences and statistics to reinforcement learning approaches to address the key challenges mentioned in the proposal.

The objectives of the proposal are three and they encompass the proposal into creating three distinct work packages to reach those objectives. The objectives detail some novel concepts, namely the idea of group RL, though not well-explained thereafter. I would argue that the other objectives are either common issues in RL research, or they refer to the specific engineering challenges of the application to Agroecology. Overall, I would assess that the objectives are somewhat ambitious, but it is not clear how the proposal will achieve state-of-the-art results.

The research methodology as presented was quite convoluted. I was not able to assess what are the key methods that are going to be used to achieve, for example, a novel paradigm as groupRL. I found that the description of the methods is quite vague. The applicant discusses how they want to go beyond the classical setup of Markov Decision Processes, and to study restless systems, proposing looking into Piecewise Deterministic Markov Processes, however, it was not clear how this would be achieved, and what methodologies are going to be derived. Additionally, the applicant considers using group-sequential experimental design methods in multi-armed bandit theory, however, it was not detailed how and why this is necessary and which are the key steps that are needed for this to be achieved, while there is no discussion on the broader benefits of such an approach. The last methodological aspect refers to explainability and the development of RL companions for Agroecology. I found that these concepts are quite disconnected and very briefly explained. While the applicant stated that the enlisted methodological concepts need to be connected, there was no clear description of the necessary steps to bring together the different methodologies.

An important note is that it was not clear for me was if the projects targets deep reinforcement learning methods, and in particular what types of methods, and was not even clear why the proposed domain would require deep solutions (as counter-reference to application of deepRL to games, as mentioned in the proposal), e.g., what are the types of data, their dimensionality, the hidden dynamics etc. In my view, the problem could be handled as a multi-arm bandit problem and I was not convinced about the need of RL methods in the domain of Agroecology.

Overall, while the proposed research contains high-risk elements, due to the lack of clear and convincing descriptions of the course of actions on how new knowledge will be developed and the lack of structure of the methodology, I assess the risks of not fulfilling key aspects of the proposed work higher than the possible gains.

Scientific Approach

The proposed scientific approach aims mainly to leverage the applicant's expertise in bandit theory for developing RL algorithms for experimental sciences. The scientific approach is divided into three sub-topics that are reflecting the three main objectives of the project, namely, sample efficiency, scalability and applicability of RL algorithms for experimental sciences. Unfortunately, the description of the proposed technical contributions are vague, convoluted and not clearly connected for achieving some groundbreaking result. It is not clear what are the issues that hinder modern deep RL methods to be applied to challenging problems. While the applicant fixates on games, there are several applications of RL that are promising, e.g., in Robotics [1], physics [2], finance [3]. It has never been explained, nor became clear throughout the project what is the scalability issues of the tackled domain of agroecology, e.g., what is the dimensionality of the problem, what are the dynamics, and why classical AI methods cannot be applicable.

I found the idea of reformulating the notion of MDPs into a PD-MP a very interesting one, but this has not been given enough space in the proposal to be explained and described concretely. While I understand that the applicant comes from a bandit theory background, throughout the method description it is not clear if the applicant has a concrete plan

for transferring the ideas from bandit theory to sequential decision making problems. Moreover, the group sequential setting that seems to be a novelty of the proposed research plan is not explained, making it very difficult to assess what the proposed method of groupRL would achieve.

The third workpackage, on the other hand, mixes up ideas about reusability and explainability without having any connection to the other approaches and tasks in the proposal. In my opinion the problems of reliability of solutions and its evaluation would be a challenge on its own. The explainability is a big issue that is not well connected to the methods developed. Lastly the companions seem an engineering solution that does not really provide a clear protocol and benchmarks for the application to agroecology.

On top of all, I was not able to assess whether the timeline and the distribution of work for the various PhD students and PostDocs was realistic, as the researchers' workload is spread across different tasks with different objectives, and is not clear if there are clear paths for the production of meaningful and novel research towards acquiring a doctorate degree.

The risk mitigation is also not sufficiently explained. In my view, the risks in this project come mainly from the fact that the applicant has more experience in bandit theory and statistics, and less experience in reinforcement learning. However, the listed collaborator in different tasks of the project could potentially help mitigate the aforementioned risk.

Finally, the project proposal was not meticulously curated. The absence of proper references made it impossible to track the referenced related works. The absence of diagrams made it also difficult to understand the various interactions and connections between workpackages and tasks. While I believe the project has interesting ideas, those have not been clearly communicated, creating a confusion about how the listed objectives are realistically unique for experimental sciences and how the proposed solutions would lead to novel scientific methodologies, even beyond the application to agroecology.

- [1] Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P. and Levine, S., 2021. How to train your robot with deep reinforcement learning: lessons we have learned. The International Journal of Robotics Research, 40(4-5), pp.698-721.
[2] Degraeve, J., Felici, F., Buchli, J., Neunert, M., Tracey, B., Carpanese, F., Ewalds, T., Hafner, R., Abdolmaleki, A., de Las Casas, D. and Donner, C., 2022. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature, 602(7897), pp.414-419.
[3] Charpentier, A., Elie, R. and Remlinger, C., 2021. Reinforcement learning in economics and finance. Computational Economics, pp.1-38.

Principal Investigator	
To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)
The PI has definitely demonstrated the capacity of performing ground breaking research, as this is evident by the list of publications in top-tier conference in Machine Learning, and his ability to supervise Ph.D. students and PostDocs. He received an award for his PhD dissertation, and a method has been named after him, showcasing the impact of his contributions. While a list of invited talks was listed, it seems that the last two years the applicant was not that active. This makes me wonder if it is related to the impact of COVID19, or whether the impact of the more recent scientific works are perceived as less important by the community. An absence of a keynote speech in some important venue is also a minus, but given that the ML field has overly grown, this can also be understandable. In terms of acquisition of third party funding, I believe the applicant has shown a good record. Finally, the applicant has consistently been

serving the community through various editorial positions. I am sure that the PI has a very good profile towards acquiring an ERC grant

Reviewer 6

Research Project

Ground-breaking nature and potential impact of the research project

The aim of this project is to develop reinforcement learning methodology in order to assist the experimental sciences and bridge the simulation-reality gap. Overall, I find the scope of the project from innovations in RL all the way to solving society's most pressing problems far too broad to be convincing. Beyond agro-ecology, it also promises potential for industry.

Scientific Approach

While I do believe that the RL innovations are appropriate, I am skeptical to what extent it will be possible to bridge the gap between simulation and realworld agro-ecology. While improving sample efficiency of RL systems is definitely useful, I am not convinced that this and the other work packages will lead to dramatic improvements in agro-ecology, industry and beyond. In that sense, I find the proposal is somewhat over-promising.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Excellent
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Very good

Comments (Optional for reviewers)

The PI has a strong track record of publications, some of which appeared in top venues like NeurIPS, ICML and Annals of Statistics.

Reviewer 7

Research Project

Ground-breaking nature and potential impact of the research project

Reinforcement Learning (RL) is one of the most promising ways to endow Artificial Intelligence (AI) agents with sophisticated decision-making skills, which achieved groundbreaking results after the introduction of deep learning methodologies. Despite promising empirical evidence, there are still significant challenges hindering the successful application of RL in real-world problems. This proposal addresses the challenge of enlarging the spectrum of applications of RL in meaningful real-world problems, with the declared purpose of assisting humans in experimental sciences. The proposal puts particular emphasis on the fact that RL has been successful only in "games", which to my understanding, is a *reductio ad unum* to imply "applications performed in simulation and with small or no impact on society". If I interpreted this correctly, I think that this is a strong statement that is not fully reflected in reality, where there are already successful applications of RL in robotics, finance, healthcare, and so on. Nevertheless, I understand the provocative intent of this statement to motivate the project.

The proposal focuses on agroecology as a representative example of complex and impactful real-world applications. The application of AI in agroecology is a very recent trend, which holds the promise of optimizing a crucial sector with high economic and societal impact. Moreover, there are no successful applications of AI in agroecology, which advocates for further research in this direction. I consider this choice very appropriate as a meaningful testbed for developing methodological advances in RL; however, I find somewhat problematic that the discussed methodological advances to RL algorithms are quite general-purpose. The main challenges described in the proposal are to make RL more sample-efficient, able to scale, and more applicable. These challenges have been pursued by researchers for many decades, and it is not immediately clear how this proposal is novel in this sense. While I agree that these properties are desirable in agroecology, which is a novel application of RL, I easily recognize them as desirable properties for all RL algorithms in the majority of applications. The overall impression is that the proposal aims at tackling the usual challenges of RL, i.e., sample-efficiency and scalability, which makes the novelty and focus quite disputable. Obviously, it would be great to substantially improve sample-efficiency and scalability of RL algorithms, but my doubt is whether all of these challenges should be tackled at once. The risk is to lose focus on a main goal and to provide incremental advances in different directions, which can certainly lead to interesting scientific contributions, but probably to poorly or moderately significant practical consequences.

An important point is that the proposal seems to describe RL as quite limited at the moment to tackle real-world problems. One could argue that, if RL is so limited, then probably it shouldn't be considered at all for tackling applications of this kind, or at least that RL has so many limits that they cannot be addressed in one project proposal. While I appreciate the ambitious goal of tackling many important challenges aiming to enable the use of RL in large-scale experimental campaigns, which remarkably suits the high-risk high-gain philosophy of the ERC, I think that the proposal could have been more optimistic on the state-of-the-art of RL to give the idea that this is a viable and worthy solution. I think that the high-gains are well motivated, with a clear description of the impact that the proposed advances could have on society. However, the high-risk is not well presented. The discussed risks are mostly about individual work packages, which I do not think should be considered high-risk. As stated before, I think that the highest risk of this project is to provide incremental advances in different direction, with a lack of focus on a central idea (more will be described later). The lack of discussion about this critical point, together with the absence of figures, schemes, tables, makes it very hard to understand what unifies the different approaches and what the main focus of the project is. Given its importance and high-risk, I consider this the weakest aspect of this proposal.

Scientific Approach

The proposal has the ambitious goal of enabling the use of Reinforcement Learning (RL) in experimental sciences. RL is arguably far from being broadly applicable in relevant real-world problems, as also stated in the proposal several times; thus, the goal of the project is definitely ambitious, suiting the philosophy of high-risk high-gain. I am certain that a successful outcome of this project could have a significant impact on society, and in particular on the important sector of agroecology on which this proposal is focused.

My concerns for this proposal are about its main focus and the risks related to it. The proposal often stresses the fact that RL has been unsuccessful in scaling to large real-world experiments until now, not only for practical reasons but also for methodological ones, e.g., the suitability of Markov Decision Processes to model problems and the inability to

handle highly stochastic behavior. To overcome these issues, the proposal aims to introduce several advances with a rigorous statistical perspective and inspired by Multi-Armed Bandits (MABs) theory. Although the advances are interesting per se, I think the proposal lacks cohesion among the proposed methods. For example, I think that the investigation of Piecewise Deterministic Markov Processes (PD-MP) to model real-world problems is highly interesting, and I would have liked to see a broader analysis of this in the proposal, in particular highlighting the connection with other tasks in other work packages. Similarly, the proposal discusses the investigation of how to improve sample-efficiency by narrowing the gap between MABs and RL; this is obviously interesting, but it is not clear how this relates to the study of PD-MP or others, e.g., explainability (WP3). The absence of figures, schemes, tables, and the unfortunate presence of several broken citations, further hinders the emergence of a clear main goal and cohesive structure for this proposal.

The methodologies described in the proposal are sufficiently novel and potentially impactful. Most of them are appropriately described; however, sometimes, only high-level ideas are presented without a concrete description of how methods will be implemented. This is not a problem, but it sometimes gives the feeling that some research directions are still unclear. Regarding the application to agroecology, I find it a very good testbed for the scope of the project. Moreover, the fact that the PI already has experience in the field is a big plus. The timescale, resources, and PI commitment are adequate. I am slightly concerned about the involvement of the postdocs, who seem a little too involved in their own research tasks and not enough in Ph.D. student supervision.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)

The PI has certainly provided evidence of scientific expertise and creative independent thinking. The previous and ongoing experience in the field of agroecology is also a big plus in providing evidence that the PI could successfully execute the project. The expertise and well-recognized works of the PI in Multi-Armed Bandits (MABs) and Reinforcement Learning (RL), often under a rigorous statistical lens, well suit the ambitious goals of the proposals. I am confident that the PI has the required technical and organizational skills, together with a vast network of connections with renowned researchers and institutes, to accomplish the tasks of the project successfully. The works of the PI on MABs and RL are particularly intriguing since it is evident that the PI strives to endow RL methods with more theoretical guarantees, as commonly done in MABs literature. This is a very important goal for RL, that could potentially have a significant impact both in the research community and in practical applications, e.g., industry, laboratories. Given this important experience and ambition of the PI, I totally believe that the PI is well-suited to accomplish the individual tasks in the work plan of the project.

Reviewer 8

Research Project

Ground-breaking nature and potential impact of the research project

The proposal aims to develop ideas of reinforced learning (RL) to be applicable in practical domains like agriculture and medicine. The current level of RL does not allow it to handle such types of tasks. The proposal is a high-risk high-gain. The proposal would strongly gain from a final reading - some of the references did not compile.

Scientific Approach

The challenges are clearly defined. The proposal will help to advance the area of RL. However, it is not very clear why RL will be the right tool in experimental science. The WP3 targeting developing agro-recommendation platform is exciting. Resorting to simulators instead of real data is understandable but still a bit disappointing.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)

PI is a renowned expert in the multi-armed bandits and RL: He is responsible for several fundamental results in this area. He has a large number of papers in the top ML conferences.

Reviewer 9

Research Project

Ground-breaking nature and potential impact of the research project

The idea of focusing on the application of RL to experimental sciences is very innovative and intriguing. The PI has already taken some steps in this direction, but the field is still largely unexplored. The project has a major potential for impact, both on the theoretical machine learning side, but also on the application. The application of agroecology is very relevant and comes with its own set of unique challenges, that will be tackled from a theoretical point of view. The societal impact remains vague and generic. The high-level ideas are sound and innovative, bringing in ideas from neighboring fields.

The objectives are ambitious, addressing numerous challenging open research questions in RL in general, and in RL for experimental sciences in particular. The overall ideas are novel and sound.

If all the challenges addressed in this proposal can be successfully solved, this will be a tremendous step forward. But I indeed see a very high risk that there will not be sufficient time to address all challenges (that's fine) and/or that the individual solutions will be "piecemeal" - addressing individual challenges in isolation and not being compatible with each other and/or not being suitable for practical application.

Scientific Approach

The submission seems to have been rushed, with quite a few typos and more importantly a huge number of references in B2 gone missing, making it almost impossible to judge the novelty, completeness of overview of the state of the art, etc.

The connection to practical applications in agroecology seems to largely rely on other projects, that are either already running or are to be setup. Relying on projects that are yet to be granted seems to be a major risk. While the proposal extensively mentions agroecology as a use-case and mentions some specific challenges, it still remains very vague how the connections will look like, how close the targeted simulations will be to something practically useful, and how much the practical experience will flow back into the theoretical developments, etc. Overall there seems to be a huge step/gap between the purely theoretical and generic WPs 1 and 2 to the application WP 3.

Heavily relying on outside expertise not only for the application, but also for some of the theoretical work seems against the spirit of a personal grant.

The risks and mitigation strategies overall are rather vague and generic, I would have liked to see a more project specific analysis.

The project is written in a way that nicely points out the challenges and ideas on a very high level, and on a very detailed level. But the middle level of detail seems to be missing, things like what kind of RL setting (model-free, model-based, model-learning) is targeted, the sim2real gap is described in a different manner than I am used to, it remains unclear why domain randomization strategies to account for uncertainty in dynamics is not applicable here (in contrast to sim2real transfer strategies in robot RL), etc.

The explainability part of the research does not seem to be very user focused, but rather a first step for RL experts.

The proposal is focused on the development of novel methodology and theory in WP1 and WP2. The topics to be addressed are very interesting and relevant, WHAT is to be achieved and investigated is nicely outlined. However, for me HOW this is going to be achieved remains very vague (beyond "extend previous results" and "incorporate things from neighboring fields") without providing any details on what the innovative ideas are to achieve this.

Overall the tasks seem extremely ambitious, with on average enough material for 3 or 4 major publications with an equivalent 1.5 years full-time work budget. I am not convinced by the comment in B1 "while apparently disconnected ... these three WPs collectively form the coherent back-bone". How the project is setup, there is no way to show that the individual parts are coherent, i.e., nicely play together rather than addressing individual sub-problems and being incompatible, or make a significant step forward in addressing the application problem. While each topic is interesting, has scientific merit, and is related to a challenge in the application domain, I did not see sufficiently how the various theoretical parts come together, let alone how practically applicable the individual (or even better collective) methods are to agroecology. While the aim to both make fundamental theoretical breakthroughs and develop methods that are practically applicable at the same time is very laudable, and on both ends are very interesting and challenging, also here I am missing the "middle".

The scientific workshops requested in the budget and their contribution to the project remain vague. I liked the idea of

having a series of RL challenges. I have doubts whether a 70% time commitment of the PI is realistic given obligations to other projects, teaching, service, etc. The other requested personnel and resources seem well justified.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Excellent
To what extent does the PI provide evidence of creative independent thinking?	Exceptional
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Exceptional

Comments (Optional for reviewers)

The PI has an excellent track record in the area of the proposal. He has more than ample experience in advising PhD students and postdocs. He also is very active in the academic community (reviewing, area chair, editor, etc.). He is currently involved in a good number of projects, that have connections (but no substantial overlap) with both the theoretic (RL and bandits) and application side (agroecology) of this proposal. His early achievement track record is excellent, with numerous high impact papers at prestigious venues and several awards. He clearly demonstrated being able to direct and execute research independently, coming up with creative and innovative ideas and directions. Finally, he is also actively involved in teaching at the MSc and PhD level.

Overall, I have no doubts that the PI is the perfect match to the proposal, and uniquely qualified to pull off this project.

Reviewer 10

Research Project

Ground-breaking nature and potential impact of the research project

Import: The project does seem to address important challenges, both in theory (sample-efficiency, reproducibility) and practice (applications in agronomy)

Ambition: Moderately high. The problems seem difficult, but well-documented and under investigation by many research groups (particularly the theory/algorithms problems)

Risk/Gain: High. The problems seem very difficult and perhaps not even "universally" solvable. Focusing on particular application domains hedges some risk and provides a suitable path forward

Scientific Approach

Feasibility: The approach of spanning theory and practice with contributions to both seems feasible, albeit ambitious

Methodology: sound

Novelty: Moderate-to-high. As discussed above, a number of ideas discussed in the proposal have been studied in the literature (batching, non-stationarity, etc.) but expanding on these ideas may require novelty. Moreover, some of the other ideas seem more novel (deciding when to intervene, for example)

Resources: Adequate and justified

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Exceptional
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Exceptional

Comments (Optional for reviewers)

The PI is a well-regarded expert in the theoretical aspects of reinforcement learning and bandits. I am less familiar with their applied work, so it is hard for me to comment on the impact here.

Reviewer 11

Research Project

Ground-breaking nature and potential impact of the research project

#Ground-breaking nature and potential impact of the research project

To what extent does the proposed research address important challenges?

StaRLux aims to develop novel, mathematically sound reinforcement learning methods tailored for experimental science domains, such as medicine and agronomy. These domains pose several challenges, including stochastic environments and dynamics, complex feature interactions, delayed feedback, and collaborations with individuals not familiar with machine learning. Each of these challenges is important and well-motivated in the proposal.

To what extent are the objectives ambitious and beyond the state of the art (e.g. novel concepts and approaches or development between or across disciplines)?

The three stated objectives, namely O1: Sample-efficient RL for experimental environments, O2: Scalable experimentation with group-RL, and O3: Applicable RL companions in agroecology, along with their respective subgoals, are indeed ambitious. To achieve these objectives, StaRLux will need to create innovative concepts and approaches. It's worth noting that these concepts could have a significant impact extending beyond the project's immediate scope.

To what extent is the proposed research high risk-high gain (i.e. if successful the payoffs will be very significant, but there is a high risk that the research project does not entirely fulfil its aims)?

I strongly agree that developing new reinforcement learning methods capable of performing effectively in stochastic environments is highly valuable. Additionally, the ability to learn policies tailored to diverse settings and individual preferences has the potential to make a substantial impact in the reinforcement learning field and beyond. Moreover, with a specific focus on the critical and socially relevant domain of agronomy, I believe that all the mentioned objectives could indeed have significant positive effects on our society.

However, I do have concerns that there is a substantial risk of not achieving the stated objectives. This concern arises because several crucial aspects of how these challenges will be addressed are not clearly defined in the proposal.

These concerns include:

- It is unclear to me to what extent the discussed reinforcement learning theory, based on bandit scenarios, is applicable to the complex experimental domains with stochastic environments.
- The claim that "... deep RL approaches are largely unable to correctly handle even small stochastic systems..." is a fundamental motivation in this proposal. However, no citations or further elaborations are provided.
- The argument that (Deep)RL is predominantly used for games is a recurring theme throughout the proposal, but it appears to overlook current work on (offline) reinforcement learning that specifically aims to solve more realistic, human-centered environments, such as Meta-World, The Franka Kitchen Benchmark, CALVIN, D4RL, and Maniskills. While these environments may not be as stochastic as those described in the proposal, they do require complex, multimodal, and/or hierarchical behavior, demonstrating the development of reinforcement learning methods beyond the scope of games.
- The description of how structured interactions will be learned and how this would positively impact sample efficiency is not clearly articulated.
- The statement "... Success in the StaRLux project is expected to significantly reduce the number of interactions required by learning agents..." lacks sufficient clarification and explanation.
- Although the project claims to address the Sim-To-Real gap, the proposal does not provide elaboration on this statement. The mere existence of a better simulator for stochastic environments does not automatically eliminate or mitigate the challenges of real-world deployment of policies learned in simulation.

Scientific Approach

Scientific Approach

To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk-high gain?

My primary concern in this regard pertains to the transfer of reinforcement learning theory developed for bandit scenarios to the intricate, stochastic settings outlined in the proposal. While the proposal frequently mentions its intention to create novel methods for experimental domains like agroecology, drawing from insights in bandit scenarios, it lacks adequate elaboration or discussion on how these complex, stochastic domains challenge the assumptions made in the bandit setting and how these challenges will be addressed or resolved.

Additionally, I have reservations about the proposed approach to imbue the novel reinforcement learning methods with explainability. It appears that this aspect may not have been thoroughly considered. Simply enhancing the feedforward/predictor model, even if it incorporates some level of counterfactuals, does not necessarily provide explainability. Attempting to infer the agent's intention based on potential future states can introduce misinterpretation and give rise to the "illusion vs. interpretability" dilemma.

Otherwise, I found the description of the challenges, current limitations, and the suggested approaches to address them to be well articulated.

To what extent are the proposed research methodology and working arrangements appropriate to achieve the goals of the project?

Aside from the reservations mentioned earlier, I believe the methodology to be appropriate.

To what extent does the proposal involve the development of novel methodology?

Every aspect of the proposal necessitates the development of new approaches and methods. This includes not only the core research endeavors but also the crucial task of verifying and testing the statistical significance of the achieved results.

To what extent are the proposed timescales, resources and PI commitment adequate and properly justified?

While I believe the proposed timeline could be feasible, I have reservations about whether the objectives outlined for the PhD candidates might be overly ambitious. On a positive note, I appreciate that the proposal incorporates research engineers who are focused on the development and upkeep of the simulation and software infrastructure.

Finally I would like to mention that it did feel a bit disheartening to see the apparently rushed state of the final proposal (including missing MANY citations and spelling/grammar mistakes).

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Very good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Very good

Comments (Optional for reviewers)

No comments received

Reviewer 12

Research Project

Ground-breaking nature and potential impact of the research project

The proposal addresses a few important challenges in reinforcement learning and its practical applications in experimental sciences, in particular, in agroecology. Specifically, the main challenges to be addressed include sample efficiency in RL for experimental environments, scalable experimentation through group RL, and its applications in agroecology.

On one hand, these are important challenges in RL research. On the other hand, these challenges are known in the research community.

To what extent are the objectives ambitious and beyond the state of the art (e.g. novel concepts and approaches or development between or across disciplines)?

The objectives are ambitious – it targets the application of advanced reinforcement learning algorithms in experimental sciences where data samples are expensive and time-consuming to obtain and the environment can be highly dynamic with uncertainties, as well as different influencing factors and objectives.

At the same time, it is less clear to the degree of the proposed research goes beyond the state-of-the-art. A key difficulty in reviewing how much the proposed research advances the field is that each research task is described rather briefly. In Part B2, the detailed work-plan has two and half pages, outlining the ideas for ten separate tasks (in three work packages). Therefore, the discussion for each task is rather limited. It mostly constitutes of high-level ideas on how one can leverage existing work to further develop a research direction. Many of such approaches are known.

To what extent is the proposed research high risk-high gain (i.e. if successful the payoffs will be very significant, but there is a high risk that the research project does not entirely fulfil its aims)?

Comments: *

The proposed research, if successful, is definitely high gain in the sense that the payoffs will be significant. The high-level target of using reinforcement learning for experimental science is highly desirable.

At the same time, for the lack of technical details, it is not clear to the reviewer that the goal is achievable through the proposed tasks.

Scientific Approach

To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk-high gain?

The proposal lacks technical details for the reviewer to confidently answer this question. I think the proposed research tasks are likely to be viable to a certain degree. For example, for T3b, the author proposes to “developing methods for predicting effects or counter-effects of actions in stochastic dynamical systems”. Yes, this is a viable research direction, but also a known direction. The proposal does not provide details on how the PI’s approach advances existing approaches. More importantly, it is not clear how much these approaches combined can lead to the desirable outcome of RL in experimental science.

To what extent are the proposed research methodology and working arrangements appropriate to achieve the goals of the project?

Again, the proposal lacks details for the reviewer to confidently answer this question. The proposal reads like ten loosely related research tasks. While each of them is likely to be a viable research topic, it is unclear how much advances one can expect from each of them and how putting together they can achieve the goals of the project.

Another challenge is that the proposal is abstract. While the PI's target is to utilize RL in experimental science, no details discussions are there for the agroecology problem. There are also no mentions of the scale of improvement needed to make the goal feasible. For example, a RL game often requires millions if not billions of steps for a RL policy to converge. This is clearly not feasible in experimental environment. What is the order of magnitude of improvement needed and what can each approach in the proposal provide for the goal to be viable? My own evaluation based on reading the proposal is that we are far from what is needed.

To what extent does the proposal involve the development of novel methodology?

The proposal contains ten tasks loosely related, covering a very broad range of RL. I am familiar with some topics, and less familiar with others. Some of the ideas or directions are known to the field (for example, the extension of bandits to RL in stochastic environments). It is possible that the PI has novelty in the intended approaches. But it was not clear from reading the proposal. There are other ideas that I have not heard of, possibly due to my limited knowledge in those components.

To what extent are the proposed timescales, resources and PI commitment adequate and properly justified?

The proposal addresses the proposed timescales, resources and PI commitment clearly.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Excellent
To what extent does the PI provide evidence of creative independent thinking?	Very good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Very good

Comments (Optional for reviewers)

The Principal Investigator (PI) of this proposal has demonstrated a strong and highly relevant background in the field, making them well-suited to lead the proposed project. In particular, the PI has strong background in both theoretical RL algorithms, as well as building a RL simulator, gym-DSSAT, for crop management. Their expertise and strong prior work align with the tasks proposed here.

My main reason of not giving a stronger rating is that I am not certain the proposed tasks will lead to the desired goals. In addition, the tasks are only loosely related, which could be a challenge for one PI to manage.

I hesitated between Recommended for funding and Not recommended for funding and chose the former as I think some good research results could come out of this project.

Reviewer 13

Research Project

Ground-breaking nature and potential impact of the research project

- The proposal addresses using RL techniques for agronomy which is a very important challenge for future food safety given the climate change conditions. The problem is also very challenging from an algorithmic ML perspective, such as how to deal with the high uncertainty and stochasticity in the environments, how to learn progressively in shifting environments, how to do a coordinated exploration for delayed reward environments etc., how to learn when to intervene and when to observe. These are hugely important aspects of RL which have not been fully addressed and could allow to scale the use of RL for experimental sciences.
- The proposal addresses these challenges from a quite theoretical perspective by developing new theory for bandit and MDP based algorithms. These goals are ambitious for theoretical RL research, but many recent RL algorithmic developments that could also benefit the agronomy challenges (such as offline RL, hierarchical RL, learning state representations in RL, etc..) were not discussed in the proposal. For some of the proposed innovations, it is also unclear how this will benefit the proposed agronomy challenges. For example, I found it hard to see where we can directly use bandits in the proposed. Yet, a major part of the proposal deals with working on RL competitions and benchmarking environments for agronomy which I think is very interesting for the community and a highly ambitious effort. Yet, a shortcoming here is that the proposal mainly addresses simplified scenarios and there is no discussion how to upscale RL for real environments where large scale data collection would be needed. I would have loved to see this discussion more at least in the impact section as data-collection is key to the success of deep RL methods.
- The payoffs of the project will be significant as lifting RL to experimental sciences could impact many industries, not just agronomy. There is also a high risk as the proposed simulated environments might not capture the reality well enough and the proposed theoretical contributions might not address the challenges of upscaled real environments or even simulated environments.
- I found the application very appealing and interesting.

Scientific Approach

- The scientific approach tackles many theoretical RL problems such as proving optimality in non-ergodic MDPs. The sketched scientific approach is feasible and the PI has the correct expertise for doing it. I am however not fully convinced whether the scientific approach, i.e. solving theoretical RL problems, will solve the main challenges stated by the agronomics environments. Yet, the theoretical research posed by the proposal is excellent and the interdisciplinary nature (e.g. combining statistical learning theory with experimental design) of the scientific approach is very appealing.
- The proposed research in theoretical RL is excellent, but I had the feeling it is missing practical considerations for using RL for agronomy. The proposed research sometimes seemed a bit disconnected to the given challenges and environments, for example for the given farmGym environment, which is a very interesting environment, practical considerations such as what is a good policy or value function architecture (use of RNNs, transformers, CNNs, GNNs, etc...?). Should we use onpolicy, offpolicy or model-based methods etc..? Here, a task on benchmarking current DRL algorithms with different policy architectures for different complexities of e.g. the farmGym environment would have been very helpful.
- Other aspects that are not considered are Meta-Reinforcement learning (which is crucial due to the different properties of each farm), offline RL with data from current agroecological strategies or hierarchical RL and investigating techniques for closing the sim-2-real gap. The proposal would have been stronger if these more practical aspects of RL would have been considered as well.
- The commitment of the PI, timescales and resources do look adequate and are well justified. The proposed research staff does look reasonable for the given project, yet, I have doubts regarding the engineers as I am not sure whether an engineer without RL knowledge can really provide good RL benchmark environments. I would rather see such a work better suited at the level of the post-docs. Also, good engineers are very hard to find for an academic salary. Hence, I would rather reduce the number of PMs for engineers and increase the PMs for PostDocs.

Minor comment:

- Many of the references in the proposal were broken (marked as "?"), which made it also harder to evaluate how much the proposal is going beyond SOTA.

Principal Investigator	
To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Excellent
To what extent does the PI provide evidence of creative independent thinking?	Excellent
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)
The PI has a very good track record in theoretical RL research and regularly publishes in top ML venues (NeurIPS, ICML). His citations are also growing steadily showing his impact in the field. The number of last authors publications, showing independent research is good, but could be increased. The PI has a good amount of experience in funding but has never led or was part of a bigger project. The number of supervised PhD thesis is also very good given the academic age.

Reviewer 14

Research Project

Ground-breaking nature and potential impact of the research project

The project is aimed to shed light on reinforcement learning for experimental science. The current main research stream is often directed to unrealistic benchmark tasks, and I sympathize with the intention very much. While I can find several recent studies addressing this research direction, the importance of RL for experimental science is still underestimated, and I think it's important to invest more research efforts into it.

The objective of this proposal is to change the research trend in the RL research community, and I think it is very ambitious. If the research trend in the RL research community is changed by this project, the impact of this research will be significant.

On the other hand, even if the results may not be so impactful, I think that the group will make a certain advances in the field, considering the capability of the PI and his group. The problems of RL described in the proposal will lead to new problem formulations, and there should be some theoretical advances in the field of RL. In this sense, I do not see a particularly high risk.

Scientific Approach

In each package, I see interesting tasks. In WP1, "when to intervene, when to observe" is an interesting question, and important in real-world applications. Similarly, the idea of incorporating expert knowledge for prediction is interesting.

In WP2, "batch coordinated exploration and joint risk" also seems unique and realistic, especially in agriculture. I'm not aware of studies in batch coordinated exploration, and it will lead to new problem formulations.

In WP3, the idea of extending finite-time guarantees to non-parametric distribution is unique and novel as well. I haven't seen an open-source RL platform that incorporates such features.

On the other hand, I see some unclear points in the plan. In WP1, it is not clear how a "large" stochastic system is going to be addressed. Also, while T1a is focused on stochastic environments, it is not clear how novel algorithms can be derived.

Although the evaluation of deep RL is often performed in near-deterministic tasks, the problem formulation of deep RL typically incorporates the stochasticity of the dynamics and agents.

In addition, it is often not clear whether the term "RL" includes both deep RL and (linear) bandit, or either of them. If it is planned to develop deep RL methods, I am not sure whether the PI and his group have sufficient prior experience with it.

Regarding WP3, how to enhance the explainability is not clear. Although I understand that the prediction of effects and counter-effects with uncertainty will be used, the novelty of such methods will be limited. I expect to have a clearer or another plan to enhance the explainability.

However, the overall plan looks reasonable and feasible, and I don't mean to criticize the overall plan.

Although there are some unclear points in the proposed plan, the overall proposal seems feasible and reasonable. The proposed research plan is mostly concrete, and we can expect a certain achievement within the research period.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Excellent
To what extent does the PI provide evidence of creative independent thinking?	Very good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)



PI and his colleagues have published many papers in top-tier machine learning conferences, and there is no doubt regarding the capability to conduct theoretical research. On the other hand, their work is mainly focused on theory, and I do not see strength in deep RL, which would be necessary to develop a scalable system. In addition, the perspective of researchers in experimental science may need to be complemented by external collaborators.

Reviewer 15

Research Project

Ground-breaking nature and potential impact of the research project

The proposal is certainly ambitious and addresses important challenges both in terms of developing improved theory and algorithms for reinforcement learning, and in applying these techniques alongside practitioners in agriculture (and to a lesser extent medicine). Deep RL methods have demonstrated great capacity for solving complex decision-making problems, but despite the rapid progress made in recent years the field is still relatively immature, and the most obvious successes so far have arisen within highly controlled environments such as games. As such, while there is potential for these methods to be very useful in combating increasingly important global societal issues such as food security and disease prevention and control, there is a pressing need for improvement to be made to the foundational algorithms which underpin RL.

The proposal has identified a number of important issues which may restrict the application of current RL approaches in the context of the problems which arise in complex, real-world domains such as agronomy (for example, approving the ability to handle stochasticity in these environments, improving sample efficiency, incorporating methods to efficiently address the high cost of observing some state variables, identifying the optimal time to take intervening action, leveraging group-RL methods to support sharing of learning across different environments and varying user objectives). The PI proposes to address these problems primarily via the incorporation and extension of ideas from the fields of statistics and multi-armed bandits. I believe there is novelty in both the nature of some of the problems being addressed, and in the methods being proposed to address them.

Importantly the proposal also addresses the need to develop strong engagement with practitioners, and identifies both the need for suitable capabilities within the RL systems themselves (such as the need for greatly improved explainability, but also the need for suitable software tools and frameworks to facilitate knowledge-sharing, and to ease the burden on practitioners of adopting these techniques.

The potential payoffs of this project are high, in two respects:

- The range of possible application areas for Deep RL is very broad - as well as the issues of agriculture and health focused on within this proposal, the possibility exists of improving decision-making across a host of other application areas, with the potential for huge economic, environmental and societal benefits. Should this project be successful in improving the ability of RL methods to cope with the issues listed earlier, this would be of undoubted widespread benefit.
- The specific application areas identified as being explored in Work Packet 3 (primarily agricultural production, although the proposal also leaves open the possibility of application within the health domain) are of undoubted importance, and if RL methods can accelerate the pace of success in experimental research in this field, then this would be a very high value pay-off.

I also view this research as being relatively high risk. The PI has identified a broad spectrum of issues which either limit existing RL methods, or where there has been limited research so far. While these are undoubtedly important problems to address, it is far from certain that this project will be successful across all aspects of the proposed objectives. Falling short in some objectives would still result in useful outputs from the project, but these could fall well short of the maximum benefits which would be obtained by a complete success.

Scientific Approach

I found it difficult to evaluate the quality of the scientific approach, as this was not an easy application to understand. Partly this relates to a mismatch between the approaches being proposed and my own expertise - my background is in RL, and so I can follow the nature of the problems identified by the proposal, and the limitations of current RL algorithms to address them. However I am less confident in my ability to judge the likelihood of success of the proposed solutions, given that they are largely grounded in areas such as statistics and multi-armed bandits, where I have much less knowledge. My own perception is that there has been limited crossover of ideas from MABs to fully sequential decision-making, but that may just be my own bias.

The other aspect which made this proposal difficult to evaluate is the quality of the application itself, which was not clearly written. While the proposal does a good job of identifying many problems to be addressed, and proposed

methods to address these problems, in many cases there wasn't enough detail provided for me to clearly understand why those methods would be expected to be successful. Maybe this was inevitable given the page limits and the ambitious scope of the project, but there was overall a lack of detail about the proposed research, or discussion of pilot studies etc which would give more confidence in the proposal. This wasn't aided by issues around the presentation of the proposal, which many typos, grammatical errors, and (in B2 especially) numerous missing citations - while I certainly wouldn't argue against funding a proposal purely on the basis of such issues, they have nevertheless impeded my ability to be fully confident in the proposal's feasibility given that I did not fully understand large parts of it.

I also note that while the PI has cast a very wide net in identifying problems which might arise in applying Deep RL methods to complex problems such as agriculture, they do seem at times to have overlooked relevant existing research. For example, group-RL would seem to have strong ties to federated RL - this has been a very active area of research over the last few years, but is not mentioned in the application. Similarly in B2, p6, [L5] the PI states that "little seems to be known about jointly addressing diverse objectives", which suggests they are unaware of the considerable amount of work in recent years on multi-objective RL, which addresses precisely this issue.

Putting aside these concerns, the overall research methodology and working arrangements do appear to be suitable. There is a strong recognition of the limitations of current research evaluation methodologies in Deep RL research (a field where "cherry-picking" of good results is rife), and the need for more robust, statistically-grounded approaches to evaluation. The resources requested (4 PhD students, 2 post-docs, 2 software/ML engineers) appear reasonable given the scope of the project. The division of work amongst these team members, and the scheduling of the tasks to be performed is appropriate. I do have some concerns about the lack of detail in the budget, which is very brief in comparison to the budgets required for research funding in my own country. In particular there is not much detail around the figure for travel (beyond some general statements about the need for travel), and it's unclear what the additional direct costs are (possibly to do with hosting workshops, but this isn't clear?).

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Good
To what extent does the PI provide evidence of creative independent thinking?	Good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Good

Comments (Optional for reviewers)

The PI is 12 years post-PhD, and from their CV appears to have been continually employed in full-time, research focused positions since then. They have been both an investigator and a PI on several funded research projects in recent years including several large projects (SR4SG, Apprenf, BIP-UP). They have a reasonable amount of experience in supervising HDR students since their habilitation, with 3 completions and 5 current students, as well as having contributed to supervision of several more students pre-habilitation. They also have experience in supervision of post-doctoral researchers. One of the strongest aspects of their track record is their capacity to form cross-disciplinary collaborations, which appears to have been a focus of their career so far, and which clearly provides a solid basis for the more applied aspects of this proposal

Their publication record is commensurate with their career opportunities, with 58 articles. I am unable to comment on the quality of the statistical journals in which they have published, but note they have published in top-tier machine learning conferences such as ICML and particularly Neurips - acceptance rates at these venues is lower than many elite journals, and publications in this area are highly valued in the ML community. There is evidence of moderate-to-good impact on the field in terms of citations of their work.

Reviewer 16

Research Project

Ground-breaking nature and potential impact of the research project

The goal of this project is to assist experimental sciences – and in particular agroecology – in massive and efficient testing in order to identify and recommend personalized best practices. To achieve this, RL is extended in basically three ways: to cope with stochastic environmental conditions without relying on predefined actions (as done by MDP models); to achieve group based (shared) RL to speed up learning from multiple environments; and to achieve practical applicability and acceptance through personalized and explainable rankings of RL strategies.

By aiming at improving the (sample) efficiency of RL, the project addresses a very important issue.

The project addresses multiple (namely, 10) challenges or “bottlenecks” related to RL, of which many would probably be worth a project on its own (on a higher level, this concerns e.g. explainability, reproducibility, coordinated exploration). In this respect, the intended project is high-risk high-gain.

There is gain for agroecology and perhaps beyond (depending on how tailored the developed methods are towards agroecology). The amount of the gain is hard to estimate in the light of the number of addressed challenges.

There is a lack of motivation why RL is needed at all and is presumably the best choice for realizing “On-Farm-Experimentation” (be it in agroecology or any other area), and why e.g. “basic” data analysis and statistics is not sufficient (once the data are available). In this respect, it would have been useful to elaborate on the advantages of an RL approach over a “traditional data science” approach.

As a side remark, unfortunately many references are missing (“?”).

Scientific Approach

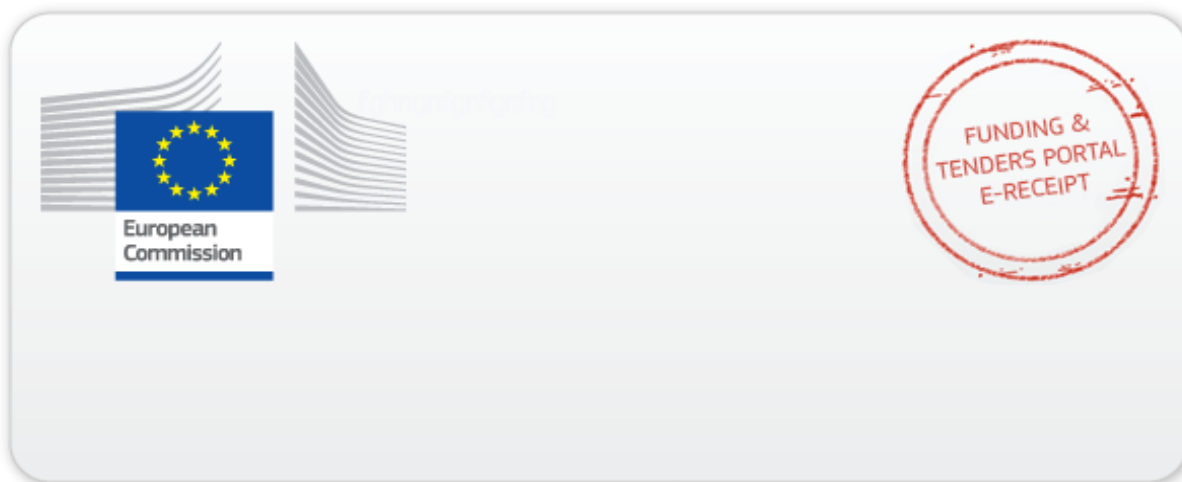
The challenges are well identified and the three work packages are aligned accordingly. Despite the fact that the work packages seem to be more or less disconnected from each other, and despite the PI’s expertise, it appears to be unlikely that all challenges can be solved. The mutual relationships of the work packages (if any) remain somewhat unclear. The rationale behind the assignment of the Postdocs to the different tasks could have been better explained/motivated.

Principal Investigator

To what extent has the PI demonstrated the ability to conduct ground-breaking research?	Very good
To what extent does the PI provide evidence of creative independent thinking?	Very good
To what extent does the PI have the required scientific expertise and capacity to successfully execute the project?	Excellent

Comments (Optional for reviewers)

The PI has broad and in-depth expertise in reinforcement learning and statistics. He has achieved international reputation and visibility, and he has published at some leading conferences (e.g., NeurIPS and ICML). He is well experienced in leading larger projects.



This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Funding & Tenders Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Funding & Tenders Portal.

(<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq>)