Thank you for inviting me to comment on the manuscript: "Adaptive Behavior of Over a Million Individual Farmers Under Consecutive Droughts: A Large-Scale Agent-Based Modeling Analysis in the Bhima Basin, India". I am an expert on agent-based land use change models. This may limit my expetise on the hydrological aspects of the presented work.

The authors use two coupled models (one ABM and one hydrological model) GEB (please spell out what GEB stands for) to model the land use of presumably 1.4 million farmers in the face of consecutive droughts over several decades. Unfortunately, the text is very rich combing description of drought modelling, theory, ABM model, and many more things. Therefore, at least for me it is impossible to understand the model(s) in detail and to put the results into context. Therefore, I am not able to appreciate the model and its results sufficiently, although the topic is timely and I guess the approach is relevant and promising.

Maybe the use of a protocol such as ODD+D (Mueller et al. 2013, Describing human decisions in agent-based models - ODD+D, an extension of the ODD protocol) could help to present the model in a more digestible way. At least the authors should summarize somewhere (maybe as a table) an overview of the properties of the agents and or agent types.

Thank you for the suggestion, this is indeed helpful as it was difficult to both cover all aspects of the model and keep a concise paper. An ODD+D has been added to the supplementary information that we hope gives better insight into the full working of the model. If you are still missing more technical insights, the paper by De Bruijn et al. (2023) discusses the technical base design of the Geographic Environmental and Behavioural (GEB) model.

The authors present many responses from their model: "Our analysis examines how these adaptations affect profits, yields, and groundwater levels, considering, e.g., farm size, risk aversion and drought perception." Maybe it would be helpful to reduce the number of responses and/or scenarios? Especially in Figures 5 and 7 I would encourage the authors to present less panels and to focus on a more narrow narrative.

We agree that some panels had unnecessary information and that 8 panels is too much, thus we reduced figure 5 to 3 panels and figure 7 to 6 and removed/trimmed paragraphs in the description of the results. Furthermore, we have moved the sensitivity analysis to the appendix.

However, our main finding is that the combination of hydrological and socio-economical factors steer the area and especially certain groups of farmers into a more vulnerable state. For this, combinations of the results of yield, crop choices, income, groundwater and wells are needed, as the interplay between those leads to this result. We also find

that being able to model the interplay of these factors is the major strength of this (new) type of model.

I appreciate that the authors follow a theory to justify their decision model. However, the authors need to guide the reader more carefully, since the SEUT is not a standard theory for all Economists and certainly not for all land use change modellers. Fishburn (1981) is a review of several theories and the list of papers suggested as examples of the application of SEUT needs to be critically revised (Groeneveld (review), Haer and Wens do not mention SEUT and do not cite Fishburn). When SEUT is introduced it also should be mentioned that the authors use imitation and "bounded rationality" (line 215) as well in their decision modelling. Later on also prospect theory is considered.

Thank you for your sharp comment. Wens (2020) compares the "economically rational" EUT with the PMT, and concludes that a more bounded rational theory covers behavior better. This was originally used as justification of using the SEUT over the EUT, but this has been lost over subsequent versions. This happened for Groeneveld as well. Haer (2020), however, does not refer to it as the SEUT, but they use the exact same theory (and calculation) of the expected utility for their boundedly rational agents in effect, where the EUT is altered by the changed perception of probabilities due to having experienced an event. The references have been changed and the justification has been expanded in the main text (line 145 to 155) and added in the ODD+D protocol (section 2.1.3).

In general, I am missing an argument why it is useful to model > 1 million agents. Other authors decided to gain knowledge by aggregating actors to agent types in their land use models (e.g. https://landchange.imk-ifu.kit.edu/CRAFTY or the work by Millington et al. https://www.jasss.org/11/4/4.html). It would be great to see an argument developed why this computational demanding approach is seen more appropriate to answer questions of land use change. This is especially critical since the authors argue that it is not computationally feasible to compute for all agents the SEUT for all 300 options ("unique crop rotations"). Would it make sense to compute less agents and therefore consider all 300 options?

There are several reasons why we decided to not aggregate agents (meaning for each farmer in the basin, we also simulate one agent, what we call "one-to-one"). First and foremost, we do not know what a representative agent for our area is (Page, 2012) and by pre-emptively aggregating agents, we may lose interactions that we were not aware existed in the first place (Page, 2012). This is especially true for an area so heterogeneous as the Bhima basin in India, where there are extreme differences in landholder size (Desai et al., 2008) , which factor through in other agent attributes such as which crops they initially cultivate (Department of Agriculture & Farmers Welfare India, 2001), their access to credit or their social factors (Hoda & Terway, 2015; Maertens et al., 2014; Udmale et al., 2015). For example, if we were to aggregate to one agent per grid cell, we would already lose out on the process where larger agents have more funds to invest within

similar budget constraints, and tragedies of the commons, where larger farmers extract more groundwater and adjacent smaller farmers are unable to access the deepening groundwater. Instead of one grid cell per agent, we could attempt to scale up all farmers, adaptation costs, etc. and reduce the total number of agents, but this would require many parameter scaling adjustments, and it is unknown if model processes and interactions (including those with the hydrology, for which spatial factors are generally quite important) would remain similar. We agree however that finding what results the model would produce with different levels of aggregation would be a very interesting future study, and we thus added it to recommendations (lines 557-559). However, for such research to be possible, we require the development of these large and efficient models that are able to simulate at this detail and scale in the first place.

The second reason for keeping more agents while arguing that it is inefficient to calculate 300 options for one agent, is that more agents do not scale linearly with computational times in GEB. Due to the high degree of vectorization in the model, many agents doing one operation can be simulated much more efficiently than fewer agents doing many operations (i.e., 1.5 million agents doing 1 action is substantially faster than 5000 agents doing 300 actions). Additionally, as we always need to simulate the full region's hydrology, fewer agents may not bring about the same computational advantages as with non-coupled ABMS. We mention that we do not aggregate agents (lines 111-112) and added additional clarification with regards to why we chose to not aggregate to the initialization section (lines 287 to 290) and to the ODD+D protocol (section 2.7.1).

I have difficulties to understand the results. To my understanding imitating the strategy of more successful agents in the neighbourhood of an agent is at the heart of the presented dynamics. In the abstract this is not mentioned: "n adaptive scenarios, farmers can either do nothing, switch crops, or dig wells, based on each action's expected utility." If my reading is correct the imitation aspect should be mentioned early on and should be discussed in a diffusion of strategy/technology context. What are the updating rules – synchronous or asynchronously? How many neighbours are considered? What is the initial trait distribution of actors. Is there a spatial structure in the inital trait distribution? Is the number of farmers constant over the years? Form Figure 4 it seems that the model does not show much variation between runs. Would have been less agents sufficient? What source/help of A.I. have the authors used for what?

Imitation in combination with calculating the utility using the SEUT is at the heart of crop switching: agents compare the expected utility of a selection of their neighbors' crop rotations and choose the rotation with the highest utility. For choosing whether to dig a well, agents do not look at their direct neighbors. However, for calculating the expected utility of doing nothing or digging a well, it is assumed agents know the "true" added value of a well. Since wells both reduce damages during drought years as well as structurally increase water availability, which affects agents differently based on, e.g., the

precipitation they receive, the crops they cultivate, etc. the "true" added value of wells is difficult to empirically predict beforehand. In other models this is often given empirically as, for example, a depth-damage curve in combination with flood maps (flooding). However, we use the agents that already had wells in the model itself to determine this added value, as we believe this gives a more accurate representation given the large differences in contexts for the agents. Therefore we don't necessarily see this as imitation of other agents, but just as a way to determine the objective added value of wells given a farmer with x crop rotation at x location. We have added more explicitly the difference between the crop switching and well adaptation to the text (161 to 171). Furthermore, the additional factors you have requested are all in the ODD+D protocol (updating rules: 1.3.1; neighbors considered: 1.3.1, 2.4.3, 2.6.3, 2.7.1; initial trait distribution: 2.1.4., added plots of distribution of initialized personal agent factors; spatial structure: 2.1.4., 2.9.1, number of farmers: line 280, 2.1.4, 3.2.1).

There is some stochasticity in the model, which was the original reason for the 60 runs (section 2.9.1 ODD+D) . It does seem this effect is rather low, and could potentially be left out for future studies. This did not have major effects on run times, as these could be run in parallel on a linux supercomputer cluster. Run times were about 20-25 hours for a parallel run and about 15-20 hours for a solo run (without spin-up).

The use of A.I. has been further clarified (lines 703-704).

The role of the "spin-up" period (21 years) needs to be explained in more detail. The model is initialized with data from some point in time (when). Given the substantial temporal dynamics of the responses (see Figure 4) the choice of the length of the "spin up" period should have a strong effect on the results? It is written that the calibration has been done in the period after the spin up from 2001 to 2010? It is difficult for me to understand the evolution of Figure 4. The starting point at 2001 is the result of the spin up period? The period from 2001 to 2010 is calibrated and after that it is the model projection?

We have further explained and addressed these questions about the roles and differences of the spin-up and run (line 327-341).

A full model run consists of a "spin-up" from 1980 to 2001, and a "run" from 2001 to 2015. The spin-up period serves to set-up accurate hydrological stocks in the rivers, reservoirs, groundwater etc., and to establish enough data points for the drought probability – yield relation. At the end of the spin-up, the model state is saved and used as starting point of the run. The start of the run in 2001 was chosen as both the IHDS (Desai et al., 2008) and agricultural census (Department of Agriculture & Farmers Welfare India, 2001) collected data in 2001. As the climate data was available from 1979-2016, the 12-month SPEI was available from 1980. Thus, the spin-up time between 1980 and 2001 was chosen to maximize the duration so that the drought probability-yield relation (the "objective drought risk experience") included as many drought events as possible. Adaptation only

occurs during the run. Two scenarios were run: one without adaptation, where agents maintained the same crop rotation and irrigation status as at the start of the model, and another where agents could change their crops or dig wells according to the decision rules outlined in section 2.3. Both scenarios use the same spin-up data.

The model was calibrated between 2001-2010 as we only had discharge and yield data during these years. We would have preferred to calibrate for the full climate data range (meaning until 2016) if the data were available, as the goal of this study was not prediction, but explanation about adaptation and risk under these consecutive droughts.

How are small, medium, and large field farmers defined in terms of hectare?

The ha cutoffs for small/medium and large farms were mentioned in the first section of results, but are now repeated in the figure descriptions for clarity.

Thus, overall I have the impression that potentially great insights are hidden in the current text. More specific and potentially less research questions could help to narrow down the story to allow easier access to the main highlights of the study. And at least for me it would be necessary to have a clearer motivation why it is beneficial to consider so many agents at the very same time.

Specific comments

Abstract: "realistically simulate" – That is maybe personal but I would avoid phrases like "realistically simulate" since it is a model and the best one can do is to model something useful in respect to the research question.

Agreed, it is still far off to be considered truly realistic. Removed the adverb.

The models are written in Python?

Yes, added it in line 104.

Lines 73/74 What means "one-to-one scale"?

Explained what the authors and De Bruijn et al. (2023) refer to as "one-to-one" scale: for every farmer in real life we have a representative agent.

Lines 78: What "simple assumptions of human behaviour"?

Expanded slightly on how behavior was represented before. Lines 77-80.

Figure 1: it is not clear that some boxes are empty – please explain.

This was done to signify the simplification. They are now removed. Line 102

Line 101: reservoir operators – is this agent type considered in the presented study? "reservoir operators" are never mentioned again.

This agent is present but not changed in this study. The ODD+D protocol discusses how the reservoir operators determine downstream and irrigation release.

Figure 2: The land cover map presents less classes (e.g. agricultural land) than used in the results? How did the authors discriminate the different crop types?

Section 2.5 in the main text describes how farmer agents are initialized. These are divided over the land use class "agricultural land". Each farmer controls their own hydrological response unit (HRU, added to section 1.2.4 of ODD+D) (De Bruijn et al., 2023), on which they can make their own land management decision. Thus each agent has their own plot of land where they have their own crop rotation (including the crop's specific characteristics and the implications of these for soil processes, evaporation, crop planting dates etc.).

Lines 132ff – Can you define in term of your indicator (SPEI) what a severe, moderate and so on drought is? I guess there are thresholds? Please specify.

Added the thresholds of the relevant categories of McKee et al. (1993) to the main text. Lines 138-140

Line 156: Why 5 km radius – is this decision based on a sensitivity analysis?

The 5 km was an error, it was actually 1 km, this is changed everywhere now. This was not based on any particular study, but on a general estimate on how far social networks go. A sensitivity analysis on more "meta" model settings (start date, radius, network size, etc.) would indeed be useful for future studies.

Line 165: C_adapt is not considered in equation 4. Does it needs to read "C_input in eq. 4 on current market prices"?

For clarity, we have changed C_adapt to C_well. The C_well signifies the costs of a well. The C_input signifies the costs of agricultural inputs (i.e. cumulative costs of seeds, fertilizers, etc.). Both signify the costs part of the equation, but indeed the line 165 was unclear, so I have added C_input there. (lines 182-184)

Equations 1-4 please explain subscripts x, d, and m.

Added the description in the text. Lines 182 to 187.

Lines 197: crop costs – crop costs C_input are dependent on the type of crop? If so how?

Yes, they are dependent on the farmers' crop type. In the preprocessing of the model, all cultivation costs are sourced from Ministry of Agriculture and Farmers Welfare in Rupees (Rs) per hectare. (https://eands.dacnet. Nic.in/Cost_of_Cultivation.htm, last access: 15 July 2022) for the full model run time. During planting, the crop that the agent is planting is taken from the crop rotation, the crop parameters are set in the model, and the cost at that specific moment in time is sourced from this premade dictionary of cultivation costs.

Equation 9: Where is the "crop coefficient" Kc used? The duration of different harvesting stages are not crop-specific?

All crop factors are crop specific, added clarification. The Kc is used to determine the crop-specific potential evapotranspiration from a reference evapotranspiration. Added this to the main text as well.

"Agent initialization": Spell out IHDS, the authors should give the ranges of the agent properties. Showing boxplots or other useful representation of the distributions of attributes would be useful. I have no intuition for example how net income is initially distributed among the 1.4 million agents and how it is developing over time.

Spelled out IHDS and added boxplots to the ODD+D in section 2.1.4.

Line 326: (7g)?

7g highlighted the crop market price evolution. It is now in the appendix, as it is not a model produced result, but is relevant for several figures in the results.

Figure 4c: Is it reasonable that one crop is going from 0.05 fraction to the dominating crop?

It is if we look at the decision rules in the model in combination with the strongly risen prices of Groundnut. However, it is not if you compare it to reality. We discuss this in the discussion, and give some explanation and recommendations to resolve it. (Lines 540 to 566). Additionally, we now added some initial argumentation in the relevant results section (lines 359-362).

Figure 5: Very busy graph (6 panels). The authors may want to focus on some panels.

Reduced the figure to 4 panels.

Figure 6 and elsewhere: Specify the unit Rs

This is the Indian currency Rupees. Added the specification.

Figure 8: Too many panels. Legend unreadable. Please focus on the important aspects.

Reduced the panel to 6 panels and increased the font size of legends on each figure.

Technical corrections

The figure labels are too small and therefore hard to read.

Increased the size of all legend/x/y/ labels and titles

Figure 4c and others: colours are too similar – hard to differ crops

The Viridis color palette was chosen to accommodate colorblind individuals. However, it is true that a continuous color palette was not the right choice for categorical data. We

now use the OKABEITO color palette (Okabe & Ito, 2002), which should still be colorblind friendly, but better suited to categorical data.

Lines 65-66 and elsewhere: Inconsistent in-text citation of Udmale et al. 2014/2015

Adjusted the references.

Line 114 and elsewhere: "95 %" -> "95%"

Replaced the instances.

Line 263: The authors refer to figure 3 in Jun et al. 2014? Jun et al. 2014 is a comment in Nature without Figures to my understanding. Please check.

Thank you for spotting the mistake! We intended to refer to Jun et al. 2014 to show which agricultural land data we used, and to refer to figure 2 (not 3 as we mistakenly did!) of our paper to show a visualization of the agricultural class in our study area. However, we see that this was quite unclear. We now clarified the text (Lines 318-319).

"Sensitivity Analysis": What are 300 distinct samples. Sampling from what distribution?

The latin.sample function from SALib uses Latin hypercube sampling. In Appendix B.4 is is further explained from what this was sampled. Lines 649-667

Line 306: Where does stochasticity enters the model(s) and how?

We added a description of the stochasticity in section 2.9 of the ODD+D protocol

Citation: https://doi.org/10.5194/egusphere-2024-1588-RC1