To Mariano Moreno de las Heras:

Attached please find our revised manuscript entitled "*Benefits of a robotic chamber system for determining evapotranspiration in an erosion affected, heterogeneous cropland*" to be considered for publication.

This work has been previously submitted to HESS, with "major/minor revisions" being suggested to apply to our manuscript. The main recommendations of both reviewers were to better discuss the proportion of measured vs. modeled data and possible uncertainties of our approach compared to other approaches estimating ET; improving our description of how calibration/validation were carried out.

We have substantially revised our manuscript according to these suggestions, taking into account all criticisms and comments. In terms of the amount of originally measured data it is true that the term "gap-filling" is misleading here, since we calculate our continuous half hourly ET fluxes (as well as the cumulative ET) as described with data predicted by our data-driven models which were calibrated and validated using our original measurements. For this reason, we have changed the terminology from "gap-filling" to "modeling" throughout the manuscript and additionally discuss the implications of our approach in more detail particularly in comparison to other approaches determining ET. Moreover, we now include a more detailed description of calibration and validation and a more detailed discussion on the advantages and disadvantages of our system.

In conclusion, we believe the implemented revisions improved the manuscript significantly and thank you very much for your consideration.

Sincerely,

Adrian Dahlmann

Review of "Benefits of a robotic chamber system for determining evapotranspiration in an erosion affected, heterogeneous cropland" by Adrian Dahlmann et al., for HESS.

Summary
This manuscript reads much more cleanly than the previous, with a clearer vision, better writing, and more concise presentation. The hypotheses are now explicit and the flow is greatly improved. While I see a number of places where the language can improve – both in grammar and also to be more concise – the work is stronger.

Minor comments
Consider making the intro more concise. I think the paragraph from L36-56 could easily lose 100 words and be more focused (e.g. focus on a few of the challenges in the first 5 lines, reduce reference to six degradation processes, avoid phrases like "indispensable prerequisite" , consider whether both "crop yields" and "related crop productivity" are needed alongside "agriculturally used".

Thank you for this suggestion. We have now shortened the introductory section by 68 words.

L60 I think T is not 90% of ET? Perhaps you mean ET is 90% of the water budget? Anyway it's known that E and T are important, I'm not sure how much detail we need here.

It is true that this sentence is not mandatory. This sentence was removed as part of the shortening of the introduction.

L140 consider starting a new paragraph at "in the following" and decide whether n-d and d are actually good labels (to me they are too short; it was difficult to go back and find this). (see also the complexity of L309; perhaps also more intuitive names could be given to the soil types)

It is true that the n-d and d abbreviations are very short and are used continuously only at a much later stage. For better understanding and comprehension, we now use these abbreviations earlier and more frequently (starting in L148).

Since the abbreviations of the soil types have already been used in other publications, we would like to keep these names for better comparability.

L161 why is a reduction in ET due to lower light not expected? In many landscapes there is a near 1:1 relationship between radiation and ET.

It is true that there is a close relationship between the PAR and ET. Our chamber system is designed to minimize the impact on the physiological activity of the enclosed vegetation. To achieve this, the chambers are made of highly transparent materials with a low PAR reduction. In addition, in our experimental setup, the selection of the calculated flux closest to the start of the measurement, the short closure times, and the ventilation of the chamber should minimize physiological changes (e.g., stomatal responses) and thus minimize changes in ET fluxes.

In similar experiments with dark chambers and chambers of this material (unpublished data), we have found that gas fluxes do not differ between dark and light, and thus no plant response is detectable in the first 5 minutes (at least for the crops we considered). For this reason, we expected a negligible effect on calculated ET due to PAR reduction. We have now added this information to the manuscript.

L165 I'm not sure "Sensor death time" is a clear concept

We now call it "sensor response time".

L275 or elsewhere can the Kc values be given?
We now added the K_c values in brackets.

It now reads: "ET$_c$ (Equ.6) was calculated from the crop factor Kc (K$_{c\ ini}$ = 0.3; K$_{c\ mid}$ = 1.15; K$_{c\ end}$ = 0.4) and the potential evapotranspiration ET0 using monthly averages (DWD 2022)."

L330 and throughout this section some numbers would be helpful – how much overestimation and underestimation (Either in mm or relative error, for example).

We have now calculated the mean error for the calibration in three different ET flux ranges (Table B4; less than 2, between 2 and 4, and greater than 4 mmol m-2 s-1) for all modeling approaches and refer to them in this section.

L334 "perform good" is not grammatically correct and should be "perform well'; perhaps "scores in the good category" would help make it both align with Table 3 and be grammatically correct.

Thank you. We now use "scores in the good category".

Some parts of the discussion could be split into clearer paragraphs and more concise language could be used throughout.

We have partially rewritten the discussion and added additional paragraphs for better understanding.

L492 "diurnal cycles" are not really explored in this work; consider shifting the focus of this a bit (or even adding some more rigorous examination of diurnal cycling in the paper / appendix).

We now present and discuss a new figure (Figure 6) with diurnal cycles in ET fluxes during the cultivation season for one sample day per month (day with the most measurements) and corresponding mean error (ME; two digits rounded).

L500-505 isotopes is a strange ending; even the CO2 fluxes mentioned by the other reviewer could be a better ending since that data exists.

We have now rewritten the conclusion to better represent what is already possible with the new system and what could be possible in the future.

It now reads: „In combination with CO$_2$ measurements, the novel FluxCrane could give new insights in ecosystem WUE in a high spatial resolution using NEE (net ecosystem exchange). In addition, coupled with the GEP (gross ecosystem production) and innovative measurements such as in-situ stable water isotopes (Dubbert et al. 2014; Kübert et al. 2020), a separation of ET into T and E would be possible to assess crop performance due to the plant specific WUE (Tallec et al., 2013) or study root water-uptake dynamics (Deseano Diaz et al. 2023; Kühnhammer et al. 2020). This is particularly relevant for regions with strong spatial heterogeneity in soils and generally low precipitation like the Uckermark and of crucial importance for the terrestrial water balance as well as the prediction of future ecosystem feedbacks (Groh et al. 2020)."

The figure captions can be improved – e.g., Fig 1 (And the plot itself) start in the upper right but then are referenced in the order a-b-d-c (and never e);
We now rearranged the figure.

Fig 2 could separate T and RH
Done.

Fig 4-5 – can remove the word "bottom" and I think the treatments are on the left and the approaches are on the top (I would call gap filling an approach; soil experiments a treatment)
Thank you. We now changed the description.

Technical comments
L18-19, consider: "To do so, a 9-month plot experiment with winter rye was conducted at an eroded cropland located in a hilly moraine landscape in Uckermark, Northeast Germany."
Done.

L33 drop the before ETsum
Done.

L62-65 is too long.
We have now shortened the first sentence there.

L68-9, remove currently dominant (and consider some citations in this area)
Done. It now reads: "At the field scale for example, eddy covariance systems provide high frequency estimates of ET of a homogeneous system (e.g., Ding et al. 2021) while manual chamber approaches are able to precisely capture multi treatment effects (<1m²) on ET at the plot scale (e.g., Hamel et al. 2015)."

L155 add "ly" to independent.
Done.

L157 "further increase the chambers bearing surface" is not clear.
We now rephrased this sentence.

It now reads: "To ensure airtight sealing during chamber deployment, steel frames with a diameter of 1.59 m and a depth of 5 cm were installed into the soil and equipped with an approximately 10 cm wide foam ring to further increase the chambers bearing surface, while deployed."

L212 consider a new paragraph at "This procedure resulted in…"
Done.

L222 consider adding concentration before measurements.
Done.

L232 a space is needed before the parenthesis.
Done.

L249 no comma is needed after case.

Done.

The authors performed a complete review of the manuscript, and answered the reviewer's question comprehensively. The manuscript has been substantially improved, however, there are still a few aspects which need further clarification. One major point is the amount of original data left after data screening. If understood correctly, about 89% of the data originates from a gap-filled strategy. This is quite a substantial number to make an argument on treatment or temporal effects. Further details see below:

Abstract
L13-16: please use the introductory sentences to describe the statement of the problem here, e.g., the devastating impact of erosion and soil dilution on ET in NE Germany and the need to improve gap-filling strategies

Done. It now reads:

"In light of the ongoing global climate crisis and related increases in extreme hydrological events, it is crucial to assess ecosystem resilience and - in agricultural systems - to ensure sustainable management and food security. In hummocky landscapes, erosion and associated topsoil dilution caused by wind, water, or tillage affect crop yields by, for example, reducing soil water storage capacity or decreasing rootability."

L35 please include a concluding sentence here.

Done. It now reads:

"In conclusion, the study could highlight effect of soil heterogeneity on dry biomass, evapotranspiration, and water use efficiency in agricultural systems, emphasizing the importance of considering soil characteristics to optimize crop productivity and resource management."

Introduction
please compare L59 with the statement in L16. How are these two numbers related?

Thank you. We found different numbers for this statement, but with the final quote it should be 100% in both cases.

Material and Methods:
L161-162: Is PAR also reduced? If so, do you expect that ET is influenced in this study (despite the reference), owing to the significant relationship in L350-352, the gap-filling approach (L237/L242) and the discussion in L406?

It is true that there is a close relationship between PAR and ET. For more details on the correlation between our system and PAR, please see the response to the other reviewer's comment.

Regarding the significant relationship between PAR and ET mentioned in the manuscript, we are only talking about the training and correlation of the model. Since we do not expect a reduction in ET due to PAR, these strong correlations do not affect the ET fluxes. We have now added this information to the manuscript.

L225-227: 10% of original data is quite low, at least as compared to EC. Is this sufficient to understand treatment effects, diurnal changes of ET and WUE? Please consider this in your discussion.

10% original data sounds very low at first glance. However, compared to other studies using chamber systems (e.g. Dubbert 2014), we were able to generate 7-10 times more data for each plot. However, it is true that the term "gap-filling" is misleading here, since we calculate our ET fluxes as described with data predicted by our models. For this reason, we have changed the terminology from "gap-filling" to "modeling" throughout the manuscript. In addition, we discuss this and additionally diurnal cycle (with the new figure 6) created with our data-driven model in more detail.

Section 2.5.2: please include information on how calibration and validation was performed (e.g., was there a training and a validation set? How was this decided? Etc.)

We now added a paragraph in section "2.7 Statistical analysis" with a detailed description of our calibration and validation.

Results:
L346/Fig. 8: Is there a possibility to run a post-hoc test after Kruskal-Wallis? This would improve the understanding on the treatment effects.

Thank you for your comment. We did the Dunn-Bonferroni test and included it in this section.

It now reads: "The subsequent Dunn-Bonferroni post-hoc test revealed only a significant difference in DM between non-eroded LV-cc n-d and eroded RG-ca d (p = 0.013). However, no statistically significant pairwise differences were found for WUE."