Reviewer 1

We agree with all seven helpful and constructive points and will incorporate them in the revised manuscript. The comments are well-targeted: several flag tonal adjustments that improve scientific rigor, two surface claims that were under-emphasized in the current draft and warrant clearer treatment, and the remainder address terminology and figure clarity.

A brief response to each follows (full text changes will appear in the revised version).

Agreed. The abstract and introduction will explicitly distinguish the precipitation-driven (climatic) diffusion limitation hypothesis under test from the structural diffusion mechanism (earthworm/soil compaction) discussed later. The two mechanisms operate on different timescales and through different physical pathways.

Agreed. The abstract will note that the in-situ VWC record (2011–2020) postdates the 2002 BES structural breakpoint and therefore characterizes post-decline rather than pre-decline soil moisture conditions.

Agreed. The same caveat will be added at the first introduction of the VWC dataset in the Methods section to establish appropriate context for the mechanistic test.

Agreed. Without direct microbial abundance data (e.g., pmoA gene surveys), the inference from the WS1 calcium amendment null result to biological control is indirect. The phrasing will be softened to "a highly plausible hypothesis consistent with existing biochemical pathways," preserving the biochemical logic while also acknowledging the limitation.

Agreed. Given the small post-break sample ($n = 4$, or $n = 2$ if truncated), "putative shift" is the appropriate phrasing. We will propagate this language consistently through the abstract and body wherever the 2011 HBR breakpoint is referenced.

Agreed. The current figure (and caption) could greatly benefit from this visual representation and breakout. Vertical dashed lines marking the 2002 (BES) and 2011 (HBR) breakpoints will be added to both panels of Figure 5 to allow direct visual alignment with the atmospheric wet deposition trends.

Agreed, and we appreciate the recognition/highlight of this point. The return-to-baseline framing will be added as a paragraph in the Conclusions (Section 5), since it offers an alternative interpretation of the long-term trend that warrants more direct treatment rather than being confined to the Discussion.

We will submit the revised manuscript with these changes incorporated.

Reviewer 2

On the overarching point (that the analyses more strongly reject a precipitation-based explanation than diffusion limitation as a process), we agree, and we will scope the claim accordingly.

The manuscript does not argue against diffusion as the proximate control on CH_4 uptake; gas transport through air-filled porosity is the rate-limiting mechanism in both the hypothesis under test and the alternatives we raise. What the manuscript tests, and does not support at these sites, is the specific attribution of the secular decline to precipitation as the driver of reduced diffusivity.

The distinction is between a reversible, climatically driven reduction in diffusivity (which should track soil moisture) and an irreversible, one-directional reduction (which would not). The observed decoupling of flux from moisture is consistent with the latter.

We will update the title, abstract, and conclusions (and other applicable areas of the manuscript) so that the claim under test (precipitation as driver) is clearly separated from diffusion as a mechanism, which we do not reject.

Major comments

1. The precipitation-driven (climatic) hypothesis under test will be distinguished throughout from diffusion as a general transport mechanism, and from structurally mediated reductions in diffusivity. We will also note that the direct VWC record (2011–2020) postdates the 2002 BES breakpoint and therefore characterizes post-decline conditions, which limits its reach as a test of pre-decline diffusion control.

2. The alternative mechanisms in the Discussion will be reframed as hypotheses consistent with the observed patterns, not as conclusions, given that no microbial, molecular, or earthworm-density measurements are presented. We will replace causal language ("indicates," "demonstrates," "supports") with "may be consistent with," "could reflect," or "is compatible with," and will remove "biological floor" as an asserted finding. We will also clarify that these candidates span two distinct mechanistic regimes that the present data cannot separate: a structurally driven reduction in diffusivity, which would preserve the transport-limited (diffusion) regime, and a reduction in oxidation capacity (e.g., methanotroph community change, nitrogen saturation), which would shift the system out of the diffusion-limited regime entirely. The manuscript will present these as open questions rather than adjudicating between them.

3. We will temper the language around the pre/post-2002 moisture–flux comparison and frame it explicitly as suggestive and subject to post-hoc inference, consistent with our reporting of p = 0.057 rather than treating it as significant. We propose to retain the analysis in the main text, since whether moisture coupling weakened across the decline bears directly on the paper's argument, but to present it with reduced interpretive weight and an explicit statement of the circularity limitation. (We also note for the referee that an interaction model fit on the full record without subsetting — Precipitation × Post-2002 — returns a non-significant interaction, p = 0.24;

this is consistent with the permutation result and is reported in the Supplement as an independent check that does not rely on the post-hoc split.)

4. The revision will add (i) bootstrap confidence intervals on breakpoint timing, (ii) a comparison across alternative changepoint detection methods, (iii) expanded penalty sensitivity analyses, and (iv) explicit discussion of the limited statistical power at Hubbard Brook given the short post-2011 record.

   These analyses have been carried out. The BES breakpoint timing has a 95% bootstrap interval of [2000, 2006] and the HBR breakpoint [2011, 2012]; the intervals do not overlap, so the asynchrony between the two sites is preserved. The BES break at 2002 is recovered by four of five detection algorithms (PELT, binary segmentation, bottom-up, and dynamic programming; the window-based method places it at 1999) and the 2002 break is present across penalty values from 0.01 to 2.0 (wider than the range currently reported), with secondary BES breakpoints (2007, 2017) appearing only at lower penalties. The HBR break is recovered by all five algorithms and is present across penalties from 0.05 to 1.5. We will report the bootstrap intervals in place of bare point estimates. On the Hubbard Brook power analysis: the observed post-2011 shift is statistically clear (pre-break n = 10, post-break n = 4; Cohen's d = 3.35; two-sample p = 0.0001), so the limitation is not that the break may be a low-power artifact. Rather, a four-point post-break regime (two if the final years are truncated) cannot have its level or stability characterized, and only large shifts (minimum detectable $d \approx 1.8$) would be detectable at all. This is why we describe the 2011 break as a putative shift requiring confirmation through extension of the record. We have also harmonized the supplementary changepoint sensitivity analysis with the main analysis so that both use the same annual series and breakpoint-labeling convention.

5. Agreed, with one clarification. We will add discussion of the inherent noise in chamber-based CH_4 flux measurements and the expectation of low instantaneous R^2 in observational datasets, with context from prior upland forest studies. The argument rests on the secular decline failing to track moisture, which constrains the driver of the decline to be something other than the reversible hydrological control. We will make this distinction between instantaneous variance and the multi-decadal trend clear.

Specific comments

1. On verification: re-running the analysis confirmed that the Appendix S3 figure of n = 301 was a code issue rather than a distinct subset. The site-name mapping was omitted from the *supplemental* script, which retained only one (ORUR) of the five VWC sites. Corrected on the full paired dataset (n = 2,415), the quadratic VWC term remains non-significant (p = 0.72; quadratic R^2 = 0.56% vs. linear 0.55%), so the robustness check returns the same conclusion (a quadratic form does not rescue the diffusion explanation), now on the **complete** dataset. We have updated the supplemental script so the code reproduces the corrected value, and will correct the Appendix S3

figure and the Methods description of the VWC–flux match (which is performed at monthly resolution: mean VWC per site-month matched to chamber measurements) in the revision.

2. Addressed under major comment 3. The $p = 0.057$ result will be treated with caution and not interpreted as significant.
3. Vertical dashed lines marking the breakpoints (2002 for BES, 2011 for HBR) have been added to both panels of Figure 5, allowing visual evaluation of the lag between peak deposition loading and sink decline.
4. Agreed, and we appreciate the suggestion. We will add a brief call in Section 4.12 inviting the broader LTER network to make soil archives available for validating the molecular predictions.

Any revision to the manuscript will have these changes incorporated.