

Thank you for the extensive work on the revised manuscript.

I appreciate the effort to respond to the questions raised by adding technical detail on the trade mechanism and providing additional information on the performance of the spatial disaggregation. That said, there remain three main points raised in the previous review that, in this reviewer's opinion, would benefit from further consideration.

First main point

The central concern raised in the previous review has not been adequately addressed in this reviewer's opinion in the manuscript.

As stated earlier: "While results are reported at both European-aggregated and individual country levels, the core focus of the paper is on comparing geographic formulations. A more systematic comparison would therefore come from contrasting standard GCAM regional results with the aggregation of the Europe-detailed version."

The authors acknowledge this point in their reply ("We agree that comparing the core GCAM version with GCAM-Europe is the central objective of these figures"), but ultimately choose not to reflect this emphasis in the main manuscript. Instead, the presentation in the charts continues to focus primarily on demonstrating the capability to produce country-level results. While this is clearly an important feature of the model, in this reviewer's opinion, this choice runs counter to the stated objective of the paper and difficult the real analysis of the model differences and strengths.

Country-level resolution is undoubtedly valuable; however, country-level results alone provide limited insight into systematic differences between model formulations. For this reason, I reiterate my previous suggestion that the main manuscript would be strengthened by presenting systematically comparable regional results (e.g. EU-12, EU-15, Eastern Europe, EFTA, non-EU Europe) when contrasting GCAM and GCAM-Europe. Country-level results would be better used to assess model realism and internal consistency, or otherwise moved largely to the Supplementary Material.

Second point: model realism and temporal focus

The manuscript would benefit more from a systematic comparison of historical and near-term results, between both model variants and against historical data, than from a strong emphasis on 2050 outcomes.

The strong emphasis on 2050 results in the figures complicates the model comparison exercise, particularly given that the scenarios differ substantially in underlying population assumptions, among other drivers. In addition, the absence of historical data for comparison limits the ability to assess model realism. While such a focus could be justified in a study primarily aimed at long-term policy assessment, it significantly hinders the paper's stated objective of comparing two model formulations.

The reviewer appreciates the authors' efforts to respond to earlier comments regarding nuclear deployment, biomass availability, and related topics regarding model realism. However, the manuscript would benefit from a more critical self-assessment of the results beyond the example points raised in the previous review.

Below is a non-exhaustive list with examples of issues that require realism checks in order to better situate the potential of both GCAM and GCAM-EU for real-world policy assessments:

- **Primary energy differences (Figure 5):**
The substantial differences in primary energy in historical years imply very different primary-to-final energy conversion efficiencies in between both models, with direct implications for emissions. These differences need to be explained, and historical values should be explicitly shown in the figure. Presenting EU sub-regional results would also help assess whether this issue is region-specific.
- **Biomass availability:**
In the reply the authors attest that they consider internally sourced biomass levels of 15–17 EJ to be consistent with literature estimates of European biomass potential. This reviewer finds this an overly generous assumption of this model framework for a reference scenario when compared to the available literature. This is of particular importance because the higher spatial disaggregated version of the model is even more generous on the biomass availability which is crucial for the future mitigation alternatives availability.
- **Nuclear power:**
It is important to show historical values to disentangle model realism versus scenario design results as argued in the authors reply for German and other regions result.
- **Wind and solar:**
As the authors reply for the wind decreased share is centered in low historical uptake, historical realism is again a center point that needs to be expressed in the manuscript. Historical deployment should be shown alongside both model scenarios to corroborate model realism on the analysis of the strong changes observed in the results.
- **Electricity final energy:**
Reporting historical and projected shares of electricity in final energy by sector would help clarify structural differences between the model variants and assess the realism of electrification levels and historical alignment.
- **Overall recommendation:**
More generally, a slightly more critical (rather than primarily descriptive) discussion of the difference between the model results and their historical consistency would help to clearly articulate the strengths and limitations of both models throughout the entire manuscript.

Finally, on Table S1 and the discussion of spatial granularity

The comparison of regional granularity presented in Table S1 is incomplete and potentially misleading. If included, it should more carefully reflect the current state of the art in EU-focused IAM analysis. For example, models such as REMIND have configurable spatial resolution and already have published applications with disaggregated EU results that go far beyond the mentioned 12 regions.

As currently presented, the table adds limited value in situating GCAM's spatial disaggregation relative to existing modeling approaches. Moreover, it should be acknowledged more clearly that country-level formulations are not necessarily the preferred option. As the authors themselves note, increasing spatial resolution can limit global assessments due to computational challenges. Many IAMs deliberately adopt a sub-regional EU aggregation to balance regional detail with global model scope for policy assessment and not due to technical limitations. On this we could cite again analysis made with EU developed models like IAMs such as REMIND, or partial equilibrium models like for example ETM-UCL, ...