Point-by-Point Response to Reviewers (Carducci et al. Paper #: NATFOOD-25071821A)
Line numbers are provided for the revised manuscript in "all markup" view, with new text added in red.

COMMENTS FROM REVIEWER #1:

Comment 1: This is a very impressive, thorough, and wide-reaching piece of work which is surely a novel contribution to the ever-expanding literature on monitoring and delivering food systems transformations. I commend the author team for the effort that went into this extensive analysis. Global indicator studies such as these often risk producing reductive conclusions with dubious implications, but I think this work threaded that needle very well. With that being said, most of my comments have to do with expanding on some of the justice and equity implications of possible interpretations of the results, rather than on the methodology itself.

Response: We sincerely thank the reviewer for their thoughtful engagement with our work and for acknowledging the novel contribution and extensive scope of this analysis.

Comment 2: Line 53: The authors state that one contribution is to understand how countries are performing. This brings up an important consideration. The food system is global. To what extent can your analysis be used to inform decisions around trade to avoid things like leakage? Wealthier countries could meet some targets, such as reduced food system emissions, by just outsourcing production overseas. I note, however, that including intensity metrics as you have done is good protection against this. Could there be any use in an indicator around the quantity of food produced domestically to safeguard against this?

Response: We thank the reviewer for this comment. The indicator framework was developed through a comprehensive multi-stage, multi-stakeholder process. The results of this work are published in our 2023 Nature Food paper: Schneider, K.R., Fanzo, J., Haddad, L. *et al.* The state of food systems worldwide in the countdown to 2030. *Nat Food* 4, 1090–1110 (2023). https://doi.org/10.1038/s43016-023-00885-9

Despite this process, there are several conceptual gaps in the framework as at the time (2021-2022), there were no trade indicators that met our criteria, which included feasibility (having recent data and being planned to be updated within the next eight years), coverage (at least 70 countries across regions and income levels) and transparency (no modelled indicators with undisclosed or untraceable methodologies). However, we plan on updating this indicator framework this year to capture any new indicators that may be available, and we will pay particular attention to trade-related indicators or other ways that future analyses can capture leakage.

Comment 3: I would like more insights into how and why the indicators and metrics used were selected than simply "global targets were drawn from the literature." For some indicators used, such as soft drink consumption for example, a %consumption metric obscures the fact that some may drink one soda per year and others one per day. Perhaps a more appropriate indicator may be liters per capita per year? Meanwhile, other critical food system impacts are missing, such as land use change emissions for agriculture (though I note cropland is included). Furthermore, other indicators, such as mobile phone usage, lack a clear logical link to food systems. While this may well be a relevant indicator, the reason for its inclusion is not immediately clear, even from the justification given in Supplementary Table 1. It would be helpful to see more detailed justification

for the inclusion of each indicator and its relevance to the food system, as well as reasons for excluding indicators.

Response: We thank the reviewer for this comment. The indicator framework was developed through a comprehensive multi-stage, multi-stakeholder process. The results of this work are published in our previous papers:

Fanzo J, Haddad L, Schneider KR, Béné C, Covic NM, Guarin A, et al. Viewpoint: Rigorous monitoring is necessary to guide food system transformation In the countdown to the 2030 global goals. Food Policy. 2021;104:102163.

Schneider, K.R., Fanzo, J., Haddad, L. *et al.* The state of food systems worldwide in the countdown to 2030. *Nat Food* 4, 1090–1110 (2023). https://doi.org/10.1038/s43016-023-00885-9

Fanzo et al., describe the architecture of the framework covering five thematic areas and their indicator domains, while Schneider et al. propose the 50 indicators which map onto the thematic areas and domains. In Schneider et al. supplementary files, there is a detailed table of excluded indicators and their justification. As noted above, data availability severely restricts the indicators that can be included in a global analysis such as this.

Importantly, the 50 FSCI indicators included are taken directly from original data sources with minimal data transformations or changing units. For soft drink consumption, this is derived from the Global Diet Quality Project / Gallup World Poll and is defined as the "share of adults who consumed a sugar-sweetened soft drink, which are generally known to be unhealthy, during the previous day." While people indeed consume soft drinks at different frequencies, this population level indicator captures patterns in consumption that enable comparisons across countries. Global quantitative data on soft drink consumption is, unfortunately, not currently available. Mobile phones per 100 people is included as an indicator from the International Telecommunications Union / World Bank , defined as "subscriptions to a public mobile telephone service using cellular technology, which provides access to the public switched telephone network. Includes post-paid and prepaid subscriptions. The relevance of this indicator is such that the number of mobile cellular telephone subscriptions has been included in the computation of food system resilience as a proxy for countries' infrastructure level as well as a direct indicator of response capacity.

In the current manuscript, we provide the updated metadata for each indicator in the Supplementary Data File 1.

The global targets for this analysis were identified through a literature review. We describe this in the results, Lines 184-185, which reads:

"Global targets were drawn from the literature, based on pre-existing international commitments, rights conventions or theoretical minimums or maximums."

We expand on this in the methods, Lines 712-733.

Comment 4: Line 87: There is a caveat regarding the relative values of global benchmarks. If a global benchmark is lower than the target, I think one sentence of clarification of this outcome

would be helpful to understand what this means for the results of those indicators.

Response: We thank the reviewer for this comment. In all cases global benchmarks are lower than the global targets. The value of benchmarks is described in line 531-540:

"In addition, while prior efforts have focused on global targets to inspire ambition (e.g., for the SDGs), this paper also provides global, regional and income benchmarks. For countries far from achieving global targets and global benchmarks, these offer more realistic milestones as comparisons are made to more homogenous groups, with more similar development levels and contextual constraints. The analysis of global targets presented here made clear that many countries and regions are unable to attain them by 2030 – and in some cases, by 2050. While bold targets are vital for sparking commitment, investment, and action, they can become a source of inertia if they are too far away that they seem unachievable. Establishing attainable benchmarks in parallel helps sustain efforts over time by reinforcing that progress is within reach, leading to more realistic action."

In Lines 587-591, we also discuss this caveat specifically:

"There are some limitations to the analysis. First, for some indicators, the global benchmark is significantly higher or lower than the global target (e.g., minimum dietary diversity for children 6-23 months, zero fruits or vegetables for both children 6-23 months and adults). While we presented the global benchmark for these indicators, it would be prudent to use the global target, given equal rights regardless of regional or income contexts."

<u>Comment 5:</u> With many tables in the results given by region, how important is it to look at country-level trends within a region? For example, are there instances where one country is significantly affecting the regional values?

Response: We thank the reviewer for this comment. Supplementary Tables 12 and 13 provide the number of countries included in the regional averages in Figures 2 and 3 (main manuscript) respectively. For regions with few countries or where indicators possess outlier countries that drive the aggregate, it would be important to look at country-level trends. We have added a paragraph in our discussion to address this comment. Lines 356-385 now reads:

"It is also important to acknowledge that a single country's performance may influence the regional average, particularly in regions with fewer countries (Oceania) or when a populous nation drives the aggregate. For example, for minimum species diversity – an indicator which reflects the percentage of agricultural land (crop and pasture) with 24 or more species – Kuwait must increase by 70% per year, in order to meet the global target (set at 100%) and global benchmark (set at 95%) by 2030, compared to the regional average of ~20% per year. In other cases, regional aggregates for food system emissions are influenced by countries with large agricultural sectors (e.g., China and India in Asia, United States and Brazil in Americas, Democratic Republic of the Congo in Africa); these values are proportional to agricultural scale and population but can obscure patterns among smaller countries within the same region.

<u>Comment 6:</u> Line 319: The usefulness of this distinction between the global and regional benchmarks in the case of African affordability of healthy diets is not clear. The point seems to be

that the African benchmark is worse than the global benchmark, such that the potential for other African countries to meet this lower benchmark is more feasible as opposed to the global benchmark. I take issue with this being framed as a "more optimistic picture." Shouldn't the objective of an equitable food system be for all to have the same standards regardless of nationality? I understand the value of creating regional benchmarks to identify regionalized points of intervention, but these regional benchmarks should be contextualized to convey their relative importance.

Response: We thank the reviewer for this comment. We discuss the rationale behind providing both targets and benchmarks in this analysis in Lines 536-548 which reads:

"While bold targets are vital for sparking commitment, investment, and action, they can become a source of inertia if they are too far away that they seem unachievable. Establishing attainable benchmarks in parallel helps sustain efforts over time by reinforcing that progress is within reach, leading to more realistic action. Particularly in more homogeneous regions, a regional benchmark may motivate more countries to achieve it. Regional benchmarks may also provide a shared reference point that encourages knowledge exchange and mutual support between higher and lower-performing countries, potentially identifying exemplars ('regional champions') that provide examples for expanding human and planetary health gains across similar contexts. This is particularly critical as food systems transformation is a shared responsibility, on which cross-country collaboration and support will be needed. At the same time, such benchmarks should be interpreted with caution and should not be used to imply that people in different regions or income groups do not have equal rights to high-performing food systems."

We have also rephrased 'optimistic' to 'realistic' in Lines 559-561*:*

"As benchmarks (whether global, regional, or income-group) are generally less ambitious, their analysis provides a more realistic picture of performance."

<u>Comment 7:</u> To prevent this framework from becoming just another framework, I think clearly outlining the intended uses and actors may help his work find its audience. Is it meant for national governments? Corporations? Farmers? Other researchers? What would implementation look like?

Response: We thank the reviewer for this comment. The intended purpose of this framework was to help guide countries (national and subnational governments) in the monitoring of their food systems pathway (developed and updated at United Nations Food Systems Summit). However, implementation would require multi-stakeholder engagement and action at all levels. National governments would need to work with private sector actors, including food corporations and agricultural businesses, to align policies and practices with sustainability goals. Farmers and farming organizations would be essential partners in implementing production-side changes while maintaining livelihoods and food security. Academic and civil society researchers would contribute to monitoring, evaluation, and adaptive learning processes. The framework provides a common set of indicators that these diverse actors can use to coordinate efforts, track collective progress, and identify areas requiring additional investment or policy attention. Operationalizing this would require context-specific strategies developed through inclusive, participatory processes at national and sub-national levels.

Finally, because the Food Systems Countdown Initiative uses global data sources, it would be prudent for countries to adapt this framework to their country context, including national data sources. We are

currently undertaking the adaptation process with 14 countries (governments, local UN agencies, key NGOs, academics etc.).

<u>Comment 8:</u> Maybe it would be helpful in this paragraph to highlight a few countries which are making good progress on some indicators, and try to identify the contextual reasons leading to their success.

Response: We thank the reviewer for this comment. We have added top performing countries (those closest to global targets and global benchmarks) in Lines 277-287:

"At the country-level, distinct regional leaders emerge based on their proximity to multiple global targets and global benchmarks. Australia leads in the Oceania region and globally, demonstrating the smallest distance to global targets or global benchmarks for 8 indicators across four themes: cannot afford a healthy diet, dietary sourcing flexibility, emissions intensity for milk and rice, experience food insecurity, food safety capacity, government accountability index and government effectiveness index. In other regions, there are varying patterns of performance. Tunisia leads Africa, ranking closest to global targets for 5 indicators, all concentrated within the Diets, Nutrition, and Health theme (all 5 food groups, cannot afford a healthy diet, minimum dietary diversity for both women and children 6-23 months and vegetable availability). Albania, Canada, and Singapore lead their respective regions, with 6, 5, and 4 indicators across multiple themes."

Understanding why some countries are closer to targets than others is context-dependent and warrants deeper examination of national policies, investments and priorities, a future research direction for FSCI. We have highlighted this in Lines 575-579:

"Our assessment re-emphasizes the need to accelerate progress on food system transformation, prioritizing and bundling policies, interventions and innovations that will affect those indicators where progress is most off track, such as food affordability, food insecurity, emissions, food price volatility, and child labor – though which indicators to target most urgently will vary by context due to local priorities."

<u>Comment 9:</u> Can this analysis be used to prioritize which indicators are most critical? Surely this will vary nationally. And perhaps the indicators "most off track" might not be the most critical.

Response: We thank the reviewer for this comment. We highlight the indicators which require intervention in Lines 571-579 which reads:

"Even with an additional 20 years of effort, targets for some indicators remain essentially out of reach, such as food systems emissions, child labor, and food insecurity.

Our assessment re-emphasizes the need to accelerate progress on food system transformation, prioritizing and bundling policies, interventions and innovations that will affect those indicators where progress is most off track, such as food affordability, food insecurity, emissions, food price volatility, and child labor – though which indicators to target most urgently will vary by context due to local priorities."

<u>Comment 10:</u> Potential next step for this research- qualitatively mapping how each indicator interacts with others

Response: We thank the reviewer for this comment. We undertook an analysis on interactions between indicators last year, which is published here: Schneider, K.R., Remans, R., Bekele, T.H. *et al.* Governance and resilience as entry points for transforming food systems in the countdown to 2030. *Nat Food* 6, 105–116 (2025). https://doi.org/10.1038/s43016-024-01109-4

<u>Comment 11:</u> To what extent do you think lack of data availability influences the results? A sentence in the limitations paragraph on the sensitivity of the results would be helpful.

Response: We appreciate this comment. In our 2023 paper we highlight the data gaps across indicators, by country. This is most definitely a limitation of our current analysis, and we've updated the limitations section to address this in Lines 605-608 which reads:

Third, important conceptual and data gaps persist in the FSCI framework, potentially affecting the robustness of this analysis (due to insufficient data across countries and indicators), highlighting the need for continued refinement of indicators (forthcoming) and new data collection to measure food systems transformation.

<u>Comment 12:</u> I think there is room to explore what role wealthier countries should play in facilitating food systems transitions in poorer countries. The country-level framing makes it appear as though each country is responsible for its own food system, when in reality the food system is global, and a great deal of exploitation and environmental degradation in poorer countries persists to meet material demands of richer countries.

Response: We appreciate the reviewer raising this important point about the interconnected nature of global food systems. We fully acknowledge that food systems are inherently global, with wealthier countries playing significant roles in shaping trade, investment, and policy influence.

However, our analysis focuses deliberately on national-level indicators and country-level progress because national governments remain the primary actors responsible for implementing food systems policies and achieving development targets. While international dynamics are crucial, meaningful action on food systems transformation begins with national commitment, national policy frameworks, and national accountability mechanisms. Countries can use global targets and the benchmarks developed, to strengthen their own national/sub-national plans and strategies, and are ultimately accountable to their own populations for food systems outcomes.

We have added a statement in Lines 544-546 to acknowledge shared responsibility:

"This is particularly critical as food systems transformation is a shared responsibility, on which cross-country collaboration and support will be needed."

<u>Comment 13:</u> Line 610: there seems to be a mistake with the parentheses

Response: We thank the reviewer for this comment. We have added a closing parenthesis.

Comment 14: I appreciate the explanation regarding the exclusion of yield indicators, but I wonder whether there's a workaround. We know that we must reduce pressure to convert land for agriculture, and that can be done by improving yields or reducing consumption of land-intensive projects. Could a yield target work if it were calculated by % change rather than as a static target? Might an indicator on meat consumption be appropriate? There are many projections in the literature which estimate requirements for shifting diets by region.

Response: We thank the reviewer for this comment. This is something we are considering in the framework update planned for next year, as this requires disaggregating the yield indicator (into specific crop commodities versus the aggregate crop group) and changing the units of the indicator, so an appropriate yield target for each commodity could be added.

Comment 15: ST2: I think it's interesting that for the category UPF sales, the most underprivileged countries come out "on top." Consider the justice implications of the countries facing the most hunger not being able to afford westernized UPFs and "unhealthy food." What is the relationship between consumption of "unhealthy food" and undernourishment?

Response: We thank the reviewer for this comment. The retail value of ultra-processed foods is defined as total sales of ultra-processed foods in the calendar year per person (USD/person). The desirable direction of this indicator is such that lower is better. In Supplementary Table 2, the countries listed are those which contribute to the benchmark calculation of the mean 20th percentile (or the best performers). Many low- and middle- income countries contribute to the benchmark calculation, meaning that individuals are purchasing less ultra-processed foods in these contexts. Whether this is due to income levels, affordability, availability or other contextual factors is unclear from this metric alone. Additionally, the relationship between income level and UPF sales is complex and not uniformly linear. While some lower-middle-income countries show lower UPF sales due to limited market penetration, we are increasingly observing rapid UPF market expansion in other low- and lower-middle-income contexts. For example:

Global benchmark of UPF sales (current US$/capita/year): 29
Global benchmark of Cannot afford a healthy diet (% of population): 2.0
Global benchmark of Experience food insecurity (% of population): 5.8

Using latest values from Algeria (lower-middle income country):

- UPF sales =166.9 current US$/capita/year)
- Cannot afford a healthy diet = 20.2%;
- Food insecurity = 18.9%

This illustrates that UPF market penetration and food insecurity/unaffordability can coexist - populations may have access to inexpensive ultra-processed products while still lacking access to affordable healthy diets.

COMMENTS FROM REVIEWER #2:

Comment 1: Superficially it looks great. However, the main text is rather vague and there appear to be major issues with the key figures (figure 1 and figure 2), as well as concerns regarding some of the indicators and confusion on units. While the standard of the language and superficial

presentation is high, I was expecting a manuscript of a much higher quality given the series of paper to which this adds.

Response: We thank the reviewer for their feedback and for identifying important areas where clarity can be improved. We agree that precision in the main text and figures is essential for reader comprehension. We have made the following revisions to address these concerns:

1. We have moved Figures 1 and 2 to the Supplementary Figure File (now Supplementary Figures 1 and 2), and have replaced with a new Extended Data Table 2 which provides the Indicator, Units, Global Target, Global Benchmark, Global Mean and Distance to Global Target (absolute) and Distance to Global Benchmark (absolute).
2. We have updated Figure 3 (now Figure 1), using the SDG method. This includes progress category terminology.
3. We have removed the comprehensive score from the results and methods.

Comment 2: The targets are claimed (L136-141) to be set with some emphasis on historical trends. This seems surprising as I would have expected the targets to be objectively set, which seems in fact to be the case, from looking at the supplementary table 1.

Response: We thank the reviewer for this comment. The statement in Lines 136-141 reflects how global targets are commonly set. In our analysis, we did not use historical trends to set targets. These were objectively set, drawing on global frameworks, international commitments and human rights. We describe this in the results, Lines 184-185, which reads:

"Global targets were drawn from the literature, based on pre-existing international commitments, rights conventions or theoretical minimums or maximums."

We expand on this in the methods, Lines 712-733.

Comment 1: Figure 1. An initial glance of this figure it looks good. For example, the split between higher better vs lower better, works well. However, the more you look at it the less the details make sense.

Response: We appreciate this comment. In considering your point, we have shifted Figures 1 and 2 into the supplementary figures (now Supplementary Figures 1 and 2), and have replaced with a new Extended Data Table 2, which provides the Indicator, Units, Global Target, Global Benchmark, Global Mean and Distance to Global Target (absolute) and Distance to Global Benchmark (absolute).

Comment 3: I found the x-axis scale (Figure 1) very confusing. It's labelled as -2.0 to 4.0 range, and a log scale. The numbers on the scale are laid out linearly, so perhaps 4.0 means 10,000 and -2.0 means 0.01 (assuming not a natural log). Clearer labelling certainly required.

Response: We thank the reviewer for pointing out this source of confusion. You are correct in your interpretation—the scale represents $\log_{10}$ values, where -2.0 = 0.01, 0.0 = 1, 2.0 = 100, and 4.0 = 10,000.

We have revised the Supplementary Figures 1 and 2 captions, which now reads:

"Indicators are separated into panels based on their desirable direction: higher values are better (left panel) vs. lower values are better (right panel). Blue circles represent current global mean values, black diamonds represent global targets, and gray lines connect each indicator's current performance to its target. Indicators are organized by thematic grouping within each panel. The x-axis uses a base-10 logarithmic (log_{10}) scale where values represent orders of magnitude: -2 = 0.01, -1 = 0.1, 0 = 1, 1 = 10, 2 = 100, 3 = 1,000, 4 = 10,000. This log scale allows comparison across indicators with vastly different units and magnitudes, as it represents multiplicative relationships as consistent visual distances. The distance between global means and targets indicates the magnitude of improvement needed to reach target performance levels."

Comment 4: Why was index approach for each metric not adopted? While, it can certainly also create challenges, it would also help greatly in the presentation, and there are approaches that have been used in similar situations. For example, having an index value of 1 being associated with the target level and 0 being ascribed to a lower limit 'bad' level. This sort of approach would certainly resolve a number of challenges in the work here.

Response: We thank the reviewer for this comment. Our indicators span vastly different scales: some are indices (0-1), others are percentages (0-100%), and others are unbounded continuous measures (e.g., emissions in kilotons CO2 equivalents). Log transformation is necessary to visualize these together, as it represents multiplicative relationships as consistent visual distances. For example, a 2-fold gap appears the same whether comparing 5 to 10 or 5,000 to 10,000. This allows readers to assess proportional distances from targets across diverse indicators rather than having the visualization dominated by indicators with larger numerical values.

While an index approach could certainly address some presentation challenges, by creating an index across all indicators, we would lose the ability to see absolute differences. The visualization and our analytical approach are designed to show where the largest gaps exist in absolute terms, which is critical information for prioritizing interventions and resource allocation.

For example, consider India's current value for fruit availability and food system emissions using a min-max indexed approach where 100 = target achieved and 0 = worst observed value:

- Fruit availability: Current value of 179 grams/capita/day with target of 400 grams/capita/day → index value of 45
- Food system emissions: Current value of 1,310,000 kt CO2eq with target of ~0 kt CO2eq → index value of 43

In an indexed approach, both indicators would appear nearly identical (45 vs 43), suggesting comparable distances from their targets. However, this standardization completely obscures that closing the fruit availability gap requires a 221 grams/capita/day increase, while closing the emissions gap requires a 1,310,000 kt CO2eq reduction—a difference in absolute magnitude of roughly 6 million-fold. These vastly different scales have fundamentally different implications for investment decisions and intervention strategies. Information about the actual scale of change required is essential for policymakers, but becomes invisible when converted to a common indexed scale. Our log scale visualization preserves proportional relationships and order-of-magnitude differences while retaining unit context, which indexing would eliminate.

<u>Comment 5:</u> The many of the values plotted (Figure 1) I couldn't understand. "Food System emission (kt CO2eq)". For a start the unit is wrong. Not sure whether this was intended to be per capita or a total. The bar appears to go from a global mean today or 4.5 (perhaps 3000) to -2.0 (0.003). Perhaps this could be a net zero food system emissions target (with carbon sequestration in soil?), but the text implies a 70% reduction. Therefore, I'm confused.

Response: We thank the reviewer for this comment. Food system greenhouse gas emissions (carbon dioxide, methane, nitrous oxide and F-gases) is expressed in kT CO2eq (AR5), which is derived from FAOSTAT. The global mean is 83,339 kT CO2eq, the global target is 0 (net-zero) kT CO2eq and the global benchmark is 572 kT CO2eq.

In Figures 1 and 2 (now Supplementary Figures 1 and 2), this converts to:

Global mean: 83,339 kt CO2eq = 4.92 on the $\log_{10}$ scale

Global benchmark: 572 kt CO2eq = 2.76 on the $\log_{10}$ scale

Global target: 0 (near-zero) kt CO2eq = -2.30 on the $\log_{10}$ scale (if we use 0.005 as the near-zero approximation)

<u>Comment 6:</u> Another example is prevalence of undernourishment, in Figure 1 similarly confusing. The target of 0 would seem reasonable, and it what is shown in extended table 1. However, target appears to be plotted at about 1.5 (perhaps 30%).

Response: We thank the reviewer for this comment. Prevalence of undernourishment (% population), derived from FAOSTAT has a global mean of 12.6%, a global target of 0% and a global benchmark of 3.3%. For prevalence of undernourishment, the values are stored and plotted as percentages and the logarithmic x-axis then positions these values at $\log_{10}(12.6) \approx 1.1$ and $\log_{10}(3.3) \approx 0.5$, respectively. This is standard R ggplot2 behavior with the data values remain unchanged, but their spatial positioning on the axis follows a logarithmic scale.

In Figures 1 and 2 (now Supplementary Figures 1 and 2), this converts to:

Global mean: 12.6% = 1.1 on the $\log_{10}$ scale

Global benchmark: 3.3% = 0.52 on the $\log_{10}$ scale

Global target: 0% (near-zero) = approximately -2.00 on the $\log_{10}$ scale (if we use 0.05% as the near-zero approximation)

<u>Comment 7:</u> Figure 2 is equally incomprehensible. Figure 1 and Figure 2 appear to both plot "Current Global Means", to allow different comparison. However, the global mean blue circles are not plotted in the same places across the figures. Not clear why. If figure 1 "is Comparing Current Global Means to Global Targets" and figure 2 is "Comparing Current Global Means to Global Benchmarks", would it not be sensible to combine them. I suspect a reader would be able to understand 3 symbols on the plot, and would allow easier comparison.

Response: We appreciate this comment and agree that simplifying the presentation would improve clarity. However, some indicators in our framework have established global targets (derived from international commitments and human rights conventions), while others have only global benchmarks (representing top performance among countries globally). We felt it was important to distinguish between these two types of reference points, as they represent fundamentally different concepts: targets reflect normative goals and commitments, whereas benchmarks reflect empirically-observed achievement. Combining them into a single figure would risk obscuring these conceptual differences and could create confusion about which reference point applies to each indicator.

Comment 8: Food system emissions in Supplementary Table 2 is also very odd. Still unclear what the unit is. But the contributing countries looks very odd. It seems to be associated with the size of the countries, and not the diet (which would drive most of the emissions). For example, Monaco and San Marino are likely to have a typical high-income diet, so hard to see why they would have low per capita emissions.

Response: We thank the reviewer for this comment. Supplementary Tables 2-4 explain how each benchmark was calculated using 1) total number of countries with available data (using the latest non-missing data point), 2) number of countries included in the benchmark (which countries and their values), 3) the year range of data points that are used to calculate the benchmark and 4) whether the mean of the 80^{th} or 20^{th} percentile was used. As discussed in an earlier response, food systems emissions units are in total kT CO2eq (not per capita), where lower values are better.

Comment 9: This makes more somewhat suspicious of all the metrics. A careful consideration that they make sense and are measuring something meaningful would have been expected. I'm quite sceptical of the all the Resilience metrics, and the contributing countries and values do not seem to make intuitive sense. Pesticide use and Functionality integrity also seem questionable.

Response: We thank the reviewer for this comment. We are unclear if the reviewer is unsure about the selection of resilience indicators, or the performance analysis on resilience indicators. The resilience indicators were selected in the comprehensive multi-stage, multi-stakeholder process published here:

Schneider, K.R., Fanzo, J., Haddad, L. *et al.* The state of food systems worldwide in the countdown to 2030. *Nat Food* 4, 1090–1110 (2023). https://doi.org/10.1038/s43016-023-00885-9

For all thematic groups, we followed the same methodology outlined in our methods section to assess current state, historical progress and required progress to 2030 and 2050 global targets and global, regional, and income benchmarks. Of note, data availability issues severely constrain the metrics available for use in a global analysis such as this. However, in the next year we plan to comprehensively review all selected indicators and would welcome any suggestions from the reviewer on alternative indicators (with sufficient data availability and quality) that we could consider in that exercise.

Comment 10: Units in this supplementary table 2 (and indeed everywhere a value with units is provided) would really help. I know the units are in Supplementary Table 2 (if not apparently correctly in all casesI), but expecting the reader to cross reference is not appropriate.

Response: We thank the reviewer for this comment. We have added a unit column in Supplementary Tables 2-4. Supplementary Data File 1 also provides the metadata on the original indicator data source

and indicator units. We are happy to provide any additional information.

Comment 11: The flow of information presented is too difficult to follow in general. I suggest rethinking all the figures and tables presented in the main text. Extend data table 1 duplicates much of the same information as the Supplementary Table 1, but no where can I find the global mean values tabulated (which seems a key value – hence plotted twice!). The lack of these values is problematic given the issues with the figures 1 & 2, making it hard to decipher what has gone wrong.

Response: We appreciate the reviewer's feedback. Our manuscript presents findings in a hierarchical structure: beginning with global-level results (current status, historical trends, and future progress requirements toward 2030 and 2050 global targets and benchmarks), followed by regional-level analyses (current status and progress requirements toward 2030 and 2050 regional benchmarks). Country-level details at the third level (current status, historical trends and progress requirements toward 2030 and 2050 benchmarks) are provided in Supplementary Data File 3.

We have now added Extended Data Table 2 in the main manuscript which provides the Indicator, Units, Global Target, Global Benchmark, Global Mean and Distance to Global Target (absolute) and Distance to Global Benchmark (absolute).

Extended Data Table 1 and Supplementary Table 1 are complementary, but not the same. Extended Data Table 1 provides the Target Value, and the abbreviated source of the Target Value. However, it does not present the justification (e.g., does the target come from an international commitment, human rights convention or theoretical maximum/minimum). Supplementary Table 1 provides this justification.

Comment 12: I think I understand the extrapolation approach, but equations 2 and 3 are I believe incorrect.

Response: We thank the reviewer for this comment. Equations 2 and 3 have been updated to align with the SDG methodology.

Comment 13: The categorization (likely challenge, potential challenge, or unlikely challenge) seems crude. I think this is just used in Figure 3. Is this required / helpful?

Response: We thank the reviewer for this comment. We have updated Figure 3 (now Figure 1) using the SDG methodology to articulate the number of countries in each progress category, based on progress ratio. The methodology adopted can be found here: https://unstats.un.org/sdgs/files/report/2024/Technical_Note_for_Progress_Assessment_2024.pdf

This has also been updated in our methods text.

Comment 14: "comprehensive score" is mentioned in the method, but is not given in the main results and hardly mentioned in the Discussion. Remove it entirely or present more fully. I would recommend the former.

Response: We thank the reviewer for this comment. We have opted to remove the comprehensive score, as per also the suggestion from Reviewer 3. The results text and methods have been removed.

Comment 15: Unfortunately, I'm going to stop reviewing this paper now, as the work presented is not sufficiently well presented for me to provide a comprehensive set of responses.

Response: We sincerely appreciate the time and effort you have invested in reviewing our manuscript. Based on your feedback and that of the other reviewers, we have undertaken substantial revisions to improve the clarity of the manuscript. Specifically, we have:

- Revised and expanded the figure legends with explicit explanations of the logarithmic scale for Figures 1 and 2 (now Supplementary Figures 1 and 2)
- Added a new Extended Data Table 2 which provides the Indicator, Units, Global Target, Global Benchmark, Global Mean and Distance to Global Target (absolute) and Distance to Global Benchmark (absolute)
- Added a unit column to Supplementary Tables 2-4
- Updated our methodology to align with the SDG methodology for Figure 3 (now Figure 1)
- Removed the comprehensive score

We hope these revisions address the concerns you and other reviewers have raised. We remain committed to making this work accessible to the broad readership of *Nature Food*, and we welcome any additional specific suggestions you may have for further improving clarity.

COMMENTS FROM REVIEWER #3:

Comment 1: This is a well-crafted, policy-relevant paper that substantially advances the conceptual and methodological basis for tracking food systems transformation. It represents an impressive step forward for the Food System Countdown Initiative and it's broad, interdisciplinary community of food systems researchers, scholars and practitioners. The paper effectively integrates diverse strands of data and policy analysis into a cohesive framework for monitoring global progress on food systems. Not an easy task— well done!

Response: We sincerely thank the reviewer for their thoughtful engagement with our work and for acknowledging the novel contribution and extensive scope of this analysis.

Comment 2: My recommendation of major revision does not reflect concerns about the overall quality or direction of the work—it points to areas where clarification or refinement would improve the manuscript. Major points of emphasis include: 1) strengthening the rigor of the comprehensive score; 2) revisiting specific indicator metrics; and 3) enhancing historical trend estimation methods. I believe that addressing these points would significantly improve the analytical rigor and practical usability of the framework, and I am confident that the paper will be highly impactful and suitable for publication with these revisions.

Response: We thank the reviewer for their thoughtful feedback, which has helped us strengthen the manuscript substantially. We have made several significant revisions to address the methodological points raised.

1. Comprehensive Score: We appreciate the reviewer's concerns about composite scoring approaches. In response to feedback from multiple reviewers, we have removed the comprehensive score from the revised manuscript to improve transparency and interpretability of our findings. This revision allows readers to engage more directly with individual indicator patterns, which we believe strengthens the practical utility of our assessment.

2. Indicator Selection: We want to clarify that the indicators used in this assessment were established through a rigorous, multi-stage, multi-stakeholder process documented in Schneider, K.R., Fanzo, J., Haddad, L. *et al.* The state of food systems worldwide in the countdown to 2030. *Nat Food* 4, 1090–1110 (2023). https://doi.org/10.1038/s43016-023-00885-9

   These indicators represent internationally recognized metrics for food system monitoring, selected based on policy relevance, data availability, methodological rigor and transparency and alignment with established frameworks including the SDGs. We believe this consensus-driven approach provides comprehensive coverage of critical food system domains while maintaining feasibility across diverse country contexts.

3. Historical Progress Assessment: Our approach follows the SDG Progress Assessment methodology (https://unstats.un.org/sdgs/files/report/2024/Technical_Note_for_Progress_Assessment_2024.pdf), the established standard for monitoring progress toward 2030 targets.

   We carefully considered the reviewer's suggestion for moving averages and conducted a sensitivity analysis to assess feasibility. This analysis examined what would occur if we required 2 data points at baseline and 2 at endline between 2015-2024. Our findings showed that this approach would result in:

   - Complete exclusion of certain indicators (e.g., food safety capacity) due to insufficient data availability
   - Loss of 10-60% of countries from the historical progress assessment for remaining indicators (e.g., cropland change, experience food insecurity, female landholdings, government effectiveness index, government accountability index, minimum dietary diversity for children, rural unemployment, rural underemployment, reduced coping strategies, social protection adequacy, social protection coverage, and zero fruits or vegetables consumption in children).

   We note that several indicators already incorporate moving averages where data systems permit (primarily FAO indicators with annual reporting) (Lines 866-871). However, for most indicators in our framework, current data availability does not yet support the more robust approaches, though we agree this would be ideal.

4. Spotlight Countries: We have added text on countries which are currently closest to global targets or global benchmarks, across multiple indicators and themes.

We believe these revisions have substantially strengthened the methodological transparency and interpretability of our work, and we appreciate the reviewer's constructive suggestions in achieving this

improvement.

Comment 3: 1) Comprehensive score (and indicator weighting)
I have some reservations about treating all indicators as equally important— both in the conceptual framing and in the comprehensive score, where all indicators are weighted equally within-group. Some indicators (e.g., malnutrition/access to safe water) capture essential human needs, while others (e.g., fruit availability) reflect dimensions of dietary quality that become more relevant once basic needs are met. A short discussion on alternative weighting or tiered prioritization would help clarify how differences in indicator weights might influence overall results and interpretation.

Response: We thank the reviewer for this comment. We have opted to remove the comprehensive score from our methods and results. There are many technical decisions in creating a composite score, like weighting indicators equally, number of indicators available per country, which can impact the results.

Comment 4: Lines 363-284: The many nuances raised in this paragraph call into question the rigor and interpretability of the comprehensive score. You note that uneven data completeness may not systematically bias results in one direction or the other, which also means there is no way to calibrate our understanding of how missing data is likely to influence the score. Is there a minimum set of indicators with consistent country coverage that could be used for a more standardized cross-country comparison? That would be a much better way to approach a comprehensive score (assuming such a set of indicators existed).

Response: We thank the reviewer for this comment. While we opted to remove the comprehensive score, the calculation required a minimum of 22 indicators; however, several countries in our sample had insufficient indicator coverage. Additionally, the 22 indicators available varied across countries, further complicating cross-country comparisons.

Comment 5: Choice of indicator metrics
a. Food systems emissions indicator (kt CO_2eq): Is this gross or net emissions (incl. removals/sinks)? National targets are often set as economy-wide net-zero targets— a complete "zeroing out" of gross food systems emissions is not usually the policy target. I recommend presenting intensity metrics (per capita / per GDP) to better reflect decarbonization pathways within broader net-zero commitments, with a clear statement on boundaries and implications. This affects interpretations around Lines 286, 312, 358, etc.

Response: We thank the reviewer for this comment. Food system emissions are in gross emissions. As a global target, net-zero emissions are the 'ideal' goal. However, we recognize nationally targets may differ. We plan to update the indicator framework in the future whereby intensity metrics could be expressed as per capita/per GDP.

Comment 6: Choice of indicator metrics
b. % share of agriculture in GDP. You note in lines 604-606 that there is no specific desirable value for this (I agree). A high share of agriculture in GDP can signal lagging economic development rather than more resilient food systems, and the percentage can be easily confounded by strong growth in other sectors— despite improvements in agriculture. I therefore strongly recommend considering agricultural value-added or production value per worker or per hectare metrics.

Response: We appreciate this suggestion from the reviewer. We plan on updating this indicator in the future to agricultural value added per worker.

Comment 7: Line 644-645: Can you provide the year range of latest values somewhere? A two-point CAGR is fragile— especially if the endpoint lands in a year with a shock, such as Covid-19 in 2020/2021. Temporal misalignment could significantly alter results and limit the rigor of cross-country comparisons.

Response: We thank the reviewer for this comment. Our approach follows the SDG Progress Assessment methodology (https://unstats.un.org/sdgs/files/report/2024/Technical_Note_for_Progress_Assessment_2024.pdf), the established standard for monitoring progress toward 2030 targets.

We carefully considered the reviewer's suggestion for moving averages and conducted a sensitivity analysis to assess feasibility. This analysis examined what would occur if we required 2 data points at baseline and 2 at endline between 2015-2024. Our findings showed that this approach would result in:

- Complete exclusion of certain indicators (e.g., food safety capacity) due to insufficient data availability
- Loss of 10-60% of countries from the historical progress assessment for remaining indicators (e.g., cropland change, experience food insecurity, female landholdings, government effectiveness index, government accountability index, minimum dietary diversity for children, rural unemployment, rural underemployment, reduced coping strategies, social protection adequacy, social protection coverage, and zero fruits or vegetables consumption in children).

We note that several indicators already incorporate moving averages where data systems permit (primarily FAO indicators with annual reporting) (Lines 851-856). However, for most indicators in our framework, current data availability does not yet support the more robust approaches, though we agree this would be ideal and could be updated in the future with more data.

We provide three supplementary data files (excel format):

1. Supplementary Data File 1: Has all the metadata per indicator (number of years of data)
2. Supplementary Data File 2: Provides the full dataset (country-level data for 197 countries, from 2000 to 2024, across 59 indicators)
3. Supplementary Data File 3: Provides the country-level performance metrics (country-level data for 197 countries (latest value and year range of data available), across 44 indicators, three metrics (current state, historical progress and required progress) for four reference points (global targets, global benchmarks, regional benchmarks and income benchmarks).

Comment 8: Why not average the first and last 2-3 years of values for each country instead of taking just two data points as first and last year? This would increase sample size and potentially decrease noise.

Response: We thank the reviewer for this comment. We have provided a response in the comment above.

<u>Comment 9:</u> Consider highlighting indicators where intervention is most urgently needed—i.e., which indicators are most off track, and where (and, conversely, where the greatest progress has occurred).

Response: We thank the reviewer for this comment. We have updated the historical progress section in the results Lines 293-299 which now reads:

For nearly all indicators, ***Fig. 1*** *shows that only a small proportion of countries with data have already met the target or benchmark. Most countries show marginal, stagnating or regressing progress in meeting the target or global benchmark. Moreover, if countries do not improve their rates of change, most are unlikely to meet these targets and global benchmarks even by 2050* ***(Supplementary Fig. 3).*** *There are three exceptions: for food safety capacity, vegetable availability and mobile phone subscriptions, close to half of countries have already met or are likely to meet targets by 2030.*

We highlight the indicators which require intervention in Lines 571-579 which reads:

"Even with an additional 20 years of effort, targets for some indicators remain essentially out of reach, such as food systems emissions, child labor, and food insecurity.

Our assessment re-emphasizes the need to accelerate progress on food system transformation, prioritizing and bundling policies, interventions and innovations that will affect those indicators where progress is most off track, such as food affordability, food insecurity, emissions, food price volatility, and child labor – though which indicators to target most urgently will vary by context due to local priorities."

<u>Comment 10:</u> I find the regional analysis too geographically broad to be meaningful. I recommend either including notable subregional examples (e.g., South Asia, Western Europe), or spotlighting a few large, illustrative countries to convey important trends/insights you want to highlight.

Response: We thank the reviewer for this suggestion. We chose UN Continental Regions deliberately to ensure the regional benchmark calculations included a sufficient number of countries for meaningful comparison. More granular regional subdivisions (e.g., South Asia, Western Europe) would result in groups of only a few countries contributing to the benchmark, which we felt was neither statistically robust nor conceptually valuable for benchmarking purposes.

We have added top performing countries (those closest to global targets and global benchmarks) in Lines 277-287:

"At the country-level, distinct regional leaders emerge based on their proximity to multiple global targets and global benchmarks. Australia leads in the Oceania region and globally, demonstrating the smallest distance to global targets or global benchmarks for 8 indicators across four themes: cannot afford a healthy diet, dietary sourcing flexibility, emissions intensity for milk and rice, experience food insecurity, food safety capacity, government accountability index and government effectiveness index. In other regions, there are varying patterns of performance. Tunisia leads Africa, ranking closest to global targets for 5 indicators, all concentrated within the Diets, Nutrition, and Health theme (all 5 food groups, cannot afford a healthy diet, minimum dietary diversity for both women and children 6-23 months and vegetable

availability). Albania, Canada, and Singapore lead their respective regions, with 6, 5, and 4 indicators across multiple themes."

Comment 11: The point that countries tend to be closer to meeting peer-performance benchmarks within their regions than to global targets is not, on its own, a particularly novel or actionable finding—it largely reflects how the benchmarks are defined. Recommend removing in favor of more important insights.

Response: We thank the reviewer for this comment. We agree that this is not a novel finding, but it is actionable in particular (Lines 436-437): "*Despite lower regional thresholds, all regions were still far from meeting the regional benchmarks for ultra-processed food sales and food system emissions."*

The multi-level benchmarking approach (global, regional, and income-group) is central to our analytical framework. It allows us to assess progress across different reference points and provides a more nuanced understanding of country performance than any single target or benchmark could offer. Many countries and regions cannot achieve global targets by 2030 or even 2050, and our analysis demonstrates where more realistic, intermediate milestones may be needed to sustain momentum and political will for continued improvement. We believe the current presentation appropriately balances the need to maintain global ambition while also providing actionable, context-specific assessments of progress.

Comment 12: You could highlight the fact that many countries have plateaued or are backsliding (e.g., lines 430-432). It's a strong call to action.

Response: We thank the reviewer for this comment. We have updated Lines 554-556 to read:

"Across nearly all indicators, most countries are not on track to meet targets and in fact are plateauing or regressing."

Comment 13: Conflating benchmarks and targets (i.e., benchmarks as targets)

Please make the distinction between benchmarks (attainable peer performance) and targets (rights-based or needs-based thresholds) clear early and often in the paper to prevent benchmarks from being interpreted as substitutes for targets. For example, low-income countries should not be normalized to lower aspirations. This is especially important where benchmark values themselves fall short of acceptable thresholds. There is some discussion on this in Lines 462 – 466, but I recommend clarifying these points earlier in the text. Perhaps you can flag which indicators you believe should not be benchmarked at all, as you note in lines 465-466.

Response: We appreciate this comment. We have clarified the difference between targets and benchmarks earlier in the text.

Lines 98-100 (Abstract): "*This paper makes three novel contributions building on the Food Systems Countdown Framework. First, we identify existing targets – based on internationally agreed commitments and normative goals and establish benchmarks— based on empirically-derived peer performance reference values at the global, regional, and income group levels—for 44 Countdown indicators."*

Lines 167-168 (Objective Statement): "*This paper makes three novel contributions building on the Food Systems Countdown Framework. First, we identify existing targets – based on internationally agreed commitments and normative goals and establish benchmarks— based on empirically-derived peer performance reference values at the global, regional, and income group levels—for 44 Countdown indicators."*

The indicators we felt could not be benchmarked were excluded in this analysis (15 indicators), Lines 754-770.

Comment 14: Yield and Productivity Analysis

Line 618-624: I disagree that there is no meaningful way to track yields as part of this effort. Developing methods to monitor the gap between potential yields and actual yields of major commodities (e.g., rice, maize, wheat, milk) would provide immense value as an indicator of food systems progress. Given that some commodity-specific indicators are used in this study (e.g., emissions intensity of milk), there is already precedent.

Response: We thank the reviewer for this comment. We agree that tracking yields is critical for food systems transformation. However, there are two issues with our yield indicators (as they are currently expressed). First, yields of cereals/fruits/vegetables are aggregate indicators that encompass multiple crop types. For example, yields of cereals considers the yields of barley, maize, millet, rice, wheat, oats, rye, sorghum, triticale, fonio, buckwheat, quinoa, canary seed and other mixed grains reported to FAOSTAT. Creating global targets or global, regional, and income-based benchmarks for aggregated crop categories, rather than individual crops, would not be appropriate given the nationally-specific nature of crop production systems and agroecological practices.

Second, the current unit of the yield indicators is in tonnes/ha, not as a ratio or gap (potential yields vs actual yields) by crop. This would require transforming the data (i.e., using crop models which are available for select crops) to create this indicator (for at least 70 countries) and creating an appropriate target and benchmark. Considering these two limitations, we have omitted the yield indicators from this paper and plan to update the indicator framework in 2026 to ensure that more granular yield indicators can be monitored.

Comment 15: Global targets may be inappropriate, but one could easily imagine yield (or yield gap) benchmarks set across agroecological zones— with spatially weighted national averages— to track productivity across similar countries over time. It's outside the scope of this paper to do this work, but crop and livestock productivity are essential food system indicators to include in any holistic food systems assessment.

Response: We thank the reviewer for this comment. For yield gap, this is greatly dependent on crop species and availability of data on optimal versus actual management practices. For many crops, these data are not available with the coverage and quality needed for this global analysis. Therefore, we will need to think critically how to include yield indicators meaningfully in the update of the framework.

Comment 16: Extended Data Table 1: Can you add a column listing the primary data source for

each indicator in addition to the sources of the target? It would be helpful to see the primary sources for all indicators in one place (either here or somewhere else).

Response: We thank the reviewer for this comment. We are unsure if you received this in the manuscript package, but Supplementary Data File 1 (excel file) contains the metadata for each indicator, including the data source.

Point-by-Point Response to Reviewers (Carducci et al. Paper #: NATFOOD-25071821B)
Line numbers are provided for the revised manuscript in "all markup" view, with new text added in red.

COMMENTS FROM REVIEWER #1:

Comment 1: I appreciate the efforts taken by the author team to address my concerns. I am happy with the modifications and clarifications made in response to my comments and those of the other reviewers. I have no further comments at this time.

Response: We sincerely thank the reviewer for their time in reviewing our manuscript.

COMMENTS FROM REVIEWER #2:

Comment 1: I am disappointed with the approach the authors have taken to rebutting in many cases the comments rather than considering and addressing them more fully. There are broadly two major issues that must be addressed prior to publication. Firstly, revise the food system emissions indicator, and secondly, revise the use of figures/tables to support findings.

In my view the most serious issues relate to the food system emissions indicator. This is an important indicator but currently is misleading and unsuitable. I made multiple comments in the last round of reviews on this topic. While, before, it was clear there was an issue with it the reason for this issue was less obvious as what the authors had done was more opaque. However, from the rebuttal comments it is now clear the food system emissions indicator is a country level absolute value, rather than per capita. Why the authors have chosen this is unclear to me.

The authors have added to the text "regional aggregates for food system emissions are influenced by countries with large agricultural sectors (e.g., China and India in Asia, United States and Brazil in Americas, Democratic Republic of the Congo in Africa);, these values are proportional to agricultural scale and population". While correct, this could be misread (as I initially did) to be in 'proportion to'. Something like "Food system emissions total scale with population and agricultural production, so the regional aggregates are heavily influenced by country size", would be less likely to be misread. However, there still appears to be no attempt to justify this choice.

This is extremely misleading and different to all the other indicators that are rates or percentages. This indicator will therefore be highly correlated to population size, and much of the analysis using it is therefore invalid (e.g. regions with countries with small populations would tend to lower regional benchmarks). For example, how should taking a mean of those values be interpreted?

The broken nature of this indicator can be seen in many ways. However, let's take a thought experiment, imagine a situation when a country splits into 2 or more countries (or the reverse). If the larger country was homogeneous with regards to the food system, country metrics for diet, etc. would be unaffected – the new and old countries would have the same values. Great, that makes sense, no progress on the food system has been made with such a change. However, what happens to the current food system emissions indicator, the absolute value gets divided up across the new countries meaning all the new countries have a much lower value, improving their scores relative to regional and global values. How does that make any sense?

While such splitting and combining of countries does happen (USSR, Sudan, Germany, Czechoslovakia), it's not common. However, I use this as a way to illustrate how country size is distorting the metric, making the comparisons/benchmarking applied in the paper invalid for this metric.

One of the rebuttals was that this is what the expert group told us. This is not particularly convincing as an argument on its own. If the work was being presented as an expert group's view on a topic (including what indicators are important) then that would be appropriate. However, it is clear that this is not the case here. The intention is clearly to establish a set of indicators and targets, rather than to report on a viewpoint of the group. Additionally, the having a 'food system emissions indicator' is absolutely right and any expert group would have something along those lines. However, I very much doubt that for most the intention was to report the absolute quantity of CO2e emissions from each country, as opposed to an intensity of emission or per capita emissions, for those countries. The rebuttal has now clarified that the indicator is gross, total food system emissions in kT CO2eq (i.e. not per capita). I therefore suspect that either during or after the stakeholder work there has been some misunderstanding or miscommunication about the intended construct. This is exactly why I previously asked about units, as that seemed a more likely error than taking such an obviously flawed metric.

Response: We acknowledge the reviewer's fundamental concern about using absolute emissions values for cross-country comparison. In response, we have developed a complementary normalized indicator while retaining absolute emissions for tracking planetary impacts. Its description is as follows:

Point 1: Production vs. Consumption Emissions

The rationale for using production emissions rather than consumption emissions stems from alignment with UNFCCC reporting frameworks. Since the Kyoto Protocol, national GHG inventories are estimated at source as production emissions, regardless of whether products are exported. This approach ensures:

1. Consistency with existing national climate commitments and NDCs
2. Actionability—countries have direct policy levers over production practices within their borders
3. Data availability—consumption-based emissions require comprehensive bilateral trade flow data with proper allocation of co-products and animal feeds through complex input-output tables, which are not yet available as time series for all foods (DeCesaro et al., 2025)

Beyond data constraints, consumption footprints relate to a fundamentally different set of interventions than production emissions. Production-based metrics guide agricultural policy, land management, and farming practices, while consumption metrics inform dietary guidance and trade policy. Given that international climate agreements operate on production-based accounting and that the expert stakeholder process prioritized alignment with existing governance structures, we maintained this approach.

We have added clarification in the Methods section (Lines 596-600) explicitly stating that emissions are production-based following FAOSTAT methodology, and discussing the implications for interpretation, particularly regarding countries with substantial food trade.

Point 2: Baseline-Normalized Emissions Indicator

The reviewer is correct that absolute emissions are inappropriate for performance benchmarking across countries of vastly different sizes. To address this while maintaining interpretability and avoiding the confounding effects of ratio-based metrics, we developed a baseline-normalized emissions change indicator.

Specifically, we calculated gross agri-food systems emissions as a percentage of each country's average emissions during the 2000-2010 baseline period (100 + 100*(GHG_total - GHG_baseline)/GHG_baseline). Under this normalization:

- 100% indicates emissions remain at baseline levels
- 50% indicates emissions have halved relative to the country's own historical baseline
- 200% indicates emissions have doubled
- The target remains 0% (complete decarbonization)

We selected normalization as ratio-based indicators introduce interpretive challenges because they contain two changing variables. For example:

Per capita emissions: Declining values could reflect genuine emissions reductions, population growth, or both. More problematically for production-based emissions, per capita metrics implicitly attribute production emissions to the domestic population regardless of whose consumption drives that production. For countries where a majority of production is exported (common among smaller agricultural exporters), dividing production emissions by population yields misleadingly high per capita values. Conversely, countries importing most of their food show artificially low production emissions per capita. Since per capita GHG metrics are conventionally interpreted as consumption footprints, this risks serious misinterpretation.

Per-unit-GDP emissions: Economic growth independent of food system changes would improve this indicator without actual progress toward decarbonization. Additionally, GDP can grow or contract due to factors entirely unrelated to food systems (e.g., sector expansion, financial crises), confounding interpretation.

Emissions intensity per unit output: We agree this is valuable and this is already included in our framework under indicators such as emissions intensity for beef, milk, rice and cereals. This captures efficiency improvements in how food is produced. However, efficiency gains alone are insufficient as absolute emissions must also decline to meet climate targets, hence the need for both indicators.

The baseline normalized indicator isolates emissions trajectories from confounding variables while enabling meaningful performance comparison. Each country is benchmarked against its own historical baseline rather than against countries of different scales, directly addressing the reviewer's thought experiment about country splits/mergers.

We use this baseline normalized indicator for all performance analyses including global, regional, and income group benchmarks (Figures 1-3, Supplementary Tables 1-3, 5-14, Supplementary Figures 1-3, Figure 28). Absolute emissions values remain in the framework and only the trend analysis (Supplementary Figure 4) as they reflect the magnitude of each country's contribution to planetary boundaries—critical information for global climate governance but inappropriate for benchmarking, as

the reviewer correctly notes. This is now documented in the methods Lines 575-600.

"To enable performance analysis of agri-food systems emissions while maintaining interpretability, we developed a complementary indicator that normalizes current emissions relative to each country's historical baseline. Specifically, we calculated gross agri-food systems emissions as a percentage of each country's average emissions during the 2000-2010 period. Under this normalization, a value of 100% indicates emissions remain at baseline levels, 50% indicates emissions have halved, and 200% indicates emissions have doubled. The target for this indicator is set at 0% (equivalent to zero absolute emissions), representing complete decarbonization of agri-food systems.

This approach to normalization was selected to avoid interpretive challenges associated with ratio-based metrics (e.g., per capita emissions, emissions per unit of agricultural output). Ratio-based indicators introduce a second variable in the denominator, which can confound the interpretation of change—for example, declining per capita emissions could reflect either genuine emissions reductions or population growth, or both. By normalizing emissions against each country's own historical baseline, observed changes reflect only emissions trajectories rather than the interaction of multiple changing variables.

We use this baseline-normalized emissions change indicator in place of absolute emissions values for all performance analyses, including calculation of global, regional, and income group benchmarks. Absolute agri-food systems emissions values remain included in the monitoring framework and trend analysis (Supplementary Figure 4) as they reflect the magnitude of each country's contribution to planetary impacts. The absolute emissions indicator tracks production-based emissions following FAOSTAT methodology, not consumption-based emissions. The FAOSTAT domain Emissions Totals includes estimates of GHG emissions from agrifood systems. These cover the emissions generated within farm gate, those associated with the land use change and the emissions from pre- and post-production food processes."

<u>Comment 2:</u> Can the authors clarify whether consumption vs production emissions have been used. I hope that a consumption basis has been used. If not, then the value of this metric is again in doubt. Food system emissions have been decreasing in Europe (SI Table 8), great, but is this just off shoring of production? More clarity on this would be highly beneficial. Specifically, please state the system boundaries of the FAOSTAT "food system emissions" measure used here: does it include land-use change emissions, upstream supply chain emissions (e.g. fertiliser manufacture, energy use), and is it strictly territorial/production-based or adjusted to a consumption basis (e.g. trade-embedded emissions)?

Response: We appreciate the reviewer's comment. This has been clarified in the Methods (Lines 596-600):

"The absolute emissions indicator tracks production-based emissions following FAOSTAT methodology, not consumption-based emissions. The FAOSTAT domain Emissions Totals includes estimates of GHG emissions from agrifood systems. These cover the emissions generated within farm gate, those associated with the land use change and the emissions from pre- and post-production food processes."

For further information, we have cited FAO's methodological note: https://files-faostat.fao.org/production/GT/GT_en.pdf

<u>Comment 3:</u> Methane. There is no separate mention of methane in the whole piece (outside of that implied in CO2eq). Given the importance of Methane and the key role of the food system in emissions of it (from enteric fermentation, rice paddies and food waste decomposition), and the international agreements on reduction (Global Methane Pledge), this seems a strange omission. For example, lower methane emissions from shifts towards plant-based diets (or at least less ruminant-based) would be an important to capture in the metrics.

Response: Methane and its main sources (enteric fermentation, rice production and food waste) are included as part of the total emissions from agri-food systems presented in the paper. Following IPCC emissions guidelines, they are presented as CO2 equivalents using the 100 yr GWP of 25. Additionally, the Food Systems Countdown Initiative also includes indicators on yields of beef, yields of milk and yields of rice. Any changes in these biophysical quantities are likely to be reflected also in the emissions estimates of any new year. We have added new text in the discussion to address this (Lines 479-489):

"Within the Environments, Natural Resources and Production theme, several indicators measure processes or outcomes of agriculture production within the farm-gate and ultimately contribute to food systems emissions. For example, the use of synthetic fertilizers and methane from ruminant livestock and rice cultivation (measured by indicators such as pesticide use, yields of beef, milk and rice, and agricultural water withdrawal), largely emit non-CO_2 gases, such as nitrous oxide. Therefore, reduction of food systems emissions requires significant transformations in farm techniques and production efficiencies in relation to water use, fertilization and livestock management. Additionally, important mitigation solutions also exist beyond the farm-gate, for instance methane capture from landfills, also a large source of this important non-CO_2 gas."

<u>Comment 4:</u> Another serious issue is I'm having trouble getting a line of sight between claims in the text and the cited figure/table. This may have been caused by the removal of the problematic figure (all reviewers raised concerns about this figure). I'm not convinced removing it, rather than addressing the concerns was the correct action.

The text (from L290) now reads "Results show that current rates of progress are insufficient to meet any global target or global benchmark (in cases where indicators lack a target) across most countries by 2030 (Fig. 1 and Supplementary Tables 8 and 9)." The 'current rates of progress' are not shown in Figure 1, so this statement requires the reader to explore in detail and compare values from SI table 8 and 9 and Fig 1. Actually, I'm not even sure if it would be possible just with reference to those figures and tables alone to make any such comment since figure 1 is just the number of countries. This is a prominent claim from the work (first sentence of 'Historical Progress' subsection), but cannot be seen with the data presented in the main manuscript and is hard even using SI.

The next sentence on 'progress' (from line 294) is also problematic: "Most countries show marginal, stagnating, or regressing progress in meeting the target or global benchmark. Moreover, if countries do not improve their rates of change, most are unlikely to meet these targets and global benchmarks even by 2050 (Supplementary Fig. 3)." However, Supplementary Figure 3 is "Number of Countries Meeting the Global Target or Global Benchmark by 2050, by Challenge Area and Theme", which does not appear to substantiate this claim.

In fact, while there is method text and Extended Data Table 3 in the main text giving the method

for the progress assessment (categorising into stagnation, regression, etc.) I can't find where the results of this are shown. Probably buried in one of the extensive supplementary files…. Found it eventually, it is SI Table 9. However, I don't believe this is referred to from the main text.

These inconsistencies, as I raised generically before, are concerning. My suggestion is to think through the key messages and ensure that the figures and tables in the main text support them. As highlighted above I do not feel this is the case currently. Additionally, where statements are made (particularly those not supported by the main figures/tables) make sure an appropriate supplementary figure/table is provided that clearly supports them.

Response: We thank the reviewer for this comment. Our progress assessment methodology (Lines 707-742 and Extended Data Table 3) directly follows the 2024 SDG Technical Note for Progress Assessment, which is the established international standard for evaluating progress towards time-bound targets. This methodology is designed to assess whether current rates of change are sufficient to meet future targets (2030, 2050) by:

1. Calculating actual compound annual growth rates (CAGR) from historical data (Equation 1)
2. Extrapolating expected 2030 and 2050 values assuming current trends continue (Equations 2-3)
3. Comparing extrapolated values to targets/benchmarks to categorize progress (Extended Data Table 3)

Our statement at Lines 260-261 reads: "Most countries show marginal, stagnating, or regressing progress in meeting the target or global benchmark" and cites both Figure 1 and Supplementary Tables 8 and 9.

- **Figure 1** synthesizes these results visually, showing the number of countries meeting targets/benchmarks by theme. While it presents aggregated counts rather than showing individual CAGR calculations (which would be unreadable for 44 indicators × 197 countries), it accurately reflects the underlying analysis.
- **Supplementary Table 9** presents the complete progress assessment results showing the number of countries in each progress category (on target, moderate progress, marginal progress, stagnating, regressing) for every indicator. This directly substantiates the claim about "marginal, stagnating, or regressing progress."
- **Supplementary Figure 3** shows projected attainment by 2050 under current trajectories, directly supporting the statement "most are unlikely to meet these targets and global benchmarks even by 2050."
- **Supplementary File 3** provides the complete performance dataset, where individual CAGR calculations can be found.

<u>Comment 6:</u> This paper claims 'no comprehensive targets have been established against which progress can be measured'. However, the food system transformation chapter for UNEP's Global Environmental Outlook-7, published an extensive set of targets across the food system covering environmental, climate and human health, based on international commitments as done here. Quite likely there are others too. Certainly, this manuscript takes what has been done previously further by operationalising them, but both some reference to previous work and nuance about what is novel would be beneficial.

Response: We appreciate the reviewer drawing attention to the recent publication of the UNEP GEO-7 food system transformation chapter. We have added text in the discussion to specifically note this (Lines 424-427):

"In addition, while prior efforts have focused on compiling global targets to inspire ambition across climate and health domains (e.g., for the SDGs, UNEP GEO-7 food system transformation chapter, EAT-Lancet Commission planetary health targets), this paper operationalized these targets and also provides global, regional and income benchmarks."

Comment 7: I don't find the figures 2 and 3 particularly helpful or informative. The calculations of the required change rate are fine, however what dominates (visually) is the colouring. The reader is implicitly encouraged with this heat map style presentation to take the message the green is easy and red is difficult. However, the thresholds for this colouring are entirely uniform (e.g. <5% green, >15% red). This takes no account of the relatively difficulty of changing these metrics or the regional/country context. At the very least this (currently implicit) assumption should be highlighted and discussed.

Response: We appreciate the reviewer's attention to visual interpretation of the required annual rates of change. However, we respectfully maintain that Figures 2 and 3 serve their intended purpose effectively and that incorporating differential thresholds based on "relative difficulty" would introduce subjective judgments that undermine the framework's objectivity.

The uniform thresholds (<5% annual change in green, >15% in red) were deliberately selected to provide a consistent, interpretable scale across all indicators. These thresholds reflect the reality of the transformation required: regardless of context, achieving a target by 2030 or 2050 from a given baseline requires specific rates of change. The color coding visualizes the magnitude of required acceleration, not feasibility or difficulty. Within the discussion we do state that acting on these rates of change are context-dependent (Lines 469-473):

"Our assessment re-emphasizes the need to accelerate progress on food system transformation, prioritizing and bundling policies, interventions and innovations that will affect those indicators where progress is most off track, such as food affordability, food insecurity, emissions, food price volatility, and child labor – though which indicators to target most urgently will vary by context due to local priorities.[20]"

Comment 8: The response to my previous comment on why not index (marked Comment 4 in the rebuttal) is not great. In fact, the arguments given are exactly why an index approach would help. For one thing the authors appear to be suggesting comparing the numeric value of fruit availability in grams/capita/day with food system emissions in kt CO2eq. This is just weird and wrong on many levels, and not what you are doing.

Re: previous comment 3. The revisions make it clearer, but is it still not 100% accurate. The reason the current version is inaccurate is the x-axis label says 'Value (log scale)", but with the values shown it is actually log10(value), it's not the same. Why not change (as I previous suggested) the axis value labels, e.g. to 10,000 or (10^4), rather than 4, this is clearer, easier to understand and does not require the explanation in the caption.

Response: We thank the reviewer for this comment. Given the normalization of the food systems emissions indicator (now expressed as percentage of baseline rather than absolute kilotonnes), the range across all indicators is substantially more comparable. We have therefore updated Supplementary Figures 1 and 2 to use linear scales rather than logarithmic scales, eliminating the axis labeling concern the reviewer raised. The x-axis now displays actual values with appropriate unit-specific formatting, making the figures more interpretable and eliminating the need for log transformation.

Comment 9: While I'm pleased that the equations have now been corrected, I'm not sure that these are needed. Most readers are likely to know what the annual compound formula is and don't need reminded. A reasonable compromise would be to move them to supplementary.

Response: We thank the reviewer for this comment. The FSCI framework is designed for diverse global stakeholders including policymakers, practitioners, and civil society organizations, many without quantitative research backgrounds. Providing these equations in the main text methods ensures transparency and accessibility for non-scientific audiences who will use this framework to monitor food systems transformation.

COMMENTS FROM REVIEWER #3:

Comment 1: Having more closely reviewed Reviewer #2's comments alongside the authors' response, I still believe the manuscript is strong overall and represents important work that merits publication. It appears that the food system emissions indicator stands out as the primary issue requiring revision (which I also had concerns about), but I do not view this as a barrier to publication if addressed directly.

I generally agree with Reviewer #2 that the use of absolute, country-level GHG emissions complicates cross-country comparison and benchmarking since it's mainly driven by population size and scale of agricultural production rather than reflecting meaningful differences in food systems performance or policy progress. I fully agree that a per-capita or intensity-based metric (e.g., per unit GDP, or per unit production) would be more appropriate.

That said, I believe this issue can be resolved with targeted revision. Specifically, I would recommend that the authors: (i) reformulate the indicator using a per-capita or intensity-based approach; (ii) clearly state their analytical boundaries (production-based vs. consumption-based emissions, agricultural production emissions vs. full food system emissions including land use change and supply chains); and (iii) explicitly acknowledge the need to track progress on more granular GHG emissions trends (e.g., individual gases like methane and nitrous oxide) as well as distinct food system processes (energy use in food supply chains vs. biogenic emissions from agricultural production). Addressing these points would, in my view, resolve one of Reviewer 2's core analytical concerns while maintaining the integrity of the broader framework.

Regarding Reviewer 2's critiques around figures and line-of-sight between claims and evidence — I agree there is still room to improve clarity and presentation in the paper. However, I see these as issues of structure and accessibility rather than fundamental flaws in the underlying analysis. Strengthening the connection between text and figures would improve readability and impact, but does not, in my opinion, undermine the valuable contributions of the paper overall.

I still remain supportive of publication following targeted revision to the emissions indicator. I do not share Reviewer #2's assessment that the manuscript is unsuitable for publication in principle.

Response: We are grateful for the reviewer's endorsement of the manuscript's contributions and appreciate the recognition that removing the comprehensive score improved transparency and analytical clarity.

Per Reviewer 2's concerns, we have substantially revised the emissions indicator as detailed in our response to Reviewer 2 above (Lines 575-600). Key changes include:

- Development of baseline normalized emissions indicator for all performance comparisons and benchmarking
- Explicit statement of analytical boundaries (production-based, following FAOSTAT methodology) in the methods
- Retention of absolute emissions only for tracking planetary-scale impacts within the trend analysis (Supplementary Figure 4), not for cross-country comparison

The reviewer recommends explicitly documenting planned refinements to signal the framework's evolution. We appreciate this suggestion and have expanded the existing Discussion section (Lines 507-511 where we note that: *"The framework is designed to be updated as new indicators and data become available. Forthcoming revisions will concentrate on incorporating advancements that have emerged since the initial indicator selection process, including alternatives for concepts in the framework where performance analysis was not possible due to the nature of the data or indicator construction."*

Point-by-Point Response to Reviewers (Carducci et al. Paper #: NATFOOD-25071821C)
Line numbers are provided for the revised manuscript in "all markup" view, with new text added in red.

COMMENTS FROM REVIEWER #2:

Comment 1: I appreciate that the authors have now acknowledged the central problem with using absolute production-based emissions for cross-country benchmarking and have introduced a baseline-normalised indicator for performance analysis. This is a substantial improvement and goes a long way towards addressing my principal concern.

However, while the normalisation of the emissions indicator substantially addresses my main benchmarking concern, the production-based nature of the indicator remains an important limitation that should be more explicitly discussed. In particular, reductions in territorial agri-food emissions may partly reflect offshoring of production rather than genuine reductions in emissions associated with food consumed by that population. I therefore encourage the authors to add a clearer caveat on interpretation where they discuss progress in emissions reduction.

Less importantly, I see these are described as "complementary indicator" – I'm not sure what is intended by the 'complementary' here. These seem core to the intention of the effort here.

Response: We thank the reviewer for this comment. We have updated the discussion:

Lines 643-658 which now reads: *"As an example, the food systems emissions change indicator used in this analysis, while an improvement from absolute or ratio-based emissions metrics (e.g., per capita emissions, emissions per unit of agricultural output), the underlying construction of the indicator still relies on production-based emissions accounting, which has several limitations for assessing genuine country progress. Observed reductions may partly reflect offshoring of production rather than actual decarbonization as countries can appear to improve by importing emission-intensive commodities, displacing their carbon footprint without reducing emissions. Conversely, countries with large agricultural export sectors may show high emissions driven by production for foreign consumption rather than domestic. This creates perverse incentives to shift to less emission-intensive activities domestically while increasing reliance on emission-intensive imports. Furthermore, countries with historically high agri-food system emissions baselines have more room to demonstrate percentage declines, while those with low historical baselines may find progress more difficult to achieve. Countries that reduce agricultural production will also appear to show progress even without efficiency improvements. Therefore, integrating a consumption-based emissions indicator would provide a more complete picture of climate progress by revealing emissions that are embedded in trade and identifying carbon leakage."*

Comment 2: As noted above, I consider this an issue that needs to be discussed more fully in the manuscript. The accounting approach taken is now much clearer, but the implications of using a production-based rather than consumption-based measure are still not sufficiently discussed. Given the importance of this distinction, I think this should be explicitly addressed.

Response: We thank the reviewer for this comment. We have updated the discussion:

Lines 643-658 which now reads: *"As an example, the food systems emissions change indicator used in*

this analysis, while an improvement from absolute or ratio-based emissions metrics (e.g., per capita emissions, emissions per unit of agricultural output), the underlying construction of the indicator still relies on production-based emissions accounting, which has several limitations for assessing genuine country progress. Observed reductions may partly reflect offshoring of production rather than actual decarbonization as countries can appear to improve by importing emission-intensive commodities, displacing their carbon footprint without reducing emissions. Conversely, countries with large agricultural export sectors may show high emissions driven by production for foreign consumption rather than domestic. This creates perverse incentives to shift to less emission-intensive activities domestically while increasing reliance on emission-intensive imports. Furthermore, countries with historically high agri-food system emissions baselines have more room to demonstrate percentage declines, while those with low historical baselines may find progress more difficult to achieve. Countries that reduce agricultural production will also appear to show progress even without efficiency improvements. Therefore, integrating a consumption-based emissions indicator would provide a more complete picture of climate progress by revealing emissions that are embedded in trade and identifying carbon leakage."

<u>Comment 3:</u> I appreciate the additional text, but I do not think this point is fully addressed. My concern was not simply whether methane is implicitly included within total CO2e (which is of course well understood), but whether the manuscript gives sufficient prominence to methane as a distinct and policy-relevant component of food-system emissions.

Given the importance of methane in food systems and current international policy attention, I think the manuscript should discuss methane more explicitly as a distinct and policy-relevant component of food-system emissions, rather than only subsuming it within total CO2e. At minimum, the Discussion should acknowledge that different gases matter in different ways for mitigation strategy, and that future refinement of the framework could usefully distinguish key gases, particularly methane and nitrous oxide, where data permit.

Response: We thank the reviewer for this comment. We have added this in our limitations section.

Line 659-663 now reads*: "Similarly, future iterations should disaggregate total greenhouse gas emissions by individual gases (methane, nitrous oxide, CO_2) where data permits, enabling alignment with gas-specific policy commitments and mitigation options."*

<u>Comment 4:</u> I do not think my original concern has been addressed here.

My point was not that the authors must adopt indicator-specific or context-specific thresholds, but that the current heatmap-style presentation strongly encourages the reader to interpret green as relatively easy/attainable and red as relatively difficult/unattainable. Given that the thresholds are uniform across all indicators and contexts, that interpretive risk should be acknowledged clearly.

The rebuttal states that the colours reflect the magnitude of required change rather than feasibility or difficulty, but this distinction still needs to be made explicit in the manuscript itself, ideally in the figure captions and/or immediately adjacent text. The sentence added in the Discussion about local priorities ('urgency') does not address this concern, and so not sure I understand why this is being used as a rebuttal in this case.

Response: We thank the reviewer for their additional comment.

Within Figures 2 and 3 captions, we have added "Color thresholds reflect the magnitude of required annual change rates (green <5%, red >15%) and do not represent assessments of feasibility or relative difficulty. Achievement of these rates depends on context-specific factors including political capacity, resource availability, and biophysical constraints that vary across countries and indicators."

Additionally in the results we have added to Lines 451-452: Moreover, this analysis does not consider the feasibility or difficulty in meeting annual progress rates or the critical synergies and trade-offs between indicators (e.g., how a need to reduce food systems emissions might affect other indicators' progress) or population trends and dynamics (e.g., the changing dietary patterns of a large emerging global middle class).

COMMENTS FROM REVIEWER #3:

Comment 1: I appreciate the thoughtful revision to the emissions indicator, and for all the efforts of the authors to respond to this last round of comments. The new baseline normalization approach to tracking food system emissions offers a reasonable alternative to the per-capita or intensity-based metrics I originally suggested.

It is not necessary to make further edits at this point, but it is worth noting that the baseline normalization may advantage countries with high historical agri-food system emissions baselines, which will have more room to demonstrate percentage declines. Also countries that reduce agricultural production will appear to show progress even without efficiency gains.

However these are reasonable tradeoffs given the interpretive challenges of the alternative approaches. Worth noting nonetheless.

I am satisfied with the revision and support publication.

Response: We thank the reviewer for their support. We have added an additional note in the methods section regarding the reviewer's point.

Lines 653-656 now reads: *"Furthermore, countries with historically high agri-food system emissions baselines have more room to demonstrate percentage declines, while those with low historical baselines may find progress more difficult to achieve. Countries that reduce agricultural production will also appear to show progress even without efficiency improvements."*

nature portfolio

Peer Review File

Food Systems Performance Against Targets and Benchmarks Reveals Urgent Gaps and a Path to 2050

Corresponding Author: Dr Bianca Carducci

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Overall comment
This is a very impressive, thorough, and wide-reaching piece of work which is surely a novel contribution to the ever-expanding literature on monitoring and delivering food systems transformations. I commend the author team for the effort that went into this extensive analysis. Global indicator studies such as these often risk producing reductive conclusions with dubious implications, but I think this work threaded that needle very well. With that being said, most of my comments have to do with expanding on some of the justice and equity implications of possible interpretations of the results, rather than on the methodology itself.

Main:
- Line 53: The authors state that one contribution is to understand how countries are performing. This brings up an important consideration. The food system is global. To what extent can your analysis be used to inform decisions around trade to avoid things like leakage? Wealthier countries could meet some targets, such as reduced food system emissions, by just outsourcing production overseas. I note, however, that including intensity metrics as you have done is good protection against this. Could there be any use in an indicator around the quantity of food produced domestically to safeguard against this?

Results:
- I would like more insights into how and why the indicators and metrics used were selected than simply “global targets were drawn from the literature.” For some indicators used, such as soft drink consumption for example, a %consumption metric obscures the fact that some may drink one soda per year and others one per day. Perhaps a more appropriate indicator may be liters per capita per year? Meanwhile, other critical food system impacts are missing, such as land use change emissions for agriculture (though I note cropland is included). Furthermore, other indicators, such as mobile phone usage, lack a clear logical link to food systems. While this may well be a relevant indicator, the reason for its inclusion is not immediately clear, even from the justification given in Supplementary Table 1. It would be helpful to see more detailed justification for the inclusion of each indicator and its relevance to the food system, as well as reasons for excluding indicators.
- Line 87: There is a caveat regarding the relative values of global benchmarks. If a global benchmark is lower than the target, I think one sentence of clarification of this outcome would be helpful to understand what this means for the results of those indicators.
- With many tables in the results given by region, how important is it to look at country-level trends within a region? For example, are there instances where one country is significantly affecting the regional values?
- Line 319: The usefulness of this distinction between the global and regional benchmarks in the case of African affordability of healthy diets is not clear. The point seems to be that the African benchmark is worse than the global benchmark, such that the potential for other African countries to meet this lower benchmark is more feasible as opposed to the global benchmark. I take issue with this being framed as a “more optimistic picture.” Shouldn’t the objective of an equitable food system be for all to have the same standards regardless of nationality? I understand the value of creating regional benchmarks to identify regionalized points of intervention, but these regional benchmarks should be contextualized to convey their relative importance.
Discussion:
- To prevent this framework from becoming just another framework, I think clearly outlining the intended uses and actors may help his work find its audience. Is it meant for national governments? Corporations? Farmers? Other researchers? What would implementation look like?
- Line 428: Maybe it would be helpful in this paragraph to highlight a few countries which are making good progress on some indicators, and try to identify the contextual reasons leading to their success.
- Line 451: Can this analysis be used to prioritize which indicators are most critical? Surely this will vary nationally. And

perhaps the indicators “most off track” might not be the most critical.
- Line 454: Potential next step for this research- qualitatively mapping how each indicator interacts with others
- To what extent do you think lack of data availability influences the results? A sentence in the limitations paragraph on the sensitivity of the results would be helpful.
- I think there is room to explore what role wealthier countries should play in facilitating food systems transitions in poorer countries. The country-level framing makes it appear as though each country is responsible for its own food system, when in reality the food system is global, and a great deal of exploitation and environmental degradation in poorer countries persists to meet material demands of richer countries.
Methods:
- Line 610: there seems to be a mistake with the parentheses
- I appreciate the explanation regarding the exclusion of yield indicators, but I wonder whether there’s a workaround. We know that we must reduce pressure to convert land for agriculture, and that can be done by improving yields or reducing consumption of land-intensive projects. Could a yield target work if it were calculated by % change rather than as a static target? Might an indicator on meat consumption be appropriate? There are many projections in the literature which estimate requirements for shifting diets by region.

Supplementary Materials:
- ST2: I think it’s interesting that for the category UPF sales, the most underprivileged countries come out “on top.” Consider the justice implications of the countries facing the most hunger not being able to afford westernized UPFs and “unhealthy food.” What is the relationship between consumption of “unhealthy food” and undernourishment?

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Superficially it looks great. However, the main text is rather vague and there appear to be major issues with the key figures (figure 1 and figure 2), as well as concerns regarding some of the indicators and confusion on units. While the standard of the language and superficial presentation is high, I was expecting a manuscript of a much higher quality given the series of paper to which this adds.

The targets are claimed (L136-141) to be set with some emphasis on historical trends. This seems surprising as I would have expected the targets to be objectively set, which seems in fact to be the case, from looking at the supplementary table 1.

Figure 1. An initial glance of this figure it looks good. For example, the split between higher better vs lower better, works well. However, the more you look at it the less the details make sense.

I found the x-axis scale (Figure 1) very confusing. It’s labelled as -2.0 to 4.0 range, and a log scale. The numbers on the scale are laid out linearly, so perhaps 4.0 means 10,000 and -2.0 means 0.01 (assuming not a natural log). Clearer labelling certainly required.

Why was index approach for each metric not adopted? While, it can certainly also create challenges, it would also help greatly in the presentation, and there are approaches that have been used in similar situations. For example, having an index value of 1 being associated with the target level and 0 being ascribed to a lower limit ‘bad’ level. This sort of approach would certainly resolve a number of challenges in the work here.

The many of the values plotted (Figure 1) I couldn’t understand. “Food System emission (kt CO2eq)”. For a start the unit is wrong. Not sure whether this was intended to be per capita or a total. The bar appears to go from a global mean today or 4.5 (perhaps 3000) to -2.0 (0.003). Perhaps this could be a net zero food system emissions target (with carbon sequestration in soil?), but the text implies a 70% reduction. Therefore, I’m confused.

Another example is prevalence of undernourishment, in Figure 1 similarly confusing. The target of 0 would seem reasonable, and it what is shown in extended table 1. However, target appears to be plotted at about 1.5 (perhaps 30%).

Figure 2 is equally incomprehensible. Figure 1 and Figure 2 appear to both plot “Current Global Means”, to allow different comparison. However, the global mean blue circles are not plotted in the same places across the figures. Not clear why. If figure 1 “is Comparing Current Global Means to Global Targets” and figure 2 is “Comparing Current Global Means to Global Benchmarks”, would it not be sensible to combine them. I suspect a reader would be able to understand 3 symbols on the plot, and would allow easier comparison.

Food system emissions in Supplementary Table 2 is also very odd. Still unclear what the unit is. But the contributing countries looks very odd. It seems to be associated with the size of the countries, and not the diet (which would drive most of the emissions). For example, Monaco and San Marino are likely to have a typical high-income diet, so hard to see why they would have low per capita emissions.

This makes more somewhat suspicious of all the metrics. A careful consideration that they make sense and are measuring

something meaningful would have been expected. I'm quite sceptical of the all the Resilience metrics, and the contributing countries and values do not seem to make intuitive sense. Pesticide use and Functionality integrity also seem questionable.

Units in this supplementary table 2 (and indeed everywhere a value with units is provided) would really help. I know the units are in Supplementary Table 2 (if not apparently correctly in all cases), but expecting the reader to cross reference is not appropriate.

The flow of information presented is too difficult to follow in general. I suggest rethinking all the figures and tables presented in the main text. Extend data table 1 duplicates much of the same information as the Supplementary Table 1, but no where can I find the global mean values tabulated (which seems a key value – hence plotted twice!). The lack of these values is problematic given the issues with the figures 1 & 2, making it hard to decipher what has gone wrong.

I think I understand the extrapolation approach, but equations 2 and 3 are I believe incorrect.

The categorization (likely challenge, potential challenge, or unlikely challenge) seems crude. I think this is just used in Figure 3. Is this required / helpful?

"comprehensive score" is mentioned in the method, but is not given in the main results and hardly mentioned in the Discussion. Remove it entirely or present more fully. I would recommend the former.

Unfortunately, I'm going to stop reviewing this paper now, as the work presented is not sufficiently well presented for me to provide a comprehensive set of responses.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
This is a well-crafted, policy-relevant paper that substantially advances the conceptual and methodological basis for tracking food systems transformation. It represents an impressive step forward for the Food System Countdown Initiative and it's broad, interdisciplinary community of food systems researchers, scholars and practitioners. The paper effectively integrates diverse strands of data and policy analysis into a cohesive framework for monitoring global progress on food systems. Not an easy task— well done!

My recommendation of major revision does not reflect concerns about the overall quality or direction of the work—it points to areas where clarification or refinement would improve the manuscript. Major points of emphasis include: 1) strengthening the rigor of the comprehensive score; 2) revisiting specific indicator metrics; and 3) enhancing historical trend estimation methods. I believe that addressing these points would significantly improve the analytical rigor and practical usability of the framework, and I am confident that the paper will be highly impactful and suitable for publication with these revisions.

Major Comments:

1) Comprehensive score (and indicator weighting)
a. I have some reservations about treating all indicators as equally important— both in the conceptual framing and in the comprehensive score, where all indicators are weighted equally within-group. Some indicators (e.g., malnutrition/access to safe water) capture essential human needs, while others (e.g., fruit availability) reflect dimensions of dietary quality that become more relevant once basic needs are met. A short discussion on alternative weighting or tiered prioritization would help clarify how differences in indicator weights might influence overall results and interpretation.
b. Lines 363-284: The many nuances raised in this paragraph call into question the rigor and interpretability of the comprehensive score. You note that uneven data completeness may not systematically bias results in one direction or the other, which also means there is no way to calibrate our understanding of how missing data is likely to influence the score. Is there a minimum set of indicators with consistent country coverage that could be used for a more standardized cross-country comparison? That would be a much better way to approach a comprehensive score (assuming such a set of indicators existed).

2) Choice of indicator metrics
a. Food systems emissions indicator (kt CO_2eq): Is this gross or net emissions (incl. removals/sinks)? National targets are often set as economy-wide net-zero targets— a complete "zeroing out" of gross food systems emissions is not usually the policy target. I recommend presenting intensity metrics (per capita / per GDP) to better reflect decarbonization pathways within broader net-zero commitments, with a clear statement on boundaries and implications. This affects interpretations around Lines 286, 312, 358, etc.
b. % share of agriculture in GDP. You note in lines 604-606 that there is no specific desirable value for this (I agree). A high share of agriculture in GDP can signal lagging economic development rather than more resilient food systems, and the percentage can be easily confounded by strong growth in other sectors— despite improvements in agriculture. I therefore strongly recommend considering agricultural value-added or production value per worker or per hectare metrics.

3) Historical trend estimation
a. Line 644-645: Can you provide the year range of latest values somewhere? A two-point CAGR is fragile— especially if the endpoint lands in a year with a shock, such as Covid-19 in 2020/2021. Temporal misalignment could significantly alter results and limit the rigor of cross-country comparisons.
b. Why not average the first and last 2-3 years of values for each country instead of taking just two data points as first and last year? This would increase sample size and potentially decrease noise.

Minor Comments:

1) Abstract
a. Consider highlighting indicators where intervention is most urgently needed—i.e., which indicators are most off track, and where (and, conversely, where the greatest progress has occurred).
b. I find the regional analysis too geographically broad to be meaningful. I recommend either including notable subregional examples (e.g., South Asia, Western Europe), or spotlighting a few large, illustrative countries to convey important trends/insights you want to highlight.
c. The point that countries tend to be closer to meeting peer-performance benchmarks within their regions than to global targets is not, on its own, a particularly novel or actionable finding—it largely reflects how the benchmarks are defined. Recommend removing in favor of more important insights.
d. You could highlight the fact that many countries have plateaued or are backsliding (e.g., lines 430-432). It's a strong call to action.

2) Conflating benchmarks and targets (i.e., benchmarks as targets)

Please make the distinction between benchmarks (attainable peer performance) and targets (rights-based or needs-based thresholds) clear early and often in the paper to prevent benchmarks from being interpreted as substitutes for targets. For example, low-income countries should not be normalized to lower aspirations. This is especially important where benchmark values themselves fall short of acceptable thresholds. There is some discussion on this in Lines 462 – 466, but I recommend clarifying these points earlier in the text. Perhaps you can flag which indicators you believe should not be benchmarked at all, as you note in lines 465-466.

3) Yield and Productivity Analysis

Line 618-624: I disagree that there is no meaningful way to track yields as part of this effort. Developing methods to monitor the gap between potential yields and actual yields of major commodities (e.g., rice, maize, wheat, milk) would provide immense value as an indicator of food systems progress. Given that some commodity-specific indicators are used in this study (e.g., emissions intensity of milk), there is already precedent.

Global targets may be inappropriate, but one could easily imagine yield (or yield gap) benchmarks set across agroecological zones— with spatially weighted national averages— to track productivity across similar countries over time. It's outside the scope of this paper to do this work, but crop and livestock productivity are essential food system indicators to include in any holistic food systems assessment.

4) Extended Data Table 1: Can you add a column listing the primary data source for each indicator in addition to the sources of the target? It would be helpful to see the primary sources for all indicators in one place (either here or somewhere else).

(Remarks on code availability)
N/A

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the efforts taken by the author team to address my concerns. I am happy with the modifications and clarifications made in response to my comments and those of the other reviewers. I have no further comments at this time.

(Remarks on code availability)
N/A

Reviewer #2

(Remarks to the Author)
I am disappointed with the approach the authors have taken to rebutting in many cases the comments rather than

considering and addressing them more fully. There are broadly two major issues that must be addressed prior to publication. Firstly, revise the food system emissions indicator, and secondly, revise the use of figures/tables to support findings.

In my view the most serious issues relate to the food system emissions indicator. This is an important indicator but currently is misleading and unsuitable. I made multiple comments in the last round of reviews on this topic. While, before, it was clear there was an issue with it the reason for this issue was less obvious as what the authors had done was more opaque. However, from the rebuttal comments it is now clear the food system emissions indicator is a country level absolute value, rather than per capita. Why the authors have chosen this is unclear to me.

The authors have added to the text "regional aggregates for food system emissions are influenced by countries with large agricultural sectors (e.g., China and India in Asia, United States and Brazil in Americas, Democratic Republic of the Congo in Africa);, these values are proportional to agricultural scale and population". While correct, this could be misread (as I initially did) to be in 'proportion to'. Something like "Food system emissions total scale with population and agricultural production, so the regional aggregates are heavily influenced by country size", would be less likely to be misread. However, there still appears to be no attempt to justify this choice.

This is extremely misleading and different to all the other indicators that are rates or percentages. This indicator will therefore be highly correlated to population size, and much of the analysis using it is therefore invalid (e.g. regions with countries with small populations would tend to lower regional benchmarks). For example, how should taking a mean of those values be interpreted?

The broken nature of this indicator can be seen in many ways. However, let's take a thought experiment, imagine a situation when a country splits into 2 or more countries (or the reverse). If the larger country was homogeneous with regards to the food system, country metrics for diet, etc. would be unaffected – the new and old countries would have the same values. Great, that makes sense, no progress on the food system has been made with such a change. However, what happens to the current food system emissions indicator, the absolute value gets divided up across the new countries meaning all the new countries have a much lower value, improving their scores relative to regional and global values. How does that make any sense?

While such splitting and combining of countries does happen (USSR, Sudan, Germany, Czechoslovakia), it's not common. However, I use this as a way to illustrate how country size is distorting the metric, making the comparisons/benchmarking applied in the paper invalid for this metric.

One of the rebuttals was that this is what the expert group told us. This is not particularly convincing as an argument on its own. If the work was being presented as an expert group's view on a topic (including what indicators are important) then that would be appropriate. However, it is clear that this is not the case here. The intention is clearly to establish a set of indicators and targets, rather than to report on a viewpoint of the group. Additionally, the having a 'food system emissions indicator' is absolutely right and any expert group would have something along those lines. However, I very much doubt that for most the intention was to report the absolute quantity of CO2e emissions from each country, as opposed to an intensity of emission or per capita emissions, for those countries. The rebuttal has now clarified that the indicator is gross, total food system emissions in kT CO2eq (i.e. not per capita). I therefore suspect that either during or after the stakeholder work there has been some misunderstanding or miscommunication about the intended construct. This is exactly why I previously asked about units, as that seemed a more likely error than taking such an obviously flawed metric.

Can the authors clarify whether consumption vs production emissions have been used. I hope that a consumption basis has been used. If not, then the value of this metric is again in doubt. Food system emissions have been decreasing in Europe (SI Table 8), great, but is this just off shoring of production? More clarity on this would be highly beneficial. Specifically, please state the system boundaries of the FAOSTAT "food system emissions" measure used here: does it include land-use change emissions, upstream supply chain emissions (e.g. fertiliser manufacture, energy use), and is it strictly territorial/production-based or adjusted to a consumption basis (e.g. trade-embedded emissions)?

Methane. There is no separate mention of methane in the whole piece (outside of that implied in CO2eq). Given the importance of Methane and the key role of the food system in emissions of it (from enteric fermentation, rice paddies and food waste decomposition), and the international agreements on reduction (Global Methane Pledge), this seems a strange omission. For example, lower methane emissions from shifts towards plant-based diets (or at least less ruminant-based) would be an important to capture in the metrics.

Another serious issue is I'm having trouble getting a line of sight between claims in the text and the cited figure/table. This may have been caused by the removal of the problematic figure (all reviewers raised concerns about this figure). I'm not convinced removing it, rather than addressing the concerns was the correct action.

The text (from L290) now reads "Results show that current rates of progress are insufficient to meet any global target or global benchmark (in cases where indicators lack a target) across most countries by 2030 (Fig. 1 and Supplementary Tables 8 and 9)." The 'current rates of progress' are not shown in Figure 1, so this statement requires the reader to explore in detail and compare values from SI table 8 and 9 and Fig 1. Actually, I'm not even sure if it would be possible just with reference to those figures and tables alone to make any such comment since figure 1 is just the number of countries. This is a prominent claim from the work (first sentence of 'Historical Progress' subsection), but cannot be seen with the data presented in the main manuscript and is hard even using SI.

The next sentence on ‘progress’ (from line 294) is also problematic: “Most countries show marginal, stagnating, or regressing progress in meeting the target or global benchmark. Moreover, if countries do not improve their rates of change, most are unlikely to meet these targets and global benchmarks even by 2050 (Supplementary Fig. 3).“ However, Supplementary Figure 3 is “Number of Countries Meeting the Global Target or Global Benchmark by 2050, by Challenge Area and Theme”, which does not appear to substantiate this claim.

In fact, while there is method text and Extended Data Table 3 in the main text giving the method for the progress assessment (categorising into stagnation, regression, etc.) I can’t find where the results of this are shown. Probably buried in one of the extensive supplementary files…. Found it eventually, it is SI Table 9. However, I don’t believe this is referred to from the main text.

These inconsistencies, as I raised generically before, are concerning. My suggestion is to think through the key messages and ensure that the figures and tables in the main text support them. As highlighted above I do not feel this is the case currently. Additionally, where statements are made (particularly those not supported by the main figures/tables) make sure an appropriate supplementary figure/table is provided that clearly supports them.

This paper claims ‘no comprehensive targets have been established against which progress can be measured’. However, the food system transformation chapter for UNEP’s Global Environmental Outlook-7, published an extensive set of targets across the food system covering environmental, climate and human health, based on international commitments as done here. Quite likely there are others too. Certainly, this manuscript takes what has been done previously further by operationalising them, but both some reference to previous work and nuance about what is novel would be beneficial.

I don’t find the figures 2 and 3 particularly helpful or informative. The calculations of the required change rate are fine, however what dominates (visually) is the colouring. The reader is implicitly encouraged with this heat map style presentation to take the message the green is easy and red is difficult. However, the thresholds for this colouring are entirely uniform (e.g. <5% green, >15% red). This takes no account of the relatively difficulty of changing these metrics or the regional/country context. At the very least this (currently implicit) assumption should be highlighted and discussed.

The response to my previous comment on why not index (marked Comment 4 in the rebuttal) is not great. In fact, the arguments given are exactly why an index approach would help. For one thing the authors appear to be suggesting comparing the numeric value of fruit availability in grams/capita/day with food system emissions in kt CO2eq. This is just weird and wrong on many levels, and not what you are doing.

Re: previous comment 3. The revisions make it clearer, but is it still not 100% accurate. The reason the current version is inaccurate is the x-axis label says ‘Value (log scale)”, but with the values shown it is actually log10(value), it’s not the same. Why not change (as I previous suggested) the axis value labels, e.g. to 10,000 or (10^4), rather than 4, this is clearer, easier to understand and does not require the explanation in the caption.

While I’m pleased that the equations have now been corrected, I’m not sure that these are needed. Most readers are likely to know what the annual compound formula is and don’t need reminded. A reasonable compromise would be to move them to supplementary.

I would encourage the authors return to the previous set of review comments and consider them more fully, with less of an apparently desire to find some means of rebuttal, but rather as a way to identify areas where misunderstandings have occurred and therefore are likely to occur again.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I thank the authors for their thorough and thoughtful responses to my comments. The revised manuscript is substantially stronger and I am satisfied with the new version.

The removal of the comprehensive score was a sound decision, and addresses many of my concerns with the first version of the manuscript. This greatly improves transparency and allows readers to engage more directly with individual indicator patterns.

The distinction between internationally agreed targets and empirically-derived peer performance benchmarks is now more clearly articulated early in the text, which will help prevent misinterpretation. The additional focus on "spotlight countries" opens the door to more nuanced analysis beyond the regional findings.

I accept the authors' explanation regarding the methodological constraints on moving averages and conducting a yield gap analysis. However, I recommend that the authors incorporate their stated next steps directly into the manuscript text, such as their plans to update indicators to include agricultural value added per worker, intensity metrics for emissions (per capita or GDP), and more granular yield indicators in future iterations of the framework. Explicitly noting these planned refinements would help document the trajectory of this work while signaling to readers that this is an evolving, living framework.

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
There have been major improvements in this manuscript, but I feel there are still issues.

Comment 1: Food-system emissions indicator
I appreciate that the authors have now acknowledged the central problem with using absolute production-based emissions for cross-country benchmarking and have introduced a baseline-normalised indicator for performance analysis. This is a substantial improvement and goes a long way towards addressing my principal concern.

However, while the normalisation of the emissions indicator substantially addresses my main benchmarking concern, the production-based nature of the indicator remains an important limitation that should be more explicitly discussed. In particular, reductions in territorial agri-food emissions may partly reflect offshoring of production rather than genuine reductions in emissions associated with food consumed by that population. I therefore encourage the authors to add a clearer caveat on interpretation where they discuss progress in emissions reduction.

Less importantly, I see these are described as “complementary indicator” – I’m not sure what is intended by the ‘complementary’ here. These seem core to the intention of the effort here.

Comment 2: Production versus consumption basis
As noted above, I consider this an issue that needs to be discussed more fully in the manuscript. The accounting approach taken is now much clearer, but the implications of using a production-based rather than consumption-based measure are still not sufficiently discussed. Given the importance of this distinction, I think this should be explicitly addressed.

Comment 3: Methane
I appreciate the additional text, but I do not think this point is fully addressed. My concern was not simply whether methane is implicitly included within total CO2e (which is of course well understood), but whether the manuscript gives sufficient prominence to methane as a distinct and policy-relevant component of food-system emissions.

Given the importance of methane in food systems and current international policy attention, I think the manuscript should discuss methane more explicitly as a distinct and policy-relevant component of food-system emissions, rather than only subsuming it within total CO2e. At minimum, the Discussion should acknowledge that different gases matter in different ways for mitigation strategy, and that future refinement of the framework could usefully distinguish key gases, particularly methane and nitrous oxide, where data permit.

Comment 4: Line of sight between claims and figures/tables
No further comment.

Comment 5
No further comment.

Comment 6: Novelty / prior work
No further comment.

Comment 7: Figures 2 and 3
I do not think my original concern has been addressed here.

My point was not that the authors must adopt indicator-specific or context-specific thresholds, but that the current heatmap-style presentation strongly encourages the reader to interpret green as relatively easy/attainable and red as relatively difficult/unattainable. Given that the thresholds are uniform across all indicators and contexts, that interpretive risk should be acknowledged clearly.

The rebuttal states that the colours reflect the magnitude of required change rather than feasibility or difficulty, but this distinction still needs to be made explicit in the manuscript itself, ideally in the figure captions and/or immediately adjacent text. The sentence added in the Discussion about local priorities (‘urgency’) does not address this concern, and so not sure I understand why this is being used as a rebuttal in this case.

Comment 8: Scaling / axis presentation
The labels and what is plotted is now accurate, which is an improvement.

However, I continue feel that an index/normalised plotting would be far clearer and preferrable. In this approach readers appear to be asked to implicitly compare between indictors, which is invalid due to the differences in units. If these were

figures in the main paper it would be important to deal with this, but as SI figures it is tolerable.

Comment 9: Equations in the main text
I note the authors' justification. I disagree, but do not consider this a major issue and will leave this to the editors and authors' judgement.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I appreciate the thoughtful revision to the emissions indicator, and for all the efforts of the authors to respond to this last round of comments. The new baseline normalization approach to tracking food system emissions offers a reasonable alternative to the per-capita or intensity-based metrics I originally suggested.

It is not necessary to make further edits at this point, but it is worth noting that the baseline normalization may advantage countries with high historical agri-food system emissions baselines, which will have more room to demonstrate percentage declines. Also countries that reduce agricultural production will appear to show progress even without efficiency gains.

However these are reasonable tradeoffs given the interpretive challenges of the alternative approaches. Worth noting nonetheless.

I am satisfied with the revision and support publication.

(Remarks on code availability)