The authors have calculated a sustainability index based on identified SDG indicators. While I overall see the importance to do something like this and agree with their general methodological approach, I see a number of weaknesses especially in the structure and writing of the paper.

- I would not include the modelling approach in this paper. According to your project description, you have these two objectives to 1) do a retrospective evaluation and 2) model the future sustainability. Accordingly, I would also publish two papers for each objective but not mix them up. Without results but just a methodological approach, the second part seems lost otherwise.

We would like to express our thanks to the reviewer for the useful comments. We agree with the suggestions about paper structure. We will just mention the future sustainability assessment in intro and conclusions (dropping Section 5) and we will extend the retrospective evaluation of SDGs, where necessary, to clarify the message.

- I am missing a sensitivity analysis for your index. You could for example vary the weighting schemes or use a different aggregation method.

Thank you for the suggestion, we performed an index score sensitivity analysis and a ranking robustness check (Table AII1). The composite index score sensitivity and the ranking robustness are obtained varying the fuzzy measures of the three pillars of sustainability; we simulate thousands of fuzzy measures sets according to the variability obtained from the experts’ survey.

In Table 1 we report the 2.5% and 97.5% centiles for both the composite score and the ranking position. The composite index sensitivity of i-th country is defined:

\[ \text{sensitivity}_i = \begin{cases} \text{low}, & \text{for } \sigma_i < \text{Mean}(\sigma) - \text{St. dev}(\sigma) \\ \text{medium}, & \text{for } |\text{Mean}(\sigma) - \sigma_i| \leq \text{St. dev}(\sigma) \\ \text{high}, & \text{for } \sigma_i > \text{Mean}(\sigma) + \text{St. dev}(\sigma) \end{cases} \]

where \( \sigma_i \) represents the standard deviation of the composite index score for i-th country.

Similarly, the ranking robustness of i-th country is defined:

\[ \text{robustness}_i = \begin{cases} \text{low}, & \text{for } \sigma_i > \text{Mean}(\sigma) + \text{St. dev}(\sigma) \\ \text{medium}, & \text{for } |\text{Mean}(\sigma) - \sigma_i| \leq \text{St. dev}(\sigma) \\ \text{high}, & \text{for } \sigma_i < \text{Mean}(\sigma) - \text{St. dev}(\sigma) \end{cases} \]

- I think it is also necessary to normalize your pillar indicators before calculating your overall index since you will otherwise have an inherent weighting scheme due to different value ranges.

It is not necessary to normalize the pillars in order to compute the overall index because all the underlying indicators were brought into the same measure unit (interval 0-1) during the benchmarking and normalization procedure. Consequently also the aggregate indices, by SDG and by pillar, remains in the same interval 0-1.

- The paper is very light on introduction and discussion as well as literature supporting the statements.

Thank you for pointing it out. We are updating the introduction following the Agenda 2030 adoption (September 2015), including the process under Inter-Agency Expert Group (identification of the most appropriate indicators across UN targets) and High-Level Political
Forum (experience and practice exchange at political and governance level). We will also compare our results to the SDG Index by SDSN-Bertelsmann Stiftung, to our knowledge the only other index available measuring the distance from achieving SDGs.

- To give the publication a bit more scientific value it would be also good to evaluate the correspondence of the three pillars against other indicators that are not considered in the SDG indicator set but still important e.g. in terms of outsourcing of environmental costs

The set of selected indicators is highly constrained by compliance to SDG targets, data availability for all considered countries and possibility to project (empirical evidences linking the selected indicator to macroeconomic variables in our macroeconomic model). Although the latter constraint is not relevant for the current paper, it is important for the coherence of the entire project.

Comparing the scores of the three pillars with other indicators, e.g. outsourcing of environmental costs, is quite risky. First, I’m not aware if there is a measure (indicator) of the outsourcing of environmental costs. Secondly, we would need data for this indicator for all the 138 considered countries. Secondly, we would need to normalize this indicator, i.e. information about sustainable and unsustainable level for this indicator. Thirdly, I would be in doubt to compare the score of “outsourcing of environmental costs” to the one of the environmental pillar because the pillar includes several dimensions that are not related to the outsourcing of environmental costs, e.g. the share of protected marine areas. Even more daring would be to compare the “outsourcing of environmental costs” to the economic and social pillar.

- Abstract: I find it a bit weird to mention the project in the abstract. I think a project can go into the acknowledgements but should not be part of an abstract of a scientific publication. Furthermore, I think the abstract should focus on what you did in this paper and the results and not on what you want to do in the future. In the end the abstract should put the results in a broader context or discuss their implications, which is completely missing here.

We agree with the reviewer that mentioning the project in the abstract is a bit weird. We will drop this reference. Our aim was to present the APPS framework in its two components: the retrospective and the prospective, but this is not the right place where to do it. We will focus to the retrospective analysis and discuss the implications.

- Introduction

  o Page 1: why future tense?

The introduction was a bit outdated; we will update it to the correct tense and with latest development of literature.

  o Page 1: I think a few sentences introducing the reader to the whole discussion about SDGs and MDGs would be good, especially since you are talking about “transition” in your title

We surely can mention the MDGs as a landmark in Sustainable Development literature. The word “transition” in our title is more related to the transition between the 2030 Agenda approval and its achievement in 2030. Because measuring the current distance from the target allows to better design the transition and direct investments where are more needed.

  o Page 1: why obviously at national level? Some might argue that sustainability can only be achieved at local level
Certainly sustainability has to be achieved at national, regional and sub-regional levels as Agenda 2030 requires. But it is also true that Governments are entitled to set “its own national targets guided by the global level of ambition” of SDG targets “but taking into account national circumstances. Each Government will also decide how these aspirational and global targets should be incorporated into national planning processes, policies and strategies” (UN 2015). Therefore countries have a recognised role in implementing Agenda 2030. The other reason behind our focus at national level is data availability. Anyway we can add a sentence to support our choice to work at national level.


- Data screening, collection and organization

o Page 1: “Each country will choose the national SDG indicators that are best suited to track its own progress towards sustainable development.” Really? If that is/were true wouldn’t it invite countries to choose those indicators which makes them look better?

We agree that the phrasing of this sentence is misleading and we will revise it. We wanted to highlight that selection process of SDG indicators is participatory “The Commission requested the Inter-Agency and Expert Group to take into account the specific proposals for refinements of the indicators made by member States during the discussion” (UN IEAG, 2017). Anyway, the Inter-Agency and Expert Group has the mandate to define the SDG indicator framework.


- Data screening, collection and organization

o Page 3: The first three paragraphs should rather go into the introduction but not in the methods

Thank you for the suggestion, we will move the paragraphs to the introduction.
Page 3, paragraph 5: You need to be more detailed, e.g. what is the geographical area? The same continent? What does “wen available and reliable” mean?

_We will try to clarify this sentence. The missing data replacement procedure depends on the used database and on the data availability. For example WDI, provide mean values for regional aggregates (more detailed than continents) e.g. Latin America & Caribbean and East Asia and Pacific. These data were used to replace the missing data of countries part of that specific regional aggregate._

Page 3: I would put the sentence about infeasibility of trend analyses in the discussion section

_Thank you for the suggestion. In this section we will just mention that the index is relative to the period 2013-2014 and we will leave to the discussion section some considerations about the impossibility to have a time series of indicators._

Page 4 and others: Do not start a sentence with Table 1 or Figure 1, but rather make a statement and put the reference in brackets behind it.

_Thank you for pointing it out. We will follow your advice._

Page 5: last sentence is unnecessary

_We believe this sentence can help the readers to follow what comes next in the paper; but we can shorten it._

- Benchmarking, normalization and aggregation

Page 6: it is true, that composite indices have a lot of advantages but also weaknesses. You should mention both, but maybe not in the methods part but rather in the intro or the discussion. I think the whole first paragraph should go into the intro or discussion but not the methods and be linked with other literature

_Thank you for the suggestion. We moved the paragraph to the discussion session._

Page 6: first sentence of second paragraph sounds wrong

_You are right, there is a typo: “get” should be replaced by “go”._

Page 6: “First, by considering the different dimensions of sustainability from the indicators listed in Table 1. Second, by building an overall composite index t summarizes the three dimensions.”

_These are not sentences_  

_Thank you, we will rephrase these lines._

Page 6: better “based on the selected indicators (table 1).”

_You are right, this rephrasing is clarifying._

Page 7: Indicators are typically ratios? Where do you get this from? They can be but there are lots of indicators that are not ratios. Normalization is a standard step when creating an index since they all have different ranges. This has nothing to do with an indicator being a ratio or not.

_You are perfectly right: indicators are not typically ratios, and can have many different measure units. It is true that in comparing country performance, we often prefer a ratio indicator that_
helps contextualising our measure, e.g. GDP per capita or renewable electricity share. When the ratios are shares, the normalization procedure is straightforward.

However, we completely rephrased the paragraph that was misleading in its original formulation.

o Page 7: The whole second paragraph revers to inverse normalization. I think you can expect the reader to understand what and why you did it from one sentence e.g. “Some indicators were inverted during the normalization process, since higher values express less favorable conditions in terms of sustainability.” This is just one example but the point is that this is also a standard step for composite index creation.

Thank you for the suggestion.

o Figure 1 is unnecessary

We dropped it.

o Page 6: Equation: iff??? Do you mean if? Furthermore, if you put equations there, you have to explain the variables.

We replaced “iff” with “for” and we briefly described the equations in the text.

o Page 7: “Defining and for all indicators is a hard task and possibly the most critical of the present analysis.” Why that? I assume that the variables refer to the minimum and maximum value of your dataset. The difficulty lies, as I understand, in your aim to monitor the progress over time. If you use the minimum and maximum value simply from your dataset for each year, it would not be possible to monitor any progress, since the normalized values always change according to the maximum and minimum value of the respective year. Is that correct? Your explanation is a bit weak on this. Since this is a real problem with composite indices, when monitoring progress over time, which is also not very often discussed, I would suggest to put a bit more effort into explaining this.

We are not using minimum and maximum values of indicators in normalising procedure, but we are opting for a benchmarking procedure which is very appropriate especially when there are some aspirational targets (SDG targets) to aim to. For example, the sustainable level for SDG 1 is 0.5 (close to zero poverty target) that is the figure characterising current statistics on poverty prevalence in developed countries. When SDG targets are less precise, we used as benchmark some EU target e.g. the 3% of gross domestic expenditure on R&D as share of GDP for SDG 9b. However, we agree with you that the explanation on this matter is quite weak and we will try to improve it.

o Page 7: “For this reason, benchmarks for sustainable/unsustainable levels have been defined for each indicator relying on specific targets set by SDGs, EU best practices (especially in environmental dimension), scientific literature, as well as the observed data.” This info is a bit thin. Add an extra column in the table and explain your source.

Thank you for the suggestion. We will try to add a column specifying the source for our benchmarks.

o Page 7, Figure 2: You don’t need 2 tables showing your indicators and another figure. Consider having just one table or one figure and a table or maximum two tables.
Figure 2 was introduced to explain the two step aggregation procedure. The first aggregation step is to indicators in each pillar, the second aggregate the three pillars into the APPS index. Anyway, we can easily drop the Figure and simply describe the process.

Page 9: It sounds as if you compared two different aggregation methods, but you basically just used two different methods to create your sub-indices and then to create your overall index. Maybe you can write it that way.

Thank you for the suggestion, we rephrased the paragraph.

Page 9: “For lack of space, we do not discuss here in detail the methodology behind fuzzy measures, the Choquet integral and fuzzy measure elicitation.” You have to explain the principles behind it. There is so much redundant and unnecessary information you can delete from the paper, but here you have to explain at least a little bit what you did.

We added a paragraph about Choquet integral. In particular, we highlighted that this approach is able to relax the preferential independence among indicators assumption (common to many composite indices) and hence to model potential interactions (ranging from redundancies to synergies) that may exist among indicators.

Page 9: So you did not use any weighting scheme for your indicators in the pillars but for the main pillars? I find it a bit difficult to assume that the environmental dimension is less or more important than the social dimension.

For aggregating indicators in each pillar, the weighting scheme is the arithmetic average. The Choquet integral is instead used to aggregate the three sustainability pillars into the APPS index. The weights used in this case reflect the preferences of 23 international experts from Farnia and Giove (2015) and Campagnolo et al. (2017).

Page 9: “A country is defined as sustainable whenever, to a certain extent, both its environmental and social dimensions are jointly satisfied and, to a lesser extent, when both its social and economic dimensions are jointly fulfilled.”?? I don’t get this sentence.

The sentence tries to explain the effects of weights used to aggregate pillars. Given that the weights for the social and environmental pillar are higher than the weight of economic pillar, having higher scores on social and environmental indicators matters more than performing good in economic indicators. The sentence is anyway misleading considering that all three dimensions are important for sustainability. We will rephrase it.

Page 10: The Möbius set in Table 3 models the above definition for all the subsets – limited to cardinality two at maximum – that can be formed from the set % = & ', (*, &)*! containing the three pillars.” Sentence wrong

We corrected the sentence; there was a problem in word visualization.

- Assessing SDGs:

Page 10: You can delete the whole first paragraph
We agree, it is not adding any information. We will drop it.

o Page 10: I find it always a bit difficult to say in a results section “unsurprisingly” or “as expected”. If you already knew the outcome, what is the point in doing it? Of course you have an expectation but this depends very much on the information you have in advance. I would stick to reporting simply your results. In the discussion you can say something whether it was surprising or not.

Thanks, we agree and will remove our judgments from result description.

o Page 11, Fig. 3 and also other figures: Caption needs more detail

We will rephrase the caption as “Country performance in the Economic pillar (0 fully unsustainable, 1 fully sustainable)”.

o Page 11, Fig. 4: use letters for sub-figures. Furthermore, in the caption you are saying you are showing normalized indicators, which you also refer to in the text but in the graph you have the SDGs. This makes it a bit confusing for the reader to follow

We will add letters for sub-figures, thank you for suggesting it. The caption is correct: the two figures show the top and bottom performers in the economic pillar displaying their performance in all (normalized) indicators belonging to the economic pillar. For example, SDG 8a coincides with “GDP per capita growth” indicator and SDG 8b stands for “GDP per person employed (PPP)” (the correspondences between SDGs and indicators can be found in Table 1)

o Page 12: “The second sustainability dimension” Delete second or sustainability, otherwise it sounds as if there are two sustainability dimensions

In general, we consider three sustainability dimensions: the economic, the social and the environmental one. In this sentence, the “second sustainability dimension” means the social one. However, we can rephrase the sentence to make it clearer.

o Page 15: Header sound wrong. Maybe The multi-dimensional composite index of Sustainability

Thank you, this header sounds more correct.

o Page 15: Footnote is unnecessary

It will be dropped in the reorganization of the section.

o Page 16: Footnote 13: Each pillar covers most countries

We will rephrase the footnote, but it is important to underline that the countries excluded from the index have some missing data in one or more pillar.

o Figure 11: Put the text below the graph in the caption. Delete blue outer line of points

We will add the description to the caption, and correct the blue outer.

o Why do you show only environmental vs economic pillar? What is the correlation coefficient

If you refer to Figure 11, this shows the correlation between the economic pillar and the sustainability index. Maybe plotting the correlation economic (or social) and environmental pillar could be more insightful. From other analysis, economic and social have typically a positive correlation why economic and environment may have a negative correlation (especially in
developing countries). We can add some plots to highlight correlation among pillars and their relative correlation coefficients.

- Why don’t you show stacked barplots for all countries with the different sustainability dimensions? This way you could show for all of them a. the total sustainability value and the shares of the different pillars

We can certainly add a figure with stacked bar plots. I’m not sure all the 45 countries/regional aggregates can fit in the figure; if not, we can analyse only a set of countries.

- Section 5: I would delete most of this section and concentrate on your results from the index. Part of it could go into the discussion as a future outlook but not before you actually discuss your approach and results from the paper

Thank you, we will do as you suggest.