Referee #3

This paper is a timely consideration of a significant issue: the status, quality, timeliness and implications of India's greenhouse emissions data, with a very useful summary of the compoient contributions and recent trends. It makes a valuable contribution and underscores the importance of making data available in a regular manner with sufficient detail to ensure its accuracy and reliability.

Many thanks for your thoughtful and helpful comments.

Specific comments:

Page 1 line 28: Given the size of some of the utility-scale PV plants, the term "small renewables" should be re-considered.

- Changes made: Changed "small renewables" to "variable renewables".

Page 1 lines 29-30: It might be worth including India's reverse auctions and innovations such as round-the-clock tenders in the list of contributory factors in renewables growth.

I'll avoid reference to round-the-clock tenders, since there's some indication that the implementation of these does not reflect the name:
https://economictimes.indiatimes.com/industry/energy/power/round-the-clock-renewable-energy-tenders-worry-developers/articleshow/76012308.cms

- Changes made: Added sentence "Development of variable renewables has been further assisted by the introduction of reverse auctions and the creation of solar parks, among other measures (Bose and Sarkar, 2019)."

Page 1 lines 32-33: In addition to the limitations of the available data listed, it could be noted that official documents are maintained by different ministries and departments, which can also lead to outright inconsistencies between different official sources, such as disagreement between the CEA and the MOSPI Energy Statistics publication concerning the quantity of coal consumed for electricity generation in the three most recent years.

I've taken some time to look at this specific inconsistency. MoSPI's Energy Statistics for coal consumption by the electricity sector (table 6.4 in the 2020 edition) are taken from the Ministry of Coal, and, since 2010, are identical to the numbers in the Coal Directories (table 4.20 in the 2018-19 edition), except for the final year, which comes from the Provisional Coal Statistics. These data represent *despatches of domestic coal to both utility and captive power generators*, not consumption at all, despite the title of both the chapter and table. It seems imported coal used by power stations is included in the 'Others plus import non-coking' column, partly explaining why this column has such large values. The supply data they use do not allow disaggregation of non-coking coal imports by using sector. Nor does this table account for stock changes at power stations. Meanwhile, the CEA data only include utility generation, not captive. So to reconcile the data in these tables one must take the utility despatch data from the Coal Directory (or PCS) and the total coal receipts less imports from CEA. These two are approximately the same, with some residual as is common with comparison of supply and use data from different sources.

The annual Energy Statistics from MoSPI is severely lacking in descriptive text, making it very difficult to determine what the data really mean.

- Changes made: Added a section to the Supplement describing this specific point on coal consumption, and referred to this section in the Introduction as "Furthermore, explanations for data are often lacking in detail, and can conflict across different datasets for reasons that are not immediately apparent (see Supplement: Coal 'consumption')"

Page 3 line 3: The first statement should be qualified to acknowledge that monthly coal consumption figures for power generation are provided by the CEA.

- Changes made: Reworded to "While monthly coal consumption by utility power stations is reported (CEA, various years-b), India does not report sub-annual total coal consumption, and apparent consumption must therefore be calculated using data…"

Page 4 line 22 to page 5 line 7: The discussion of the seasonality of emissions is very important as it is such a strong factor in India's data. Accordingly, some extension of the discussion might be worthwhile, for example, to consider how monsoonal weather affects a) production and supply, hindered by weather affecting logistics, b) demand and consumption, for example decreased construction activity or abrupt decreases in air conditioning load as rains relieve extreme heat conditions that normally occur in May. In addition, the substantial increase in hydroelectric and wind generation that accompanies the monsoon and suppresses coal consumption could be noted here as well as on page 9. The benefit of an extended discussion is that it would guide readers who may wish to analyse seasonal changes in emissions in terms of other economic and meteorological data.

I agree. I've added some of these ideas to the discussion here.

- Changes made: Added the following sentences: "These emissions patterns largely result from the effects of the monsoon's heavy rains, driving a decline in industrial, construction and transportation activities. Coal emissions are also driven down by the displacing effect of higher power generation from both hydropower and wind during the monsoon season."

Page 6 lines 16-18: The apparent omission of power plant coal stockpile changes in the Energy Statistics publication might also be mentioned.

- Changes made: Added sentence: "Furthermore, IEA's data exclude changes of stocks at power stations, which exhibit large swings (SI Fig 10)."

Page 10 lines 12-22: This paragraph considers data revisions and errors. Although it is correctly stated that coal statistics from CIL and SCCL undergo only minor corrections, it might be noted that data from captive power plants and other users are much more sporadic and provided only in summary form. Capturing total coal consumption could benefit from more systematic and timely data on non-CIL and non-SCCL data.

While I would readily have agreed with this statement at the time it was made, I have since discovered that the Ministry of Coal has recently started reporting provisional year-to-date production and offtake explicitly including both captive and other mines at https://coal.nic.in/content/production-and-supplies. I have added a module to my code to check this page and calculate the differences to give monthly production, and will include these in my published dataset. The Internet Archive provided the website in May 2020, allowing the estimation of June's production by difference. I hope the publication of these year-to-date data will continue, but there's never any guarantee.

Moreover, there is also the Monthly Summary to Cabinet (https://coal.nic.in/content/monthly-summary-cabinet), which has included a statement on the production of captive mines since the September 2017 edition. I have added these to my published dataset.

The more general point about having to fill in gaps in available data is made in the preceding paragraph.

- Changes made: none.

Page 11, lines 1-12. The Conclusions are entirely appropriate and relevant. Given the importance of timely and accurate data, and the multiple shortcomings noted in the body of the paper, a useful

addition to this section could be a brief set of key recommendations that could guide efforts to better coordinate, accelerate and improve India's collation and publication of energy and related statistics.

- Changes made: Added paragraph to the Conclusions: "India publishes more energy data than many other developing countries, providing a wealth of information for management, policy analysis and scientific research. Nevertheless, there remains significant room for improvement in the quality of these publications. Possible avenues for such improvement include: (i) Publishing more data in machine-readable formats, rather than just as tables in PDF documents or in web-page tables, (ii) Providing a way for the public and researchers to ask questions about or report errors in data, establishing direct contact with those responsible for the data, to facilitate crowd-sourcing of quality assurance, (iii) Encouraging collaboration in data preparation and presentation across ministries to prevent errors creeping into reports, (iv) providing more documentation of reported data, (v) Reducing use of manual copy-pasting and typing, and automating as much as possible with both automatic and manual quality assurance, (vi) Standardising the use of important terms (e.g. 'consumption') across reports from different departments to prevent confusion, (vii) Making available older, non-electronic reports (e.g. Monthly Abstract of Statistics), online through use of digitisation."