**Reviewer 2**  
  
  - Overall, the authors did well to address the comments from the initial reviewers. However, as outlined below, I believe a few issues still need to be addressed before publication can occur:

In the Methods Section (~ lines 138-149), the manuscript would benefit from additional elaboration as to why only titles, abstracts, and keywords were analyzed (as opposed to full documents)? Was this simply a time/computational constraint? One could envision articles that have a relatively broad description of urban climate change mitigation and GHG accounting in their abstract, but actually get much more into the detail and nuances of aspects of transportation, buildings, waste, and/or urban form in their main body. If an analysis of the full documents is deemed infeasible for this manuscript, the authors should consider positing how analysis of full documents may (or may not) change the final results and implications.

The reviewer makes a good point. We add a footnote to Line 142 raising this issue and referring to a discussion in the SM (section 2, paragraph 4):

The method is applied to abstracts, titles and keywords only. We do not have the comprehensive library coverage to obtain all the documents, and there are major computational barriers to converting pdfs into word documents while preserving text structure (this is a necessity for the analysis; we have identified no tools that do it accurately). Performing the analysis on full texts would likely increase the number of topics (due to the greater quantity and variety of words), potentially revealing more detail on the existing broad set of topics, or new information on topics systematically absent from abstracts (e.g. methods and data). However, this may come at the expense of a manageable and interpretable number of topics.

In the Methods section, related to the automated content analysis (~lines 140-149), the manuscript would benefit from additional clarification as to how the ‘query keywords’ were selected and/or tested. Where they manually chosen by the authors or based on Non-negative Matrix Factorisation? If manually chosen, how was this process accomplished? For example, terms like ‘degree day’ would appear to be relevant and helpful in determining papers related to the topics of “Thermal Comfort” and/or “Heating & Cooling Systems.” Yet, ‘degree day’ does not appear to have been one of the key search terms. The point of this example is not say that ‘degree day’ should have been added to the keywords – I acknowledge that there is no shortage of words that could have been added to the analysis and a line needs to be drawn somewhere. Instead, it is simply to highlight the need for clarity on how certain words were included while others were excluded. In addressing this issue, some of the text from Sections 2 and/or 3 of the SI might be more beneficial if included in the Methods Section of the main text. At the very least, the Methods section should include a more explicit reference to the relevant sections of the SI.

We agree this basic point requires more elaboration in the text. We are unable to state why certain words may have been included or rejected, as this involved a very large number of terms and was, in the end, based on the authors’ intuition. However, we can provide clarity on the procedure, which followed a number of discrete steps (lines 128-140):

To identify the urban mitigation literature covered by these four domains, we develop a structured search query for the Web of Science (WOS) literature database. We identified an initial set of keywords based on the authors’ expertise, then iteratively refined these through WOS searches and by reviewing random samples of the obtained documents until approximately 90% were deemed relevant (by two authors, independently). The search query includes specific combinations of keywords for each domain (e.g. “bicycle infrastructure provisioning”), as well as more generic strings (“low-carbon transport”). We aim to identify mitigation-relevant studies as well as mitigation-focused studies (i.e. papers on urban congestion policies that do not directly refer to emissions reductions would be included). The full search methodology is detailed in the Supplementary Materials (SM) Section 1. Using the search query we acquire a document set. This is largely comprised of journal papers, but also includes conference proceedings and book chapters. For each document, we obtain the title, abstract, keywords and list of references.

And in the SM text (section 1, paragraph 2):

Several discrete steps were taken to develop the literature search. (1) We first identified four overarching urban topics judged to directly or indirectly contribute to urban climate change mitigation, as widely discussed in the literature: transportation, buildings, waste management, urban form and city-level mitigation plans and policies. (2) Based on the authors’ expertise, we specified a list of keywords, grouped by each topic area. (3) We used the keywords on the WOS platform, identifying more relevant terms, and removing terms that delivered spurious results. (4) We downloaded the titles and abstracts, for each topic assigning random samples of 25 documents to the authors. (5) We developed a simple criteria to rate documents as either ‘relevant’ or ‘not relevant’: “is this document relevant for an assessment of urban demand-side climate mitigation, in the fields of transportation, buildings, waste management, or urban form?” (6) Two authors then reviewed their random samples, which were cross-checked for consistency. (7) We repeated steps 3-6 until approximately 90% of the documents in each topic were deemed relevant. For transparency, the full search query is provided below.

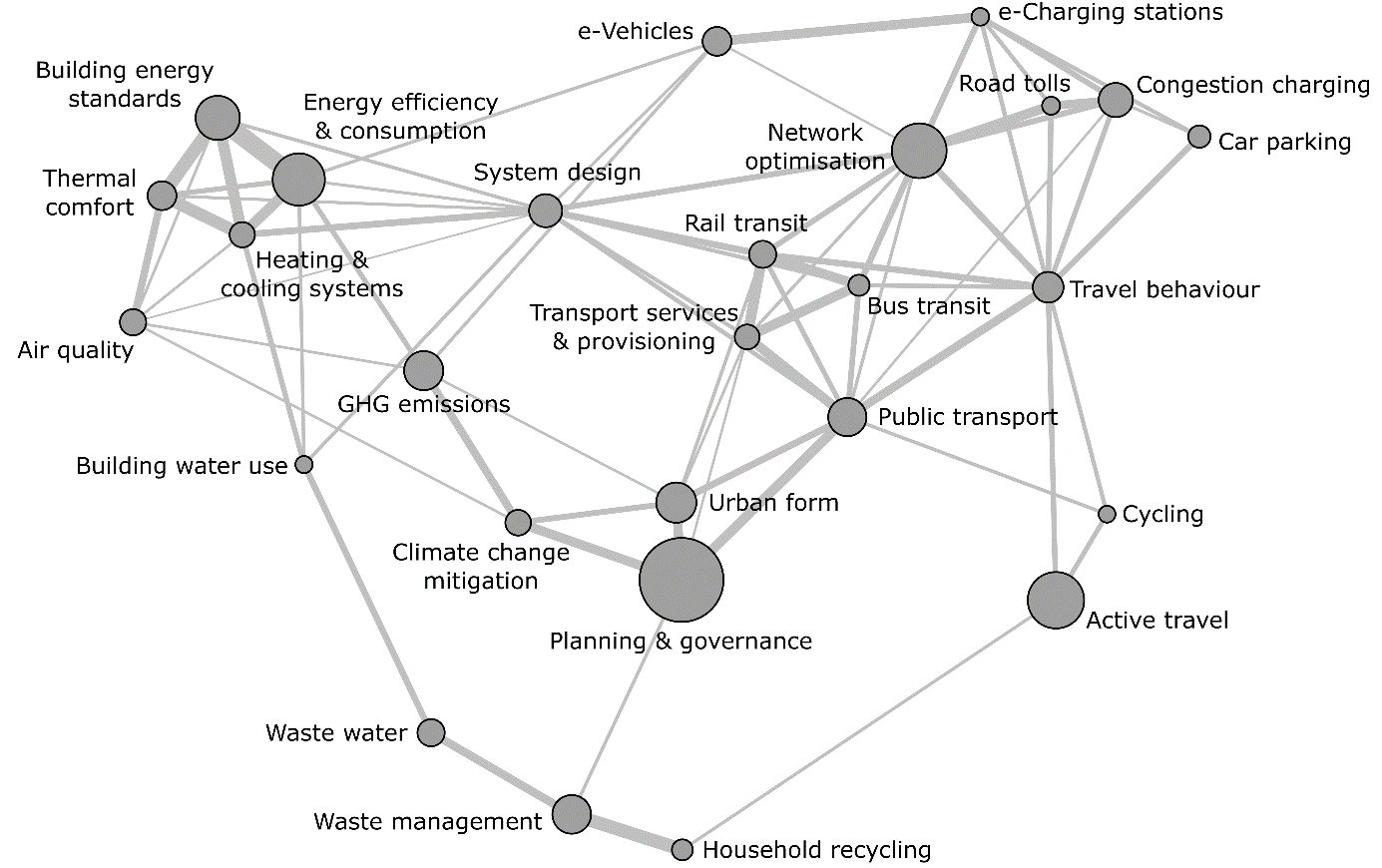
We already acknowledge the restrictive wording and scope of the query in the SM (section 1, paragraph 5):

Our search results in a fairly restrictive selection of papers that is by no means comprehensive, but is sufficient to provide an overview of the major topics of urban climate change mitigation. Obviously, diverging emphasis and wording in the search query could be equally plausible. For example, a broader focus on infrastructure provisioning may be relevant (e.g. electricity access, affordable housing), but this is perhaps better suited to a sustainability themed assessment of the urban literature, rather than our narrower climate change mitigation focus.

The results and caption related to Figure 2 need to be clarified/revisited.

* The topic of “Building Energy Performance” is mentioned in the caption but it does not appear anywhere in the figure. Did the author’s mean to reference “Building Energy Standards” in the caption?
* “Green Infrastructure” is referenced in the caption, but does not appear anywhere in the figure, nor is it listed as one of the 27 core topics in Table 1.
* Does the thickness of each line represent the level of correlation between two topics (i.e., a thicker line indicates a stronger correlation, etc.)? If so, this point should be made more clearly. If not, the clarity and ease of interpretation of the figure would be improved if such a change were made (on the assumption that such a change would be relatively straightforward in the graphical software used by the authors)

Thanks for catching these mistakes and needed clarifications. Here is the revised caption and figure, showing the correlation weights more clearly:



**Figure 2: Correlation structure of urban mitigation topics.** Each node is a topic, scaled by the marginal distribution (see Table 1); each line represents a correlation, based on the co-occurrence of two topics within document abstracts. Lines are scaled from the highest correlation of 0.33 (between building energy standards and Energy efficiency & consumption) to the lowest of 0.025 (between Air quality and System design). The visualisation is generated using the force-directed algorithm ForceAtlas2 in Gephi [39].

For the Methods section, related to analysis of the document references (~lines 150-155) and the results in Figure 3, were there any publications that bridged multiple clusters (e.g. a paper that was almost equally split across Clusters 2 and 3)? If so, the manuscript would benefit from a description of any mechanism(s) used to ultimately categorize a publication in one cluster versus another.

(Max to do)

The authors should ensure that a finalized version of SI is submitted with the revised manuscript. The current version of the SI still includes “tracked changes” and makes interpretation difficult (e.g., Figure 2 appears 3 different times).

Done (we kept tracked changes initially so that the keyword changes in the search query could be easily seen).

It appears that Figure 2 of the SI needs to be updated. It describes 111 references from the IPCC assessment reports, but the revised version of the manuscript includes 129 references from the IPCC assessment reports, correct?

(Max to do)

The manuscript would benefit from further elaboration and clarification related to Figure 3 in the SI.

* What time periods are associated with the ‘Assessment Periods’?

We added the following sentence to the caption:

**Figure 3: Topic scores by IPCC assessment period. Each cell shows the sum of each topic’s scores.** Cells are coloured by the score as a proportion of the sum of all topic scores for documents written during each assessment period. Topics are shown in descending order by percentage growth from AR5 to AR6. These values are given on the right-hand side. As indicated in Figure 1, the assessment periods correspond to 1991-1995 (AR2), 1996-2001 (AR3), 2002-2007 (AR4), 2008-2013 (AR5), 2014- (AR6).

* Are topic scores calculated using Non-Negative Matrix Factorisation? If so, Section 3 of the SI should probably be moved above this figure and explicitly referenced. If not, additional elaboration is needed on how the topic scores were calculated.

As suggested, we have changed the order of the SM (1. Literature search, 2. Topic modelling, 3. Additional figures).

* Clarify what is meant by “Each cell shows the sum of each topic’s scores” in the caption. Does this mean the scores are summed for each Assessment period, summed cumulative across assessment periods, or something different?

 (Max to do)