

1 We thank all the reviewers for their time, valuable and encouraging feedback and recommendations for improvement.
2 Answers to specific comments appear below.

3 **R1:** *“While it is clear that SDM and SDDM generally outperform other models in terms of perplexity, the authors should*
4 *comment more on ideal settings for each approach as there is no clear winner between the two based on experimental*
5 *results.”*

6 SDDM should be superior to SDM when the time-group partition of the data allows for sufficient number of documents
7 per time-group slice to learn local topics. On the EJC dataset, majority of the groups have very few articles leading to
8 lower quality local topics. This observation is briefly summarized in lines 301-302 of the main text - we will emphasize
9 this point more clearly in the revised version.

10 **R2’s comments regarding clarity of the presentation**

11 We appreciate the suggestions for improving the clarity of the paper. If the paper is accepted, the 9th content page
12 (allowed for the camera-ready version) will help us to incorporate these suggestions and improve the flow of the paper.

13 **R3:** *“Elaborate on data set preparation”*

14 Regarding vocabulary sizes, for the Wikipedia corpus, we followed similar procedure for vocabulary truncation as
15 in Online Learning for Latent Dirichlet Allocation (Hoffman et al., 2010; they also analyzed over 3mil Wikipedia
16 articles and truncated vocabulary to 7995 words as stated in their footnote 4). We describe the Wikipedia vocabulary
17 preparation in lines 146-150 of the Supplement and we will add the reference to clarify our choice of the vocabulary.
18 On the EJC dataset we applied relatively standard vocabulary truncation steps, i.e. removing very common (over 99%
19 documents) and rare (under 1% documents) words, removing short and stop words, and stemming. This procedure is
20 described in Supplement section 7.1. To argue that the resulting vocabulary size of 4516 words is appropriate, we may
21 compare to the vocabulary size of 4253 words used by Hoffman et al. (2010) on the 350k documents Nature corpus (see
22 footnote 3 in their paper; we also note that Nature corpus is not public leading to us choosing the EJC corpus instead).

23 **R3:** *“Extend qualitative analysis. For this, the quantitative analysis of hyperparameter influence could be pushed into*
24 *the supplemental.”*

25 We thank the reviewer for the suggestion and we will add additional qualitative examples in the revised version.