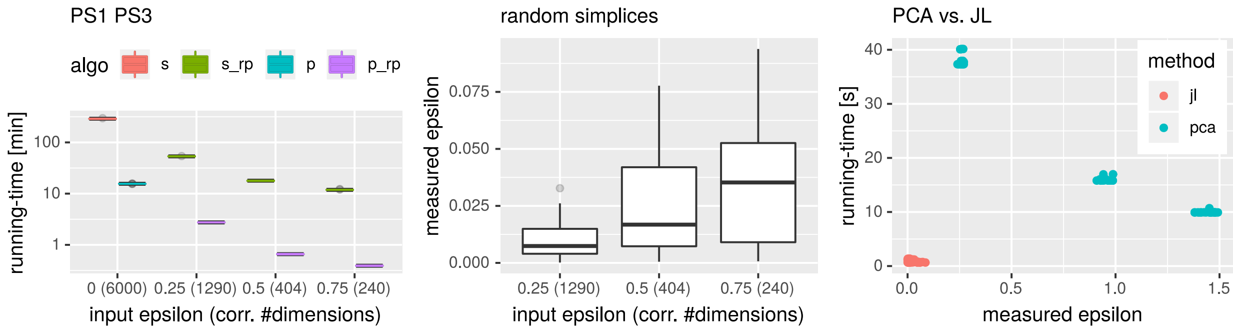

We thank the reviewers for their detailed comments and suggestions, which we have addressed below.

(#1) novelty of sub-sampling: Please note that Thm. 11 and 12 involve a *candidate* set and a *witness* set. While the existence of sublinear witness sets are immediate from Indyk's work, his candidate set is the entire input of size n. Our own arguments lead also to sublinear (even constant) *candidate sets*: In Thm 11 this is mainly by triangle inequality, but in Thm 12 it is a more complicated argument (sketched in the text and detailed in the supplement) assuming a natural distribution of distances around the true median. Meanwhile the experiments show outstanding performance.
(#1)(#3) additive error: We believe that the additive error is necessary. Put two curves $p = p_1p_2, q = q_1q_2$ with Frechet distance 1 whose points are vertices of a rectangle with $\|p_i - q_i\|_2 = 1, \|p_1 - p_2\|_2 = \|q_1 - q_2\|_2 = \gamma$ for *very large* γ. The above distances (esp. $\|p_i - q_i\|_2 = 1$) between points will be distorted by at most $(1 \pm \epsilon)$ so the Frechet distance will also be within $(1 \pm O(\epsilon))$. Insert another point z in the middle of one curve. z has distance $\approx \gamma/2$ to each other point for reasonably large γ. Now JL has its mass concentrated in the interval $(1 \pm \epsilon)$ but inspecting the concentration inequalities nearly half of this mass is between $(1 + \frac{\epsilon}{100})$ and $(1 + \epsilon)$. Thus with reasonably large probability the error will be $> \frac{\epsilon\gamma}{100}$ which is additive since γ is unrelated to the original Fréchet distance of 1.
(#2) abscence of stated ground truth / lack of clustering metric and baseline: An evaluation of the meaningfulness of the k-center objective for a real world problem is out of scope of this work. Our starting point is: given that k-center gives meaningful results, how well does our approximation resemble those results? Thus, from our point of view, there is no abscence of ground truth. We ran the Buchin et al. k-center algorithm to obtain clusters. Those were confirmed to expose a meaningful structure in the data. This acts as our ground truth. Our focus, as stated in **Q3**, lies in a comparison of the algorithms quality with/without embedding. To answer **Q3** we wrote "In about 99.75% of the 400 experiments for the DELTA data set the same center-set was identified..." The exact clusters and centers are available in the supplement.
(#2) decoupling of the contributions of the embedding and the parallelization: In Fig. 2. the random projection and the parallelization speed up the computations by a factor of 10 *each independently*. *Both together* yield a speedup of factor 100. The requested *decoupling* is thus already presented. We will do our best to make his clear. Cf. left plot.
(#2) theoretical gain and quality vs speed: Indeed since the algorithms run in linear time with respect to the dimension d, the gain is a factor of $(4\epsilon^{-2} \log n)/d$. We will add this explicitly to the discussion. It is a good idea and we will add quality vs speed trade-off plots comparing to other baselines like PCA, see reviewer (#3) and right plot.
(#2) further data sets: We will simulate further data to outline the running time performance on high-dimensional data, e.g. from the $(d + 1)$-dimensional simplex for exceptionally large d, see middle plot.
(#3) target dimensions and constants: For the dimensions of the original data we will add a table, see minor comment of (#2). For the target dimensions we indeed use the asymptotic formula with a factor of 4, i.e., $k = 4\epsilon^{-2} \ln n$. We will add this and the resulting values of k for all experiments. An extensive empirical study [1] showed that a factor of 2 almost never fails. Another factor of 2 was added to be absolutely sure corresponding to $\delta = \frac{1}{4}$.
(#3) Fréchet distance and sequence information: Proposition 7 and Lemma 8 can be used to prove an error-bound for the continuous DTW, which we will add. However our focus is on a physical application that involves numerous sensors (≈600, even more in the future). These sensors, though of same type, have varying noise and bias due to aging, radiation, calibration, and manufacturing variations. Our focus lies in finding outliers among time-sequences and thus need measures sensitive to outliers. Sum-based DTW is not applicable since it is sensitive to bias/noise and averages out single heavy outliers. The sequence information does matter, because the notion of an outlier is relative to the single curves noise/bias level. Further, Fréchet distance (already stated in our introduction) induces a linear interpolation. This interpolation is crucial because sensors are sometimes not available for short periods of time (due to radiation, heat etc.). Using a discrete distance measure like the discrete Fréchet or DTW, this could induce an unboundable additive error, which would require us to pre-process the data and search for sampling gaps and is not desirable.
(#3) PCA: Empirically, we have started evaluating the quality and running-time of the PCA vs. the random projection, see right plot. We will also try to obtain theoretical results, which is part of future work.

[1] Suresh Venkatasubramanian and Qiushi Wang. The johnson-lindenstrauss transform: An empirical study. In *Proceedings of ALENEX*, pages 164–173, 2011.