We thank the reviewers for their time, helpful feedback, and advice. We are pleased that overall, reviewers praised the clarity, rigour and contributions of the work. We are encouraged that reviewers acknowledged novelty (R3, R4) and appreciated our work as a principled contribution in the development of machine learning methods for non-Euclidean data (R1, R2). We thank reviewers for their kind words, and hope to address any remaining concerns below.

(R1, R2, R3) Uni-modal distributions. We thank the reviewers for raising this question. The goal of our two experiments with unimodal target distributions is to demonstrate specific pathologies of previously introduced methods. We believe this is most clearly demonstrated by a unimodal distribution at the point (or the limit) of the pathology as it removes additional modelling artefacts that are introduced through more complex distributions. We further evaluate the capacity of our method to model highly multi-modal distributions in our real-world experiments in which we can show substantial improvements to baselines. We will clarify the different purpose of these experiments in the paper.

We agree with (R2) that "the stereographic projection is known to lead to problems in its singular point." Yet, current SOTA methods are actively using such a parametrization (e.g., Gemici et al. (2016)) and we believe it is important to show and evaluate this aspect. Moreover, most manifolds of interest similarly have a non-Euclidean topology, so applying previous *projected* methods on these manifolds would also yield a similar pathology.

(R2, R3) Real-world data. Our experiments on real-world data are motivated from problems in climate and earth science, hence leading to empirical assessments on $\mathbb{S}^2$. We believe these experiments to be informative since they show that our method is a) scalable and b) can fit highly complex and multi-modal distributions more accurately than previous methods. Moreover, please note that our model can straightforwardly be applied on higher dimensional manifolds. We agree with reviewers that our model would be strengthened by an additional experiment on a high dimensional manifold. To achieve this we will run a fourth experiment by first computing hyperbolic multi-dimensional embeddings of WordNet graph data, and then fitting our model to the obtained empirical distribution.

(R4) Optimal transport (OT) and flows. As reminded by (R4), the field of OT is core for the study of flow – known as *transport map* – transforming one distribution into another. Indeed, we develop this connection in Appendix D.1 and show that the dynamical formulation of OT still holds in the manifold setting. Regarding (R4)'s concern that our method "can not model the transformation for a white noise to multi-mode distributions": Our method is theoretically sound as it able to model multi-modal distributions as long as supports are connected (see Cornish et al. (2019, Theorem 2.1)), and is empirically shown to model well multi-modal earth data (cf Table 3 and Figure 6).

(R1, R2) Computational aspects. We thank (R1) for suggesting to expand on the limitations of our method. As reminded by (R2), projections required by our approach can increase the computational cost. With our current implementation, we empirically find that this additional cost amounts to $\sim 20\%$ for the Poincaré ball and $\sim 30\%$ for $\mathbb{S}^2$. This cost can be further reduced by only projecting the output of the ODE solver steps. We thank (R2) for the suggestion and we will rigorously compute and include wall-clock time comparisons in the next draft.

(R1, R2) Empirical comparison to related methods. We thank (R2) for suggesting the mixture of von Mises-Fisher (vMF) baseline. Indeed, we found in our early empirical assessments that high multi-modality (e.g., as occurring in the different the earth datasets) would prevent a mixture of vMF distributions from being a competitive baseline. We will re-run this baseline and include its performance in Table 3.

We also agree with (R1) that comparing our method to Rezende et al. (2020) would indeed be valuable. Unfortunately, the code and necessary experimental details have not yet been released, preventing us from a detailed comparison.

(R2, R4) Additional manifolds. We agree with (R4) that applying the proposed method to more exotic manifolds is an exciting direction. We are currently exploring applications on several Lie groups such as orthogonal or positive definite matrices. As reminded by (R2) we indeed assume that the manifold is known beforehand. We agree that learning the manifold is an exciting task, but indeed a harder one as shown by the field of topological data analysis.

(R2) VAE experiments. We thank (R2) for suggesting to leverage our model in a VAE setting. We agree that this is indeed a promising application of our method (e.g., see also Bose et al. (2020)) and plan to explore this in future work.

(R2) Naive Euclidean method. We thank (R2) for highlighting the potential risk for non-careful readers to compare manifold-valued densities against $\mathbb{R}^D$ valued ones. We updated the introduction to better refer this.

References

Bose, A. J., Smofsky, A., Liao, R., Panangaden, P., and Hamilton, W. L. (2020). Latent Variable Modelling with Hyperbolic Normalizing Flows. *arXiv:2002.06336 [cs, stat]*.

Cornish, R., Caterini, A. L., Deligiannidis, G., and Doucet, A. (2019). Relaxing bijectivity constraints with continuously indexed normalising flows. *arXiv preprint arXiv:1909.13833*.

Gemici, M. C., Rezende, D., and Mohamed, S. (2016). Normalizing Flows on Riemannian Manifolds. *arXiv:1611.02304 [cs, math, stat]*.