We thank the reviewers for their efforts in reviewing and their feedback, considering the paper "well written", "[a technically] clean contribution with a clear advancement" as well as "clear, rigorous, and self-contained". Furthermore, R2 commented that our "methods mixing geometric deformation models and DL may interest the NeurIPS community."

The main reviewer concerns revolved around (see also additional responses below):

1. *Modeling assumptions (R1/R4).* Our model is based on spatio-temporal velocity fields (rather than stationary velocity fields (SVF)), because this allows us to *preserve theoretical properties.* Importantly, no registration approaches exist which model and estimate metrics which jointly deform with space. Our model, containing LDDMM as a special case, is therefore *a first theoretical contribution in this direction, demonstrated on a practical large-scale registration task*.

2. *The overall formulation, contribution and performance of our approach (R1/R4).* Our approach is related to [23] in the sense that we estimate a metric, but we do so in an LDDMM (not SVF) setting. Further, as in our approach (but not for [23]) the metric moves with the deforming space *we derived an entirely new registration model (RDMM).* While our model can be estimated via standard optimization, we use deep learning approaches to predict the metric *and* the initial momenta and show the benefits of such a DL approach (see Fig. 7). We do not use previously computed ground-truth momenta but we backpropagate through the discretized RDMM equations and the network predicting the momenta.

3. *Parts of the writing that were difficult to follow.* (R2/R4). We will simplify and improve our notation, will better explain the proposed optimal mass transport penalty, and will clarify the explanation of our learning approach.

R1: Modeling. *Do time-dependent velocity fields matter?* Yes, from a theoretical point of view (see 1 above). Practically, when very large deformations (e.g., lung motions or large shape changes) are to be modeled. This might not directly be observable in overlap measures, but will impact deformation paths. We will clarify this. *Should the regularizer/prior not be independent from data?* Instead of putting a prior directly on a displacement, velocity, or momentum field, *we put the prior on the metric itself*. This prior on the metric is in the reviewer's sense data independent, but the data drives which one of the metrics is selected in the end.

R1: Technical questions. *How is the velocity field bounded?* Our goal is diffeomorphic registration, but there are, of course, cases (e.g., sliding) where one would not want to use such a model. Velocity fields are bounded by predicting the pre-weights. Weights are then computed via smoothing of these pre-weights which assures that velocity fields stay bounded. *Symmetry loss: does invertibility imply symmetry?* RDMM aims at diffeomorphisms. Hence, we compute a transformation (and its inverse) from source to target space. However, this does not mean that swapping source and target images (A→B vs B→A) *also* swaps the transformations. To encourage this, we add the symmetry loss.

R1: Experiments. *Missing validation test/set discussion.* We use the same approach as the cross-subject experiment of [33]: *training (2,800 pairs), validation (50 pairs) and testing (300 pairs);* all separate. We will clarify this. *Results are similar to the baselines.* Our focus was not on outperforming current state-of-the-art registration approaches. We agree with R1 that *our work should be considered a more theoretical contribution opening up further research* for the estimation of spatially-temporally-varying regularizers. We will clarify this in the final manuscript.

R1: Positioning. *Does it fit in NeurIPS?* We think so. While we apply our model to medical registration, it is much more general. One could easily, for example, modify our model to transform densities. This could then be used for generative models, for example. We will make such connections more explicit.

R2: Technical questions. *Why are there foldings?* Our model assures diffeomorphisms in the continuum. However, as numerical discretization will only approximate this continuum solution, foldings may occur. As the computation of the Jacobian is also discretized, it is itself only approximately correct. We will add this to the discussion.

R4: Technical questions. *How are the initial momenta learned / differences to [23] / value of the DL approach?* The equations underlying RDMM are discretized and this model is appended to the deep networks predicting the initial momenta and pre-weights. Hence no ground-truth momenta are necessary. [23] only considers an SVF model and uses numerical optimization to estimate SVF initial-momenta. Instead, we predict the pre-weights *and* the initial momenta with deep networks, within this new RDMM model which allows for an estimated regularizer to move with the deforming image. Fig. 6 compares various optimization/learning-based and spatially varying/non-varying methods. *Fig. 7 illustrates the benefit of DL, which results in more anatomically meaningful regularizers (init_w (learn b)) than estimation via pairwise optimization (init_w (opt b)), which is much noisier.* *Method will be limited by what it has seen during training.* We agree. In many cases this is desired, as it amounts to a statistical deformation model preventing transformations considered too unusual. We will highlight the possible limitations for capturing unusual deformations.

R4: Experiments, references, others. *The experiments use 160^3 and 200^3 images - is this a realistic?* Our approach could scale to larger images, but we opted for slight downsampling to reduce computational requirements and GPU memory demands. *Spatially-varying method has been studied … in MICCAI'13, MedIA'15.* We will include these references. Note that in these references metric estimation is in a fixed atlas-space; whereas we address general pairwise registration. *Is it really end-to-end training?* It is not end-to-end with respect to inclusion of the affine step. But the non-parametric registration is (given the affine initialization). We will clarify this.