We thank the reviewer for their feedback on our manuscript, and their carefully considered comments.

Our interpretation of the reviewer's concerns and comments is that there is a basic confusion (likely common to both TopREML as proposed in this paper, and to the use of TopKriging) about the conceptual hydrologic model that underpins both approaches and their treatment of network topology.

Our major contribution in TopREML is the application of the REML statistical interpolation method to a spatially-hierarchical system. The conceptual treatment of both hydrology and network topology in TopREML is essentially the same as that proposed within TopKriging.
We therefore primarily focused this manuscript on a geostatisical argument.

We suggest that most of the areas of concern and confusion that the reviewer mentioned can be clarified by adding an introductory section to the TopREML paper that (i) outlines the conceptual hydrologic model that underpins TopREML (and TopKriging), (ii) describes how this conceptual model leads to the treatment of network topology by TopREML and TopKriging, (iii) highlights the conceptual differences between the interpolation schemes in TopREML and TopKriging, and (iv) offers explicit clarification of points related to time-averaging and network topology.

Here we firstly provide a sketch of the content proposed for this additional section of the paper, and then address the reviewer's comments individually, illustrating how the new section would address these comments, and noting any additional revisions needed. The blue text indicates our proposed (draft!) amendment to the paper the text will be completed by a conceptual illustration clarifying the relation between the conceptual hydrologic model and the interpolation procedure. Bolded text in the reviewer response section indicates our proposed changes to the manuscript in response to the reviewer's comments.

1.0 Conceptual hydrologic model

TopREML and TopKriging have a very similar conceptualization of runoff generation along a river network. Runoff is assumed to be generated at point scales on the landscape, from where it is routed to a channel and measured at a gauge. Runoff observations made at any individual gauge can be theoretically separated into a "local" contribution, derived from a never-previously-gauged catchment area, and an "upstream" contribution that is observed within the channel at the upstream gauge(s).

The goal of both TopREML and TopKriging is to isolate the local effects within observations made at a gauge. These local effects then form a basis for spatial interpolation.

For the local effects to form a suitable basis for spatial interpolation, variations associated with temporal correlation (e.g. travel time effects) need to be removed. This is achieved by performing the spatial separation on time averaged data – or "runoff signatures" – with the proviso that the time averaging window is much greater than the characteristic catchment and channel response timescales. For this reason, both TopREML and TopKriging are typically performed on seasonal or annual average values of the flow metrics being considered.

1.1 How do TopKriging and TopREML differ?

Although TopKriging and TopREML share a common conceptual framework, they differ in terms of the geostatistical interpolation process that is used. This difference has several flow-on effects in terms of (i) how spatial randomness is conceptualized, and (ii) how local versus upstream contributions are separated in the methods.

TopKriging
TopKriging relies on the use of Kriging for interpolation. Kriging operates on the assumption that a random spatial process governs the point-scale behavior. Thus in TopKriging, runoff generation is assumed to follow a random point process on the landscape. TopKriging requires an estimate of the correlation structure of this point process. This estimate is made by (i) accounting for the fact that observations of runoff cannot be made at the point scale, and so must be areally aggregated (changing the equations that describe the semivariograms), and (ii) that overlapping contributing areas of nested sub-basins must be removed from the ultimate description of the point process. It is part (ii) that leads to the topological structure in TopKriging - network topology is introduced implicitly by the identification of these overlapping areas of nested basins, because this overlap results directly from the hierarchical structure of the river network.

TopREML
TopREML does not require information about a spatially random point process, but solely uses information from the gauges. It makes the assumption that at long averaging timescales a water balance can be applied to runoff observations in the basin, so that runoff observed at a downstream gauge is the aggregate of all upstream contributions and inputs from the local drainage area. This assumption can be used to separate the local contributions from each gauge in the basin, which form a basis for interpolation. The interpolation method assumes that: (i) the runoff signatures of local flow generation regions that are close to each other (in Euclidian space) are more likely to be identical and (ii) runoff signatures actually observed at the gages are the sum of the individual (unobserved) contributions of upstream local runoff areas.

Although TopREML does not require that runoff generation follow a point process (unlike TopKriging), the treatment of spatial correlation in TopREML is compatible with the existence of such a process (and thus with the underlying assumptions of TopKriging). The purpose of Appendix A is to demonstrate this compatibility.

1.2 Clarification of key theoretical concepts

Treatment of time:
As highlighted above, interpolation in TopREML is applied only to time averaged properties of the flow, specifically with an averaging window that is much greater than the intrinsic response timescales of the basins being modeled.

This treatment of time has several specific consequences:
a) TopREML cannot achieve real-time forecasting or cope with temporally correlated properties. It is suitable for the regionalization of time-averaged and statistically stationary runoff properties – or "runoff signatures"- to ungauged catchments. This is similar to most existing applications of Topkriging (e.g., Skoien 2006, Lahaa 2013).
b) TopREML predictions are insensitive to non-stationarities derived from to catchment response timescales, provided the averaging window used to derive the flow signatures is much larger than those timsescales.
c) Provided the averaging window is large enough for the flow behavior to have reached a statistical steady state (i.e. pseudostationary state), the water balance assumption used to separate local from upstream runoff behaviors is valid
d) All covariance/correlation arguments made in the paper should be read as referring to the spatial, and *not* the temporal, correlation of the runoff signatures.

Network topology:
Geomorphological considerations of the topology of a river network generally focus on the channels, and lead to an intuitive conceptualization that topological interpolation should focus on distance metrics computed solely along flowpaths. However, neither TopKriging nor TopREML utilize flow path length as a distance metric. Instead, both methods consider network connectivity as realized in the connections between contributing drainage areas. This is because drainage areas represent a conservative feature (they are constrained to sum to the total catchment area). In contrast, flow path lengths do not follow any simple conservation approach, making their use in geostatistical predictions theoretically troublesome. The reliance on area connectivity is perhaps more obvious in TopREML (where Euclidian distances between catchment centroids are used as a distance metric) than in TopKriging (where the areal connectivity is introduced within the regionalization procedure), however both methods conceptualize the network topology similarly.

Specific response to reviewer's detailed comments

Part of the confusion arises from the derivations presented in Appendix A: the averaging presented appears to use Eucledian distance and not flow distance, which is needed if one wants to capture stream network structure. How can this be the basis for the kriging carried out here?

The purpose of Appendix A is purely to demonstrate that TopREML and TopKriging are compatible – that if indeed (as TopKriging assumes) runoff is generated as a spatial point process on the landscape, the covariance structures used in TopREML would be appropriate approximators to the resulting covariance between the gauges (which see the areally accumulated flow properties). Euclidian spatial averaging is also assumed in TopKriging for this purpose.
We note again that the stream network structure is captured in both methods by consideration of contributing areas, not channel distances, and that provided the time averaging windows are large enough to average over travel-time variations, this should not be problematic.
We anticipate that the proposed new text in Section 1.1 "TopREML" and Section 1.2 "Network Topology" would serve to clarify this point in a revised manuscript.

Likewise, I do not understand the context of Appendix B. Why are we talking about events here? What signature are we trying to regionalize here?

One of the motivating factors for this study was a desire to regionalize some of the key parameters in a model for the flow duration curve (see Botter, 2007), which is driven by a statistical representation of the waiting time between positive runoff increments. Thus, the mean waiting time (or the inverse of the frequency) of the runoff events is a signature of interest for regionalization. Because this frequency is not a conservative variable (that can be isolated directly from a water balance argument), we provide this additional appendix to demonstrate how it can be modeled in TopREML.
We propose mentioning Botter's model specifically as a motivating factor when introducing the runoff frequency as a signature of the flow regime.

Both in the title and within the body of the text, the paper talks of runoff signatures. What do they mean? My understanding of signatures is that they are aspects of runoff variability extracted from observed runoff time series: flood frequency curve, flow duration curve, or the regime curve (mean seasonal runoff) etc. The authors do not go to any more specifics, and so I am confused. I do know that each of these signatures can be distributed across the network (including their moments). So, which of these signatures is being predicted here in an ungauged basin context?

"Signatures" specifically refer to time-averaged and statistically stationary runoff properties (see point 1.0 above).
We anticipate that the initial explanations provided in the proposed new section 1.0 "Conceptual model" and 1.2 "Treatment of time" will assist with

communicating this point. We will also include a more explicit description when introducing the example signatures to be regionalized.

> *The authors state that they make a water balance assumption to enable the spatial averaging, is this not akin to a steady state assumption? If so, how would steady state apply to any of the above signatures? In the case of event hydrographs, don't you have to deal with timing delay between upstream and downstream locations? In the case of flood frequency curves, can you assume steady state for the same return period? In the case of flow duration curve, can you assume steady state for the same frequency?*

The reviewer is correct that there are some important hydrological variables for which it would be problematic to make a steady state assumption, and that these variables should not be used in TopREML. Indeed, if a water balance assumption cannot be reasonably made, then TopREML's application will be difficult, because local and upstream influences will not be reliably separated.
We anticipate that the proposed discussion in section 1.0 "Conceptual model" and 1.2 "Treatment of time" will clarify this point. However we will also delineate some inappropriate applications of TopREML in the revised manuscript.

> *What is a point process? I can understand that precipitation is a point process, and runoff generation can also be a point process if it is estimated in a small pixel, but runoff leaving any point is already accumulated over an area upstream (whether on a hillslope or on a stream network). When I used to work in this area, I used to frame it as an averaging in space-time that accounts for the time needed for water droplets to arrive from wherever it is generated.*

A spatial point process is a statistical model of spatial randomness for which any one realization consists of a set of points in time/space. We agree that any measurement of runoff is NOT a measurement of a point process, but an aggregated measure of that process. This is why TopKriging requires a regularization step – the observations of the point process are made only in aggregate, but estimates of the point process are needed. TopREML doesn't require that runoff be considered as a point process, but it is compatible with such estimations, and thus with the underlying assumptions of TopKriging (as per Appendix A).
We consider that this point would be adequately made in the proposed new introductory section

> *I am concerned that I do not see distance measured along the network figure prominently in the presentation anywhere. How about the time delay? It is possible that the authors are indeed using this feature but it is not presented explicitly. I want clarification.*

We agree that we need to clarify this point further. Time delay is neglected by focusing on temporally averaged flow signatures. Real-time or short-time interpolation via TopREML would be inherently biased because the proposed interpolation is only through space and not through space-time. As discussed previously, the connectivity used in TopREML is based on the hierarchy of contributing areas measured at a gauge, not flow path length. This is in part because contributing area is sufficient for the time-averaged methods used here, and also because it avoids the theoretical difficulties posed by attempting to incorporate non-conservative flow path lengths.
We consider that this point would be adequately made in the proposed new introductory section

> *The presentation is currently dominated by kriging language, but the authors should weave in hydrological language – and motivate the assumptions made hydrologically.*

As noted in our initial response to this review, the major contribution of TopREML is in terms of an alternative geostatistical method, not an alternative hydrologic conceptualization. We hope that the proposed new introductory section that introduces the underlying conceptual model would address the reviewer's point. However, since many of the statistical contributions here are fundamentally statistical and not hydrological, we anticipate that geostatistical language will continue to feature prominently in the revised manuscript.
We anticipate that by flagging that the contribution of TopREML is primarily geostatistical rather than based on a change in conceptual hydrological models, readers would understand the reliance on geostatistical argumentation.

> *I really would like to see even more clearly the separation from top-kriging both in terms of problem formulation as well as results.*

The key differences between the two methods in terms of problem formulation and assumptions are listed in the introduction section of the manuscript, and we anticipate that the proposed new introductory section will make these differences even more explicit.

The key differences are:

1. Treatment of network topology: TopREML uses water balance considerations to constrain the covariance structure of runoff and to account for the stronger spatial correlation between flow-connected basins (p1360 l.25). Topology is implicitly accounted for in Top-kriging through the regularization of overlapping catchment areas (Lahaa 2013, p673 §4).

2. Runoff generating process: TopREML assumes that runoff signatures are generated by spatially correlated homogenous IGAs (p1361 l.5). Top-kriging assumes an underlying spatially correlated point-scale process (Lahaa 2013, p673 §3).
3. Estimator: TopREML uses a restricted maximum likelihood estimator to estimate variance parameters (p1361 l.5), while Top-kriging uses a kriging (method of moments) estimator (Lahaa 2013, Eqn 3).

These three points are mentioned in the current manuscript on page 1359 (key assumptions and shortcomings of Topkriging) and pages 1360 – 1361 (distinguishing characteristics of TopREML). Finally the performance of these two methods is repeatedly compared throughout the result section (Figures 4 and 5).

In spite of claiming to capture river network structure I do not see the picture of a river network presented in either of the applications – this makes it less appealing hydrologically. Finally, again as a hydrologist, I would like to see real results of actual signatures being predicted, in addition to the current focus on performance and uncertainty.

Thank you for this suggestion.
We will include in the revised manuscript maps of the river network in the Austrian case study with appropriately color-coded representations of the observed and predicted mean summer flow at control gauges.

References:

1. G. Botter, A. Porporator, I Rodriguez-Iturbe and A. Rinaldo. Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. Water Resources Research, 43(2), 2007

2. G. Laaha, JO Skøien, F. Nobilis, and G. Bloöschl. Spatial prediction of stream temperatures using top-kriging with an external drift. Environmental Modeling & Assessment, 18(6):671–683, 2013.

3. JO Skøien, R Merz, and G Blöschl. Top-kriging–geostatistics on stream networks. Hy- drology & Earth System Sciences, 10(2), 2006.