The Wasserstein distance as a hydrological objective function

Jared C. Magyar[1,*] and Malcolm S. Sambridge[1]

[1]Research School of Earth Sciences, Australian National University, Canberra, Australia
[*]Now at Physics, School of Natural Sciences, University of Tasmania, Hobart, Australia

Correspondence: Jared Magyar (jared.magyar@utas.edu.au)

Abstract. When working with hydrological data, the ability to quantify the similarity of different datasets is useful. The choice of how to make this quantification has direct influence on the results, with different measures of similarity emphasising particular sources of error (for example, errors in amplitude as opposed to displacements in time and/or space). The Wasserstein distance considers the similarity of mass distributions through a transport lens. In a hydrological context, it measures the 'effort' required to rearrange one distribution of water into the other. While being more broadly applicable, particular interest is payed to hydrographs in this work. The Wasserstein distance is adapted for working with hydrographs in two different ways, and tested in a calibration and 'averaging' of hydrographs context. This alternate definition of fit is shown successful in accounting for timing errors due to imprecise rainfall measurements. The averaging of an ensemble of hydrographs is shown suitable when differences among the members is in peak shape and timing, but not in total peak volume, where the traditional mean works well.

1 Introduction

1.1 Motivation

A fundamental aspect of hydrology is understanding the distribution and movement of water through the Earth system. It is therefore necessary to quantify the similarity of a pair of spatial (e.g. precipitation fields) or temporal (e.g. hydrographs) distributions of water. The goal of this quantification may simply be to gauge the semblance of the distributions. In other, more complex cases, it may act as an objective function for parameter estimation, model calibration, or data assimilation, where the goal is to minimise the discrepancy between a model output and observed data. Again still, we may be interested in the 'average' distribution of water among an ensemble. Each of these varied tasks unify in the requirement of some measure of similarity between pairs of distributions, with the characteristics of this choice important for the quality of the result.

Such a quantification varies in nomenclature according to discipline and purpose, but some common terms are objective, response, or misfit function. While it is abundantly clear that there is great need for comparative measures in hydrology, there remains ambiguity in selecting that which best quantifies discrepancy for any given application. In this selection process, we are led back to the more fundamental query: *what is a good 'fit'?* The misfit function that we use must be intrinsically linked with what we perceive as a 'good fit' for the application and purpose at hand.
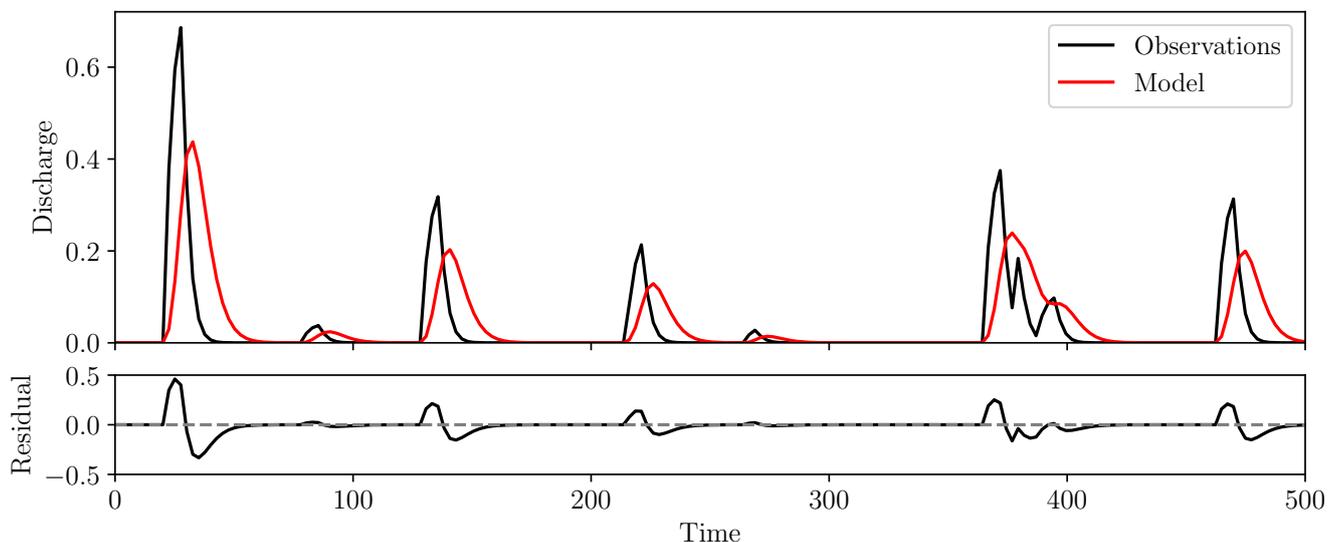
Figure 1. Two hydrographs (top panel, red and black) and the point-wise residuals between them (bottom panel). Both timing and amplitude differences are observed, but these are combined into a single amplitude error to give large residuals, even though the human eye might perceive these two distributions as being quite close.

A commonly used metric for comparing both hydrographs and spatial fields is the root mean square error (RMSE), given by Eq. (1) where f and g are the functions under comparison, and have known values at the N points $\boldsymbol{x}_i$.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}[f(\boldsymbol{x}_i) - g(\boldsymbol{y}_i)]^2}{N}} \tag{1}$$

For the RMSE, f and g are compared by studying the difference in amplitude at specified points. While varying in the details, there are a variety of metrics, such as the Nash-Sutcliffe efficiency (NSE) and L^2 distance, that are similar in their derivation and compare the amplitude of the functions at specified locations. We will refer to the metrics of this type collectively as 'point-wise' misfit functions, as each datum is compared only with that equal in time and space. These are undoubtedly simple and computationally efficient, but they do not satisfactorily account for spatial or temporal *displacement* of features (Roberts and Lean, 2008; Wernli et al., 2008; Farchi et al., 2016). Figure 1 displays such an occurrence, where displacement is also a major source of difference between the two synthetic hydrographs, rather than purely amplitude, yet for point-wise misfit functions such differences are forced to be represented by amplitude alone.

The cause of the large residuals in Fig. 1 stems from the underlying assumption that 'like' features are already aligned when the misfit is computed. In essence, it is built upon the assumption that errors are in strictly in amplitude rather than displacement (Lerat and Anderssen, 2015). In the case of hydrographs, even small time-shifts can yield high misfit in steep rising limbs and recessions; a consequence of this amplitude-based misfit interpretation (Ewen, 2011; Ehret and Zehe, 2011).
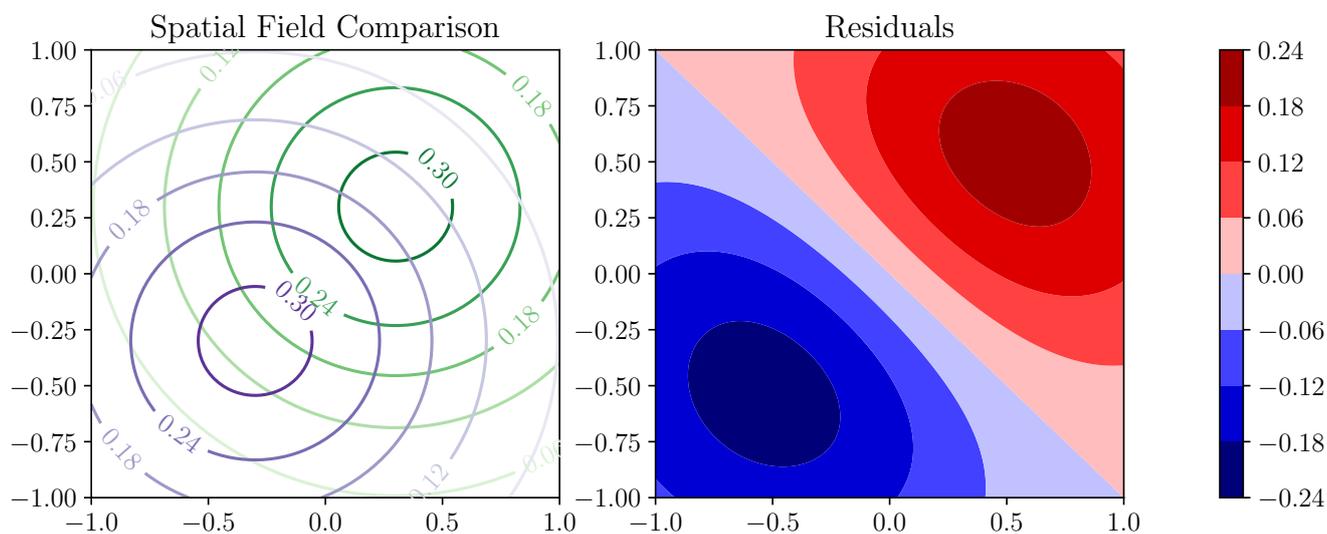
Figure 2. Synthetic example of the comparison of modelled (purple) and observed (green) spatial fields in the left panel. Right panel shows the residuals (green field minus purple field), which displays the double penalty effect as error is in displacement rather than amplitude.

40   As we can expect errors in both timing *and* amplitude when modelling streamflow, point-wise metrics alone are frequently inadequate, and therefore require some type of multi-objective function (Yapo et al., 1998).

    In a more general, multivariate setting, an analogous issue can occur. The misalignment of features displayed in Fig. 2 leads to the 'double penalty' effect, where large errors are assigned for small displacements errors, this occurring both where the model (purple in Fig. 2) has predicted the high amplitudes, and where the high amplitudes are observed (green) (Wernli et al.,

45   2008; Farchi et al., 2016).

    This poor quantification of fit under the influence of the displacement of features has led to a range of other 'scores' being used when validating forecasts from numerical weather predictions (see, for example, Roberts and Lean 2008; Wernli et al. 2008). These attempt to account for both the differences in particular features of each field, and the distance between these identified features.

50   An inadequate measure of fit can also lead to spurious local minima in the parameter estimation objective function, making local optimisation algorithms ineffectual (Engquist and Froese, 2014). These are especially common when the recorded signal is multi-modal. In these cases, displacement of such multi-model features can lead to the 'peaks' becoming incorrectly aligned and producing a locally low misfit (see Fig. 3).

    Whether the misfit is unsuitable due to incorrect quantification of timing errors (Fig. 1), the double penalty effect (Fig. 2),

55   or local minima (Fig. 3), the key downfall of a point-wise metric can be traced to the same source: function comparisons occur in complete isolation, independent of all other points. However, 'cross-domain' comparisons are fundamentally necessary to recognise the temporal or spatial displacements which generate the problems. These types of issues are not exclusive to hydrology. For example, seismic signals can be compared via differences in amplitude, or by considering differences in travel
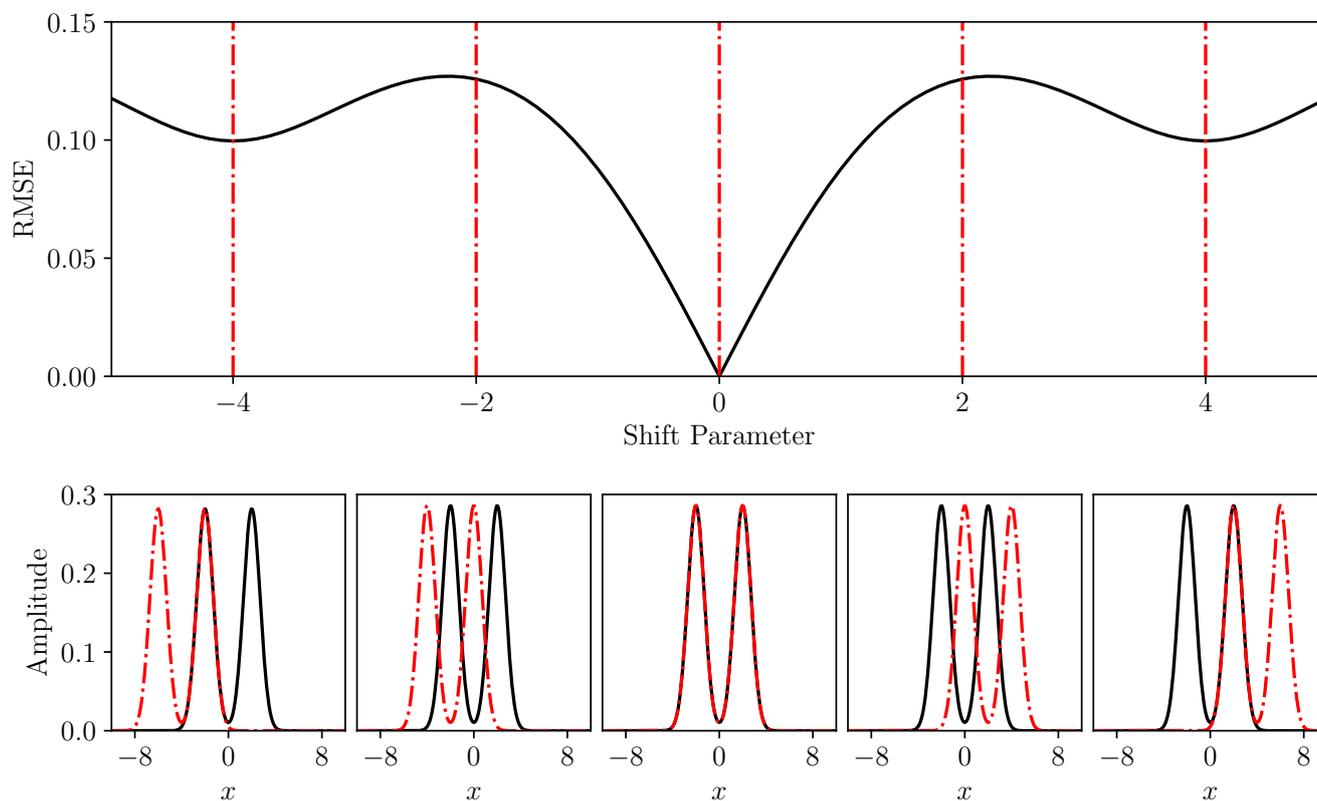
Figure 3. Root mean square error (top) between double Gaussians at a variety of time shifts (bottom, solid black and dash-dot red). Bottom panels correspond to shift parameters indicated in top panel with dash-dot red lines. In this example the situations shown in the far left and and far right lower panels, with shift values of -4 and 4 respectively, both represent local minima in the RMSE function above, because the black curve has just a single peak aligned with the red dash-dot curve in each case, whereas the global minimum at shift value 0 has both peaks aligned.

time of seismic phases and thus timing errors. Local minima in the misfit function frequently arise when using amplitude based misfit functions, this being the root of the infamous 'cycle-skipping' problem commonly encountered in full waveform inversion, resulting from similar misalignment issues to those displayed in Fig. 3. Due to these similarities, we are motivated to explore the Wasserstein distance, derived from the field of *optimal transport* that has found recent use in seismology for quantifying waveform similarity in cases where amplitude-based misfit functions are inadequate (Engquist and Froese, 2014; Métivier et al., 2016).

While briefly addressed in Ehret and Zehe (2011), to the best of our knowledge the Wasserstein distance has had limited exposure in hydrology. It has been found beneficial for parameter estimation problems in geophysics, due to it comparing data holistically, rather than in isolated pairings, with this key property being what we seek to exploit (Engquist and Froese, 2014;

Métivier et al., 2016). It has also been considered in the comparison of atmospheric contaminant fields (Farchi et al., 2016) and in some diagnostic data assimilation tasks with a similar rationale (Feyeux et al., 2018; Li et al., 2019).

70   Much like techniques considered in Roberts and Lean (2008) and Wernli et al. (2008), the Wasserstein distance is displacement-based, allowing more appropriate misfit quantification under temporal or spatial shifts. Its definition is derived from the concept of 'transporting' one mass distribution onto the other, so naturally accounts for any such shifts. Consequentially, small timing errors for hydrograph comparison do not incur large penalties resulting from the misalignment of peaks in flow, and is a more suitable measure of fit in the presence of timing errors than the RMSE or Nash-Sutcliffe efficiency. Use of the Wasserstein dis-

75   tance bears some similarity to the dynamic time warping technique of Ouyang et al. (2010), the series distance (Ehret and Zehe, 2011), the hydrograph matching algorithm of Ewen (2011), and the cumulative distribution method of Lerat and Anderssen (2015).

In this work, rather than detailing highly specific applications with complex models, we instead introduce methods that are suited for a range of tasks in (but not limited to) hydrology. The main goal is to highlight some of the key concepts of

80   optimal transport, and how they may find use in hydrology via simple, illustrative examples. The limitations of its use are also discussed, and where further research is required. We also limit our experiments to one-dimensional problems, in particular the comparison of hydrographs. However, Sect. 1.2 details optimal transport in a more general sense, as it conserves its useful properties in multiple dimensions so can also be used for comparing spatial fields.

1.2   Optimal transport

85   Optimal transport (OT) provides the numerical machinery for interrelating density functions. As distributions of water can be well represented as a density function with appropriate scaling, we suggest that OT gives the required tools for making more satisfactory comparisons than can be achieved by point-wise misfit functions.

Before the more modern interpretation with regards to density functions, OT originated with the work of Monge (1781), who was tasked with finding the most cost-effective way of filling a hole with a pile of dirt. OT is hence fundamentally concerned

90   with reshaping one distribution of mass into another whilst minimising the requisite effort.

Little progress was made on generalised solution methods for this problem until it was rephrased by Kantorovich in 1942. This forms the foundation of many modern computational methods (Peyré and Cuturi, 2019). However, we will introduce OT in a form closer to the original Monge definition, as this is based on continuous functions. While measured discretely, distributions of water have an underlying continuity in time or space so this description is appropriate.

95   The central aspect of Monge's studies was an *optimal transport map*, acting as a directory between each location in the pile of dirt (known as the source) and an associated location in the hole (known as the target). By following this optimal map, the source distribution can be rearranged into the target distribution with minimal work. The relation between the source, target, and transport map is presented in Fig. 4.

Key restrictions for transport are mass conservation, meaning the source and target must have equal mass, and non-negativity,

100   meaning that there cannot be negative mass at any point in either. These properties have led OT to being framed as between probability densities, as these naturally obey both requirements. If we denote the source as f and the target as g, these functions
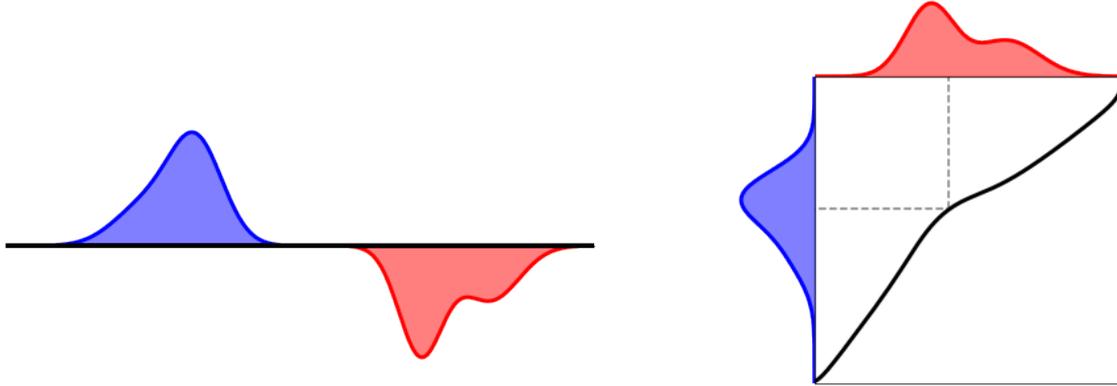
Figure 4. The left panel shows Monge's initial formulation of OT, with a source mass distribution (blue) and a target to be filled (red). The right panel shows the solution to the OT problem, with a transport map (black) transporting the source to the target. Dashed line shows transport for a particular point.

must satisfy Eq. (2) and Eq. (3).

$$\int_{\mathcal{X}} f(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{\mathcal{Y}} g(\boldsymbol{y}) \, d\boldsymbol{y} = 1 \tag{2}$$

$$f(\boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{x} \in \mathcal{X}, \quad g(\boldsymbol{y}) \geq 0 \text{ for all } \boldsymbol{y} \in \mathcal{Y} \tag{3}$$

105   Here, $\mathcal{X}$ is the space over which f is defined, and $\mathcal{Y}$ is the space over which g is defined. In the case of hydrographs, these will be the time window over which the streamflow was measured, while for spatial fields, they will be the spatial extent of the recordings.

With the overarching goal of transporting f to g with minimal work, we must first define some notion of 'effort' through the cost function, $c(\boldsymbol{x}, \boldsymbol{y})$, giving the amount of work it takes to transport one unit of mass from $\boldsymbol{x}$ to $\boldsymbol{y}$. With this cost function in

110   hand, the *total cost* of following some transport map T will be given by Eq. (4).

$$C(T) = \int_{\mathcal{X}} \overbrace{\underbrace{c(\boldsymbol{x}, T(\boldsymbol{x})) f(\boldsymbol{x}) \, d\boldsymbol{x}}_{\text{cost for a single point}}}^{\text{cost across } all \text{ points}} \tag{4}$$

Finding the *optimal* transport map (denoted as T^*) therefore amounts to minimising Eq. (4), subject to the additional requirement that this map transports f *exactly* to g. Needless to say, this is a complicated functional minimisation problem, so will not be discussed in depth here as a simple analytical result is instead used. Direction to alternate methods is also provided in Sect.

115   2.1.

Once found, the optimal transport map can be used to define the p-Wasserstein distance; the central concept of interest for this work. The p-Wasserstein distance (or simply, the Wasserstein distance) is defined using the p-norm raised to the p-th power as the cost function, giving the cost of Eq. (5) where D is the dimensionality of the densities and subscripts are vector

components.

$$120 \quad c(\boldsymbol{x}, \boldsymbol{y}) = \sum_{d=1}^{D} |x_d - y_d|^p = \|\boldsymbol{x} - \boldsymbol{y}\|_p^p \tag{5}$$

The choice of p reflects how severely mass transport is to be penalised. That is, if p is large, mass transport across long distances will require more 'work' relative to small transport distances and thus be avoided if possible. Upon finding the optimal transport map under the cost function Eq. (5), the Wasserstein distance is given by Eq. (6).

$$\mathcal{W}_p(f, g) = \left(\int_{\mathcal{X}} \|\boldsymbol{x} - T^*(\boldsymbol{x})\|_p^p f(\boldsymbol{x}) \, d\boldsymbol{x} \right)^{\frac{1}{p}} \tag{6}$$

125 The Wasserstein distance has many favourable metric properties for probability densities, a deeper discussion of which can be found in Ambrosio et al. (2008); Villani (2009). However, for the purpose of the present discussion, it is sufficient to state that similar mass distributions require little work to transport between and thus have a low Wasserstein distance in comparison to dissimilar distributions, which require more work for transport. We will pay particular attention to the square of the 2-Wasserstein distance, $\mathcal{W}_2^2$, due to its convex properties that make it well suited for parameter estimation purposes (Engquist

130 and Froese, 2014). We therefore set $p = 2$ in Eq. (5) and Eq. (6).

Armed with the Wasserstein distance we can, for example, gauge the similarity of two hydrographs, or a pair of precipitation fields. Again, note the key property that this is a *global* comparison, considering the total cost of transporting the *entire* source to the target, rather than locally quantifying the distance using point-wise comparisons. It therefore naturally accounts for displacements through the transport map and and assigns a more appropriate penalty for this type of error. We can see

135 this through the simple numerical experiment in Fig. 5, where the Wasserstein distance provides a consistent interpretation of displacement, even when the features are not overlapping. It also mitigates the occurrence of local minima, due to cross-domain comparison. In a parameter estimation context, this provides a smooth, convex function well suited for gradient-based local optimisation methods.

These properties with respect to displacements have encouraged application in atmospheric chemistry and seismology alike

140 (Farchi et al., 2016; Engquist and Froese, 2014; Métivier et al., 2016). However, while the Wasserstein distance naturally measures differences in shape and displacement, it is blind to differences in total mass between the source and target as they are scaled to unit mass prior to computation. Potential methods to counter this are proposed in Sect. 3 for applications where this may prove to be problematic.

1.3 Wasserstein barycenters

145 Beyond a straightforward measure of fit, the Wasserstein distance can also be used to define an 'average' of density functions, known as the *Wasserstein barycenter*. The barycenter is the distribution that is the closest, in a Wasserstein sense, to the distributions we are finding the average of. This makes it the density that minimises a weighted sum of Wasserstein distances. If the ensemble of densities are denoted g_i and the barycenter is f, then f must obey both the requirements of a density Eq. (2)
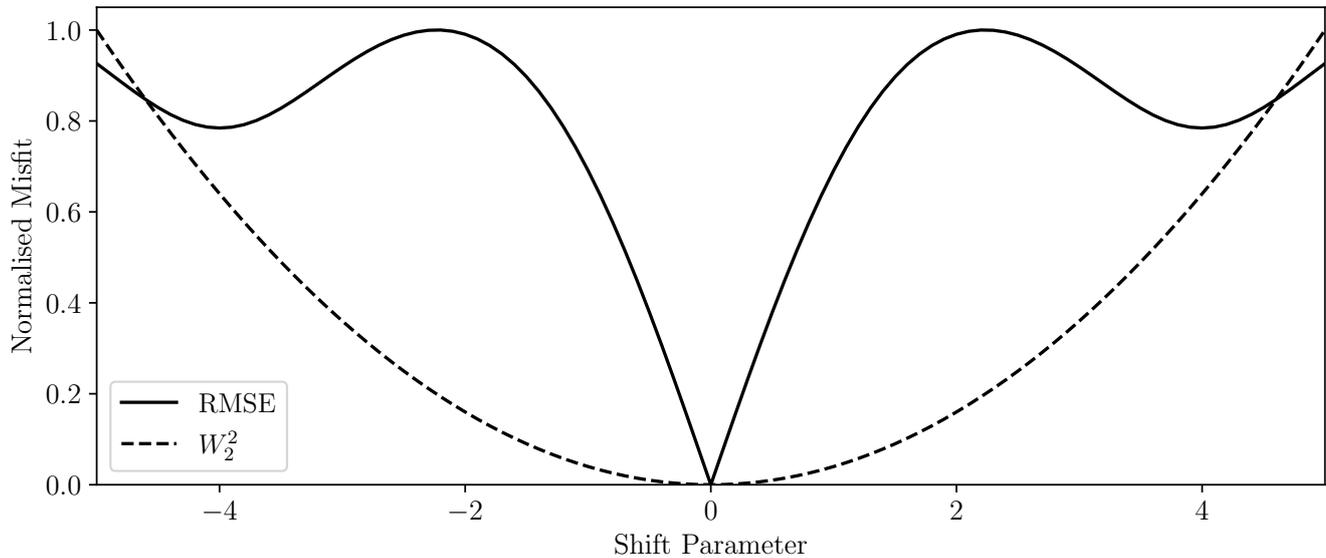
Figure 5. Comparison of RMSE and squared 2-Wasserstein distance for double Gaussian problem displayed in Fig. 3. All misfits are scaled to maximum of one for visualisation purposes.

and Eq. (3) whilst minimising Eq. (7), with the relative weights of the component densities given by λ_i.

$$150 \quad \min\left(\sum_{i=1}^{n}\lambda_i\mathcal{W}_2^2(f,g_i)\right), \quad \sum_{i=1}^{n}\lambda_i = 1 \tag{7}$$

As OT makes cross-domain comparisons, the Wasserstein barycenter gives a more intuitive intermediate distribution when densities are separated in time or space. We can explore this through a simple example in Fig. 6.

Taking the arithmetic mean at each point of the two distributions gives a poor intermediate representation. Both of the densities have a single peak, but the intermediate in Fig. 6 has two, so has not captured the characteristics of either. However,

155 the Wasserstein distance provides a more satisfactory average conserving the shape of the peak. This again can be traced to how we define similarity. The conventional mean is defined with respect to amplitude, while the Wasserstein barycenter is an average with respect to the *displacement pathway*, also known as the *geodesic* (Ambrosio et al., 2008). Section 3 uses the Wasserstein barycenter to find the average hydrograph among an ensemble, each perhaps derived from a different climate or hydrological model, allowing conservation of peak shape and number in the case where equivalent peaks are equal in volume.

160 2 Computational methods

2.1 A brief survey

Now that the favourable properties of the Wasserstein distance for mass transport problems have been explored, we must discuss the computational methods. This aspect is perhaps the greatest difficulty in large-scale implementation. There are a variety of
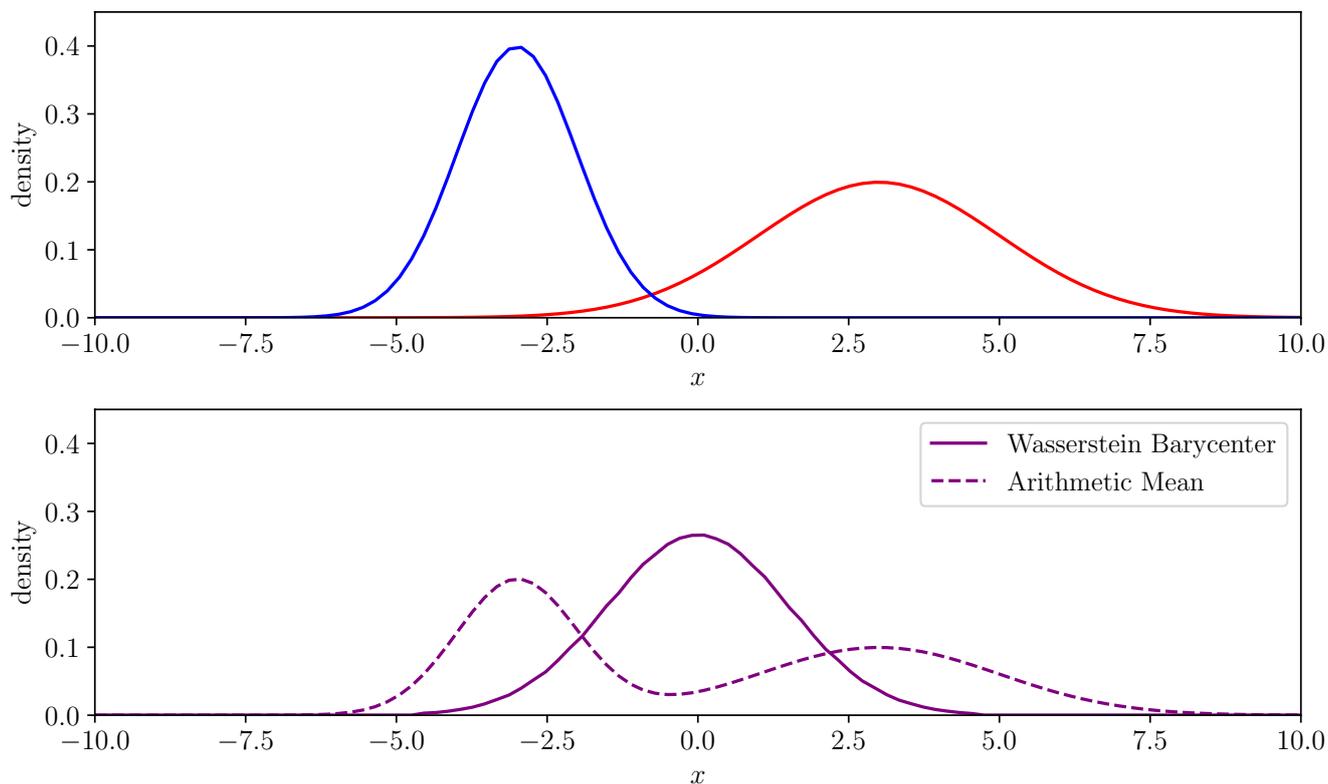
Figure 6. Two Gaussian density functions (blue, red, top panel) with one spatial/temporal dimension x. The Wasserstein barycenter of these densities (solid purple, bottom panel) and the mean (dashed purple, bottom panel).

available techniques, some of which have been explored previously in similar applications, whilst others are a source of future

165   discussion. A brief survey follows, with further details given in the associated references, many of which have unexplored potential for hydrological applications.

It was shown by Brenier (1991) that minimising Eq. (4) can be transformed into a partial differential equation when the squared Euclidean distance is used as the cost function. This can then be solved numerically, an example of one such solution given in Benamou et al. (2014). This was the preferred method of Engquist and Froese (2014); Yang et al. (2018) for comparison

170   of seismic waveforms, although it is a technique not frequently seen in applied studies. The applicability of the method in seismology implies that this solution may be appropriate and warrant further consideration in the future for hydrology.

The more commonly used approach, especially for machine learning and graphics applications, is to use the Kantorovich formulation and thus consider the source and target densities as discrete probability distributions, for which a linear program can be solved to obtain a transport plan (a generalised version of the transport map). While this is effective for a broader

175   range of problems than the Monge-Ampere PDE that is specific to the squared Euclidean distance cost function, solving the linear program is computationally expensive and scales poorly with the size of the problem. This burden can be somewhat

alleviated using the entropically regularised approach pioneered by Cuturi (2013), which has become the foundation for modern machine learning applications. Further efficiency can be achieved on regularly spaced data using the convolutional entropic regularisation developed in Solomon et al. (2015). Also derived from the entropically regularised approach are the stochastic
180   optimisation methods of Genevay et al. (2016) and Seguy et al. (2018). These are particularly useful when samples can be drawn from the source and target densities.

There appears to be a bias in the literature towards discrete approaches, seemingly due to many machine learning problems not having a continuous interpretation, leaving the Monge-Ampere and related continuous approaches meaningless. We therefore should not be discouraged from using continuous methods for problems in the Earth sciences with spatial or temporal
185   data, despite the relative lack of previous applications in the present literature. Indeed, we can see previous success of these methods in Engquist et al. (2016) and Yang et al. (2018).

2.2   One-dimensional transport

The method used in this work is structured around the special case of transport in one-dimension, which accommodates a highly computationally efficient solution. When the source density f and target density g are both one-dimensional, the optimal
190   transport map between them is given by Eq. (8).

$$T^*(x) = G^{-1}(F(x)) \tag{8}$$

Here, we have used F and G to represent the cumulative distribution functions (CDFs) of the source and target, these defined according to Eq. (9).

$$F(x) = \int\limits_{-\infty}^{x} f(x')\,dx', \quad G(y) = \int\limits_{-\infty}^{y} g(y')\,dy' \tag{9}$$

195   Upon substituting the optimal map Eq. (8) into Eq. (6) to obtain the p-Wasserstein distance, a change of variables gives the simple form Eq. (10).

$$\mathcal{W}_p(f,g) = \left(\int\limits_{0}^{1} \left[F^{-1}(s) - G^{-1}(s) \right]^p \, ds \right)^{\frac{1}{p}} \tag{10}$$

This reveals that, in the one-dimensional case, the Wasserstein distance is a measure of discrepancy between the inverse CDFs of the source and target. A visualisation of the one-dimensional Wasserstein distance under this interpretation is given in Fig.
200   7.

In this study, we use the Nelder-Mead algorithm for optimisation, so the gradient of the Wasserstein distance is not required. However, if a gradient based method were to be used, the derivative for the one-dimensional case can be derived directly from Eq. (10). Indeed, it was shown by Sambridge et al. (2022) that both the Wasserstein distance and its derivative can be computed through a sorting algorithm followed by a dot product. Engquist and Froese (2014) and Métivier et al. (2016) have also made
205   use of the derivatives of the Wasserstein distance for parameter estimation, although did not make use of the one-dimensional form.
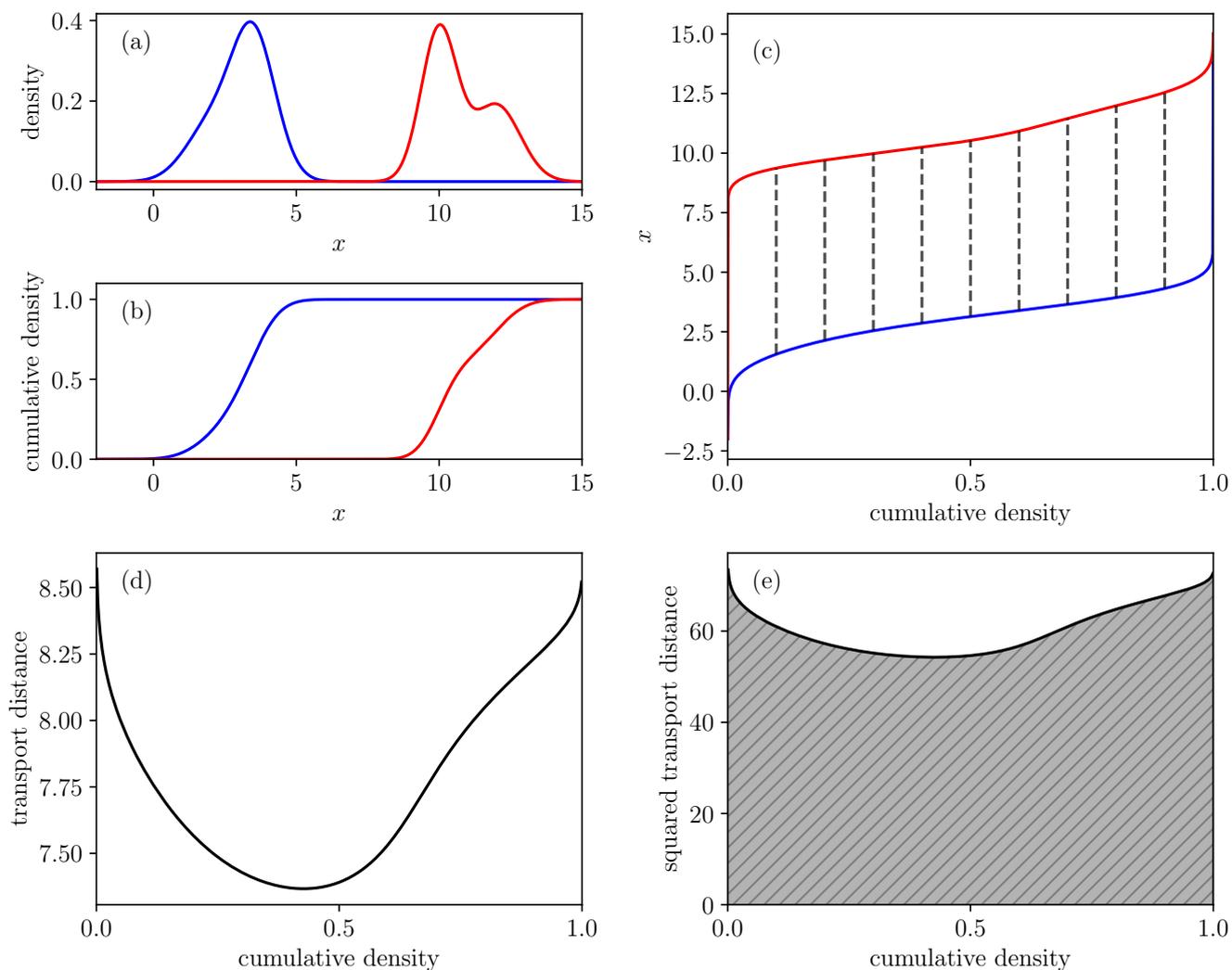
Figure 7. Process of moving from densities to 2-Wasserstein distance in one spatial/temporal dimension x. (a) Density functions from Fig. 4. (b) The cumulative distributions. (c) The inverse cumulative distributions with dashed lines showing transport pairs. (d) The difference between the inverse cumulative densities. (e) The area under the squared difference, giving the squared 2-Wasserstein distance.
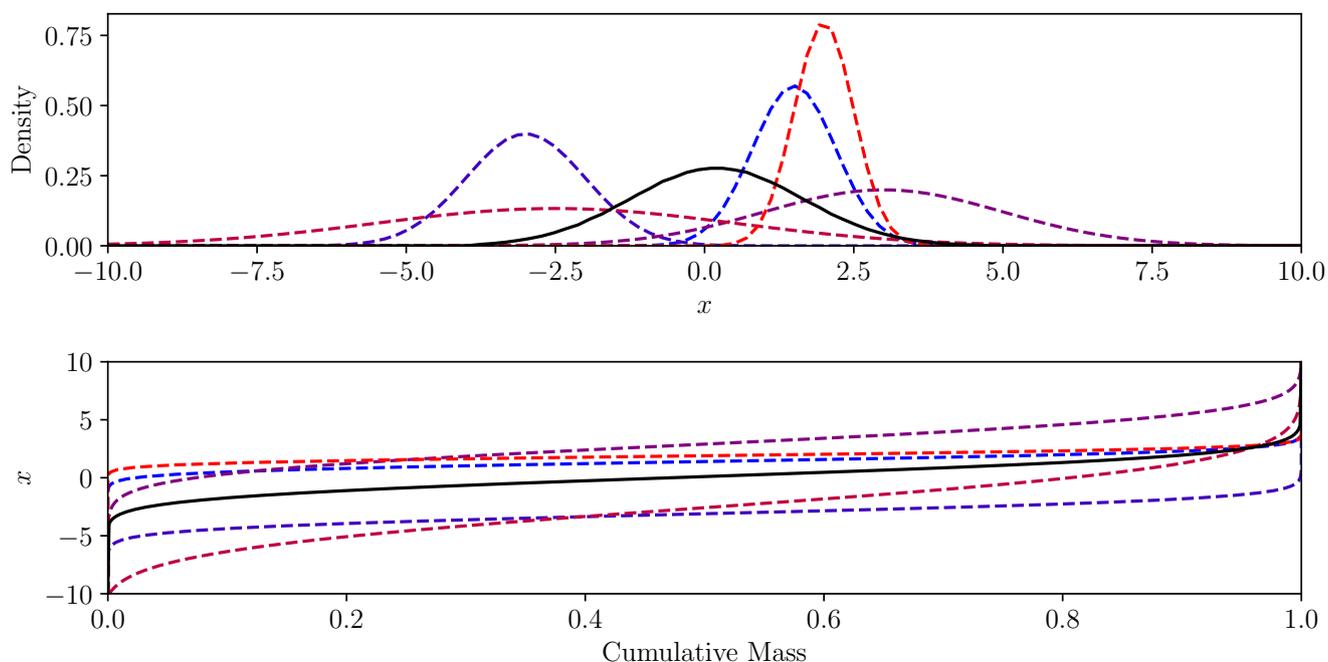
Figure 8. Method of finding barycenter (solid black) from 5 Gaussian densities (dashed). Top panel shows density functions, bottom panel the inverse cumulative distributions. The mean of the inverse cumulative distributions is the cumulative distribution of the barycenter.

Much like the Wasserstein distance, finding the Wasserstein barycenter of a series of densities is vastly simplified in one dimension. Rather than performing some type of optimisation scheme to minimise Eq. (7), we can find the barycenter directly from the inverse CDFs. Keeping the notation from Eq. (7), we have Eq. (11) (Bonneel et al., 2015).

$$F^{-1}(s) = \sum_{i=1}^{n} \lambda_i G_i^{-1}(s), \quad s \in [0,1] \tag{11}$$

210

That is, the inverse CDF of the barycenter is the weighted sum of the inverse CDFs of the component densities. Finding the barycenter then becomes a problem of converting an inverse CDF evaluated at a set of discrete points into a density function. Interchanging the (s, F^{-1}) coordinate pairs converts from the inverse CDF F^{-1} to the CDF F. A finite difference differentiation method can then be used to move from the CDF to the density function f. Finally, this can be interpolated into

215 a smooth density function.

Overall, the Wasserstein barycenter definition of average is equivalent to the histogram interpolation of Read (1999), although the author appears to have proposed this idea independently. The relation between the densities and their barycenter is displayed in Fig. 8.

Again, it should be made clear that the computational results discussed here apply *only* to the one-dimensional case. Different

220 methods must be used for multivariate problems, a selection of which are given in Sect. 2.1. There is also the potential to use the alternative *sliced Wasserstein distance* (Rabin et al., 2011) or the more general continuous extension known as the

Radon Wasserstein distance (Bonneel et al., 2015), which both extend the one-dimensional results using one-dimensional projections of the multivariate mass distribution. The Radon and sliced Wasserstein distances display similar properties to the true Wasserstein distance, and are a potential alternative to the more computationally expensive methods.

225 ## 2.3 Adaptation to hydrology

OT is defined only for probability densities. For the applications we envisage, the non-negativity requirement will naturally be obeyed, as we are measuring inherently positive masses of water. The restriction of unit mass requires a little further consideration. We could take one of two differing philosophies: modify the data so it is compatible with OT, or redefine OT such that it works with the type of data used. Here, we will consider both. Firstly, we can scale the water distributions such that

230 they have unit mass, and *then* compute the Wasserstein distance. The total mass of the densities is given by Eq. (12).

$$M_f = \int_{\mathcal{X}} f(\boldsymbol{x})d\boldsymbol{x}, \quad M_g = \int_{\mathcal{Y}} g(\boldsymbol{y})d\boldsymbol{y} \tag{12}$$

The scaling to unit mass is then given by Eq. (13).

$$\hat{f}(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{M_f}, \quad \hat{g}(\boldsymbol{y}) = \frac{g(\boldsymbol{y})}{M_g} \tag{13}$$

Though this indeed makes the data compatible with OT, it also means that the Wasserstein distance becomes completely

235 insensitive to differences in total mass. A multi-objective approach can be used to allow both shape (Wasserstein) and total mass (additional penalty term) to be accounted for, giving a misfit of the form Eq. (14).

$$\mathcal{W}_{2,\gamma}^2 = \mathcal{W}_2^2(\hat{f}, \hat{g}) + \gamma h(f, g) \tag{14}$$

Here, γ provides the relative importance of the total mass error compared to the Wasserstein distance, while h is the specific form of the penalty term. The quadratic function Eq. (15) is a simple choice for this penalty.

240 $h(f, g) = (M_f - M_g)^2 \tag{15}$

By selecting an appropriate weighting factor, we can develop a balanced misfit measure that contains the favourable properties of the Wasserstein distance, whilst also removing the insensitivity to total mass differences. Note however that this weighting is a *choice*, so may require some tuning procedure to attain the desired balance between the two aspects of fit.

We will now take the alternative approach of making a modification to OT to suit our hydrological purposes. The proposed

245 result is derived for one-dimensional data, although could potentially be extended to the multivariate case in the same manner as the Radon Wasserstein distance (Bonneel et al., 2015). We can build the new definition upon the foundation that we can compute the cumulative mass of a non-negative function, regardless of whether this mass sums to one, and that this function will be monotonically increasing. The key step of one-dimensional OT is that points of equal cumulative mass are matched in the optimal map (see Eq. (8) and Fig. 7). A similar concept with cumulative densities of arbitrary mass can be designed.

250 Inspired by the ordinary Wasserstein distance, we can map the points of half cumulative mass together (medians in the case of

probability densities). To ensure these points are always aligned regardless of total mass, a shift can be applied such that the cumulative densities are centred at this halfway point using Eq. (16) and Eq. (17).

$$F_s(x) = \left(\int_{-\infty}^{x} f(x')dx' \right) - \frac{M_f}{2} \tag{16}$$

$$G_s(y) = \left(\int_{-\infty}^{y} g(y')dy' \right) - \frac{M_g}{2} \tag{17}$$

255 For the matching of points of equal cumulative mass, the inverses of these are required. As we are working in the time window $[0, T]$, we can limit the inverse to this range, giving the piece-wise functions Eq. (18) and Eq. (19).

$$\overline{F}^{-1}(s) = \begin{cases} 0, & s < -\frac{M_f}{2} \\ F_s^{-1}(s), & -\frac{M_f}{2} \le s \le \frac{M_f}{2} \\ T, & s > \frac{M_f}{2} \end{cases} \tag{18}$$

$$\overline{G}^{-1}(s) = \begin{cases} 0, & s < -\frac{M_g}{2} \\ G_s^{-1}(s), & -\frac{M_g}{2} \le s \le \frac{M_g}{2} \\ T, & s > \frac{M_g}{2} \end{cases} \tag{19}$$

260   In a transport sense, we can see from Fig. 9 that this amounts to transporting the excess (or deficit) mass to (or from) the boundaries, which act as a source or sink to maintain the mass balance. In this case with a temporal domain, the boundaries are the start and end of the time window. The excess or deficit mass is therefore associated with times before $t = 0$ or after $t = 500$, whichever is closer in time. This makes it an alternate implementation of the method discussed in Farchi et al. (2016), who also proposed using having sources or sinks at the boundaries of a spatial domain to complete the mass balance. Note that,

265 like the traditional one-dimensional Wasserstein distance, this is a result that can be computed efficiently without the need to solve a partial differential equation, linear program, or iterative scheme.

Now that the mapping and form of the inverse cumulative distributions have been proposed, the modified Wasserstein distance under the interpretation that it is the total transport cost can be defined according to Eq. (20).

$$\mathcal{HW}_2^2 = \int_{-\infty}^{\infty} \left[\overline{F}^{-1}(s) - \overline{G}^{-1}(s) \right]^2 ds \tag{20}$$

270 We choose to label the misfit Eq. (20) as the *hydrograph-Wasserstein distance*.

While Eq. (20) is expressed with infinite bounds, the integrand is only non-zero over the interval $[-M_{\max}/2, M_{\max}/2]$, where $M_{\max} = \max\{M_f, M_g\}$, so can easily be computed numerically over this range.

The adjustment of barycenters for distributions of arbitrary total mass is much simpler. We can use the scaling defined in Eq. (13) and find the barycenter of these distributions, then scale the result such that its total mass is the average of the composite
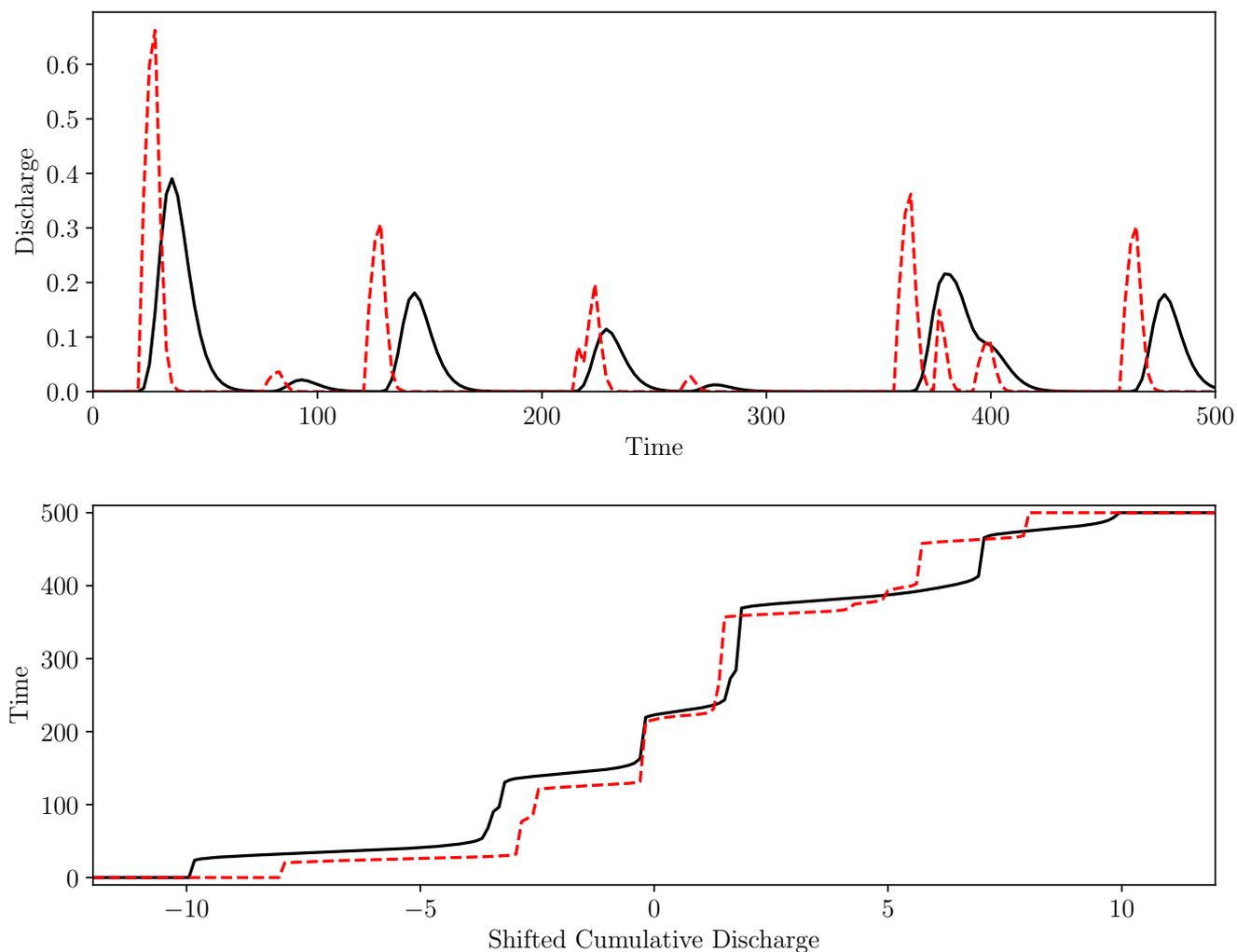
Figure 9. Visual representation of computing the hydrograph-Wasserstein distance. Top panel shows two synthetic hydrographs (dashed red and solid black) between which we want to compute the distance. Bottom panel shows the inverse shifted cumulative functions defined in Eq. (18) and Eq. (19). Transport occurs between points vertically aligned in this bottom panel, forcing excess mass to be mapped to the boundaries.

275 distributions. This is equivalent to replacing the Wasserstein distance in Eq. (7) with the penalised distance Eq. (14) under a quadratic penalty Eq. (15).

3 Rainfall-runoff modelling

3.1 Calibration

As a first application, the calibration of conceptual rainfall-runoff models using the Wasserstein distance as a minimisation
280 objective was considered. This experiment is similar to that found in Lerat and Anderssen (2015), but uses all model parameters in the hydrological model.

A conceptual rainfall-runoff model gauges the relationship between a rainfall time-series (hyetograph) and streamflow time series (hydrograph) for a particular watershed. By calibrating the model parameters to best capture this relationship, the model can be used for forecasting future streamflow under chosen rainfall conditions, or to infer properties of the watershed itself.
285 Automatic calibration methods seek to do this by minimising the discrepancy between the simulated streamflow from the rainfall-runoff model and the streamflow that was observed at gauging stations (Nash and Sutcliffe, 1970).

As hydrographs are a time-series, the efficient one-dimensional techniques described in Sect. 2.2 were used. The application of the Wasserstein distance to this problem bears some similarity to the technique of *dynamic time warping* (DTW), an application of which to hydrology can be found in Ouyang et al. (2010), the series distance developed in Ehret and Zehe (2011), the
290 cumulative distribution method in Lerat and Anderssen (2015), and the hydrograph matching technique in Ewen (2011). These also strive to allow timing errors to be better accounted for when comparing hydrographs.

As was used by Lerat and Anderssen (2015), a non-linear storage model was employed to test the misfit functions. The storage in the model evolves according to the differential equation (21).

$$\frac{dS}{dt} = m_3 r(t) - q(t) \tag{21}$$

295 The outward streamflow flux is given by the non-linear model Eq. (22).

$$q(t) = \left(\frac{S(t)}{m_1}\right)^{m_2} \tag{22}$$

Using some initial volume of water in storage, we can update the storage and compute the streamflow at each time step for any given set of model parameters and hyetograph. This gives three model parameters (m_1, m_2 and m_3) that must be calibrated using given rainfall and runoff measurements. To work in a controlled environment while testing the main characteristics of
300 the Wasserstein distance, synthetic rainfall events were used and the model used to generate the expected streamflow. We then attempted to recover the model parameters used to generate the synthetic observations.

As precipitation is a more uncertain measurement than streamflow, we subjected the synthetic rainfall event to timing errors, recalling that these types of errors are poorly represented when using point-wise misfit metrics (Ehret and Zehe, 2011). Figure 10 shows the true synthetic hyetograph, the observed hyetograph corrupted by timing errors, and the streamflow generated by
305 the true rainfall using the storage model defined in Eq. (21) and Eq. (22).
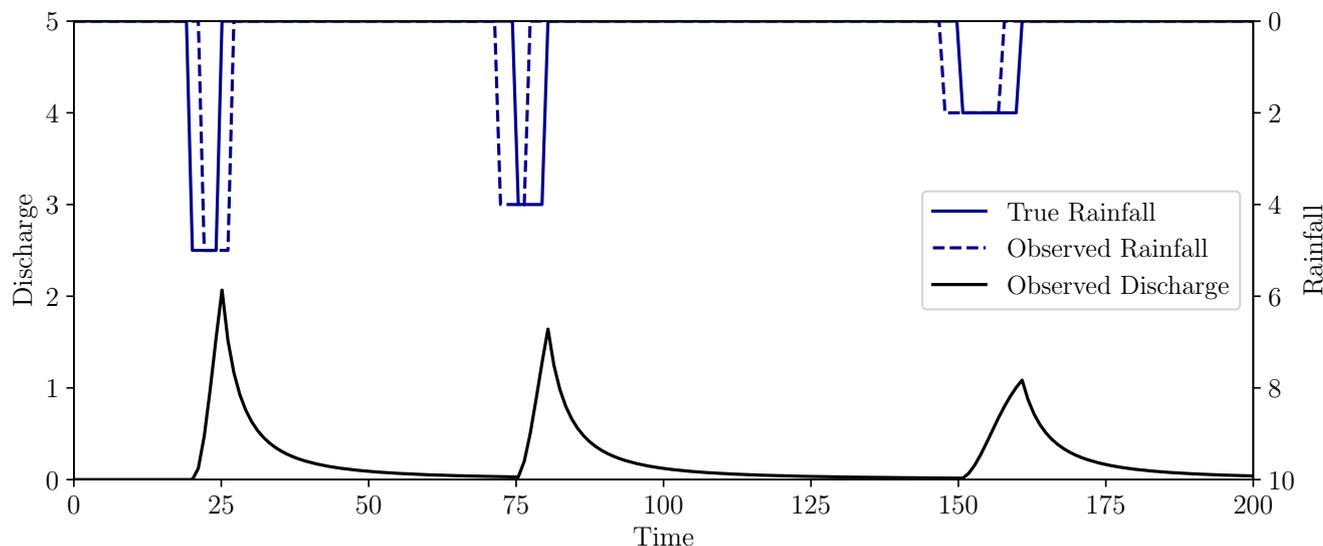
Figure 10. True rainfall (solid blue) with timing errors in observations (dashed blue), and observed hydrograph (black) produced from true rainfall and storage model described by Eq. (21) and Eq. (22) with model parameters $(10, 2, 0.7)$.

We then calibrated the three model parameters to this streamflow, using the erroneous rainfall measurements with a variety of misfit functions. Optimisation was performed with the RMSE, Wasserstein distance ($\gamma = 10, 100$) and the hydrograph-Wasserstein distance using the Nelder-Mead algorithm, implemented through the scipy Python package (Nelder and Mead, 1965; Virtanen et al., 2020). Figures 11, 12 and 13 show the misfit as a function of the model parameters, along with the true model parameters and the optimised parameters under each misfit.

The Wasserstein-based distances provided a misfit minimum closer to the true model parameters than the RMSE. A greater understanding of this can be garnered by examining the hydrographs produced by the calibrated model Fig. 14.

The Wasserstein-based distances better recovered the model parameters, and visually give a better hydrograph fit. The RMSE calibration underestimated the maximum flow peaks, in favour of more sustained lowered flows. The reason for this can be traced back to Fig. 1, where misaligned steep peaks gave large residuals. This can be lessened by having wider peaks, even though this gives a visually less satisfactory hydrograph and poorer estimates of model parameters. We can conclude that for this simple synthetic case study, that the Wasserstein distance has promising attributes with respect to timing errors in hydrograph fitting.

Of course, this only shows the efficacy of the Wasserstein distance for one particular rainfall event. Random synthetic rainfall events were therefore generated using the method described in Appendix A, with the model calibrated to each event. The calibrated model parameters are compiled into boxplots in Fig. 15.

Whilst still having some variability and bias, the Wasserstein-based distances performed significantly better across the 500 trials with a median result closer to the true value for all parameters and lowered variability. The ordinary Wasserstein distance
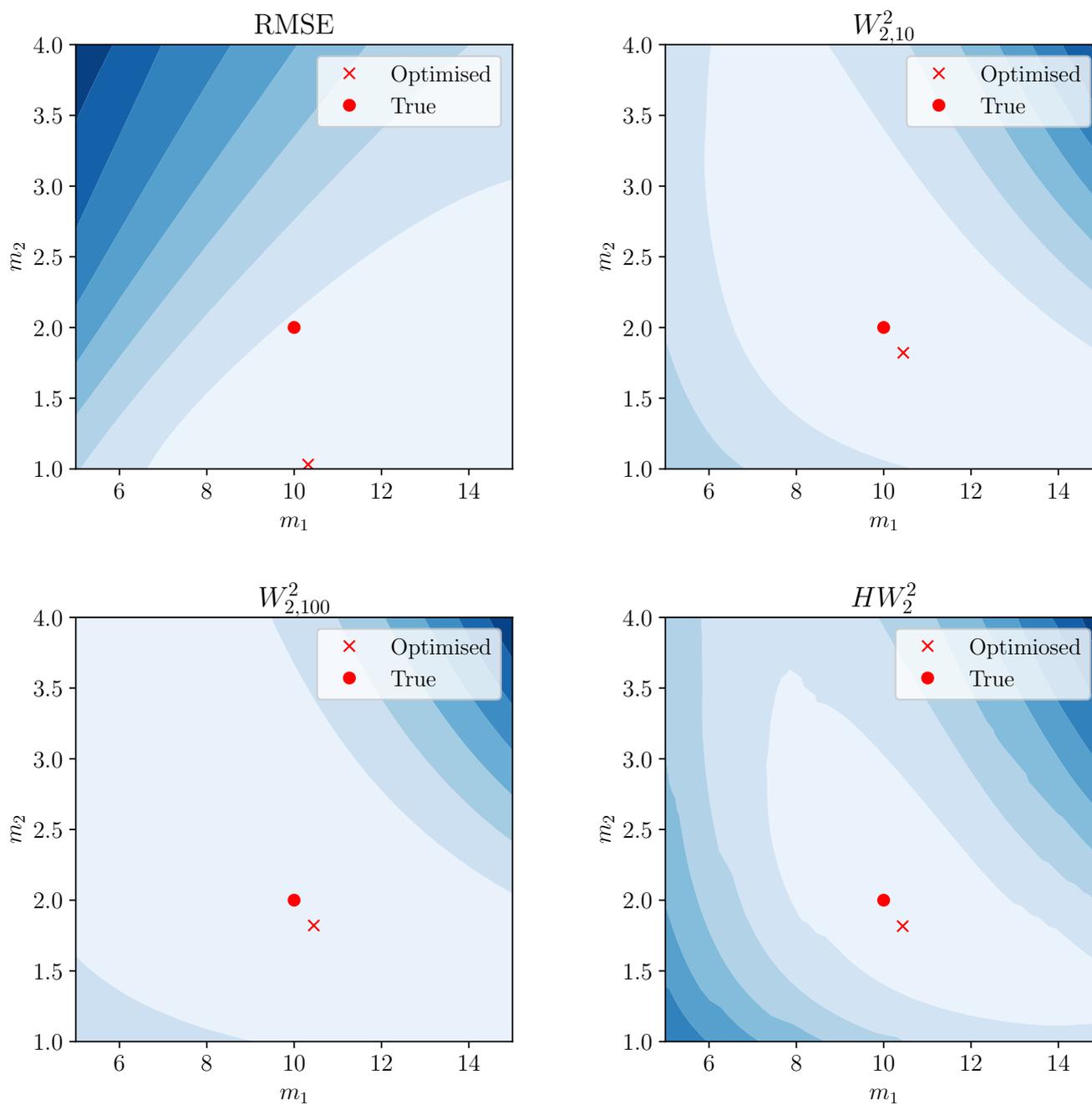
Figure 11. Misfit surface as a function of m_1 and m_2 for each objective. Red circle marker represents true model parameters and red cross marker represents minimised solution from Nelder-Mead algorithm.
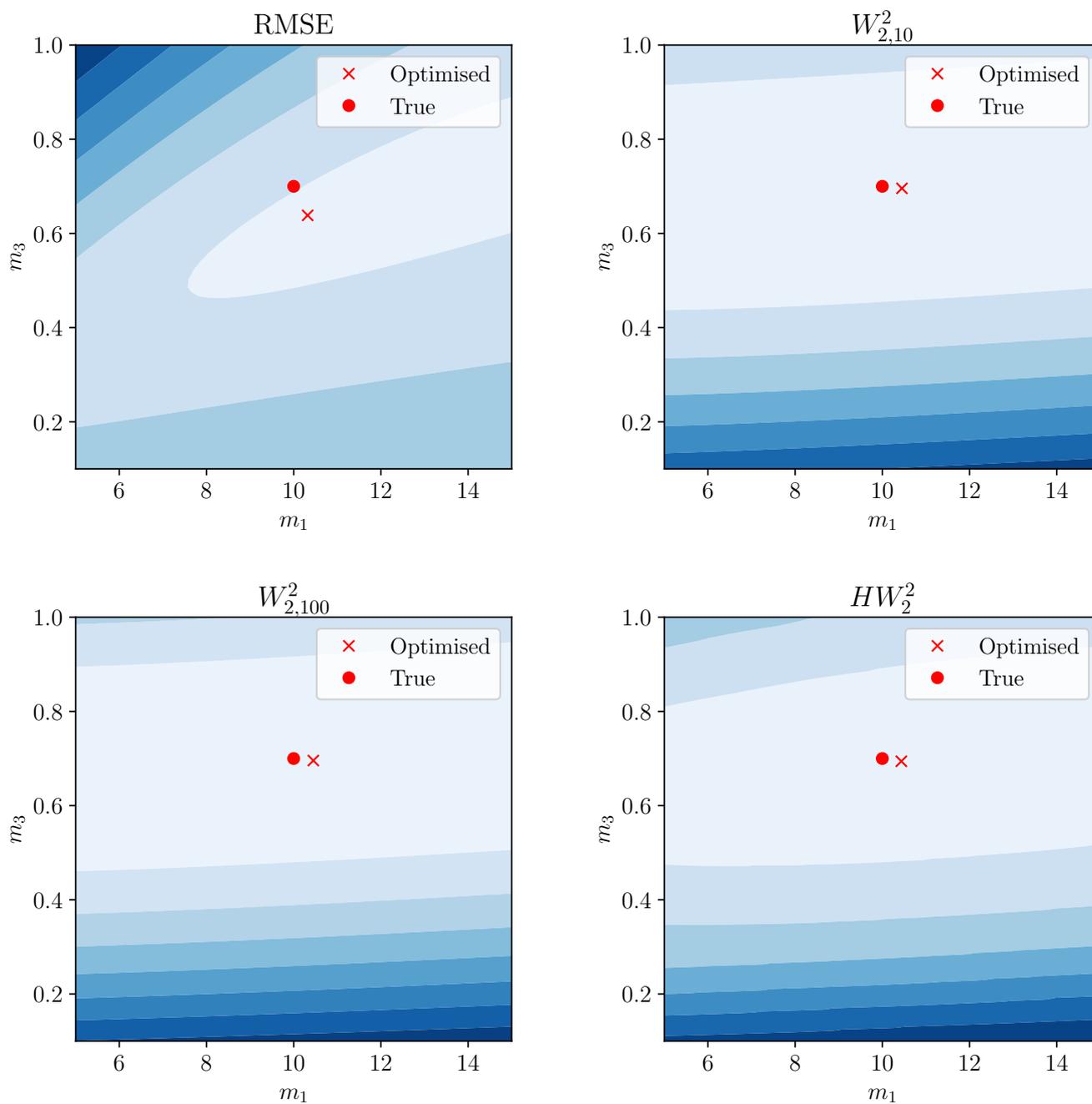
Figure 12. The same display is Fig. 11, but considering model parameters m_1 and m_3. Red circle still denotes true parameters and red cross the minimum of this surface.
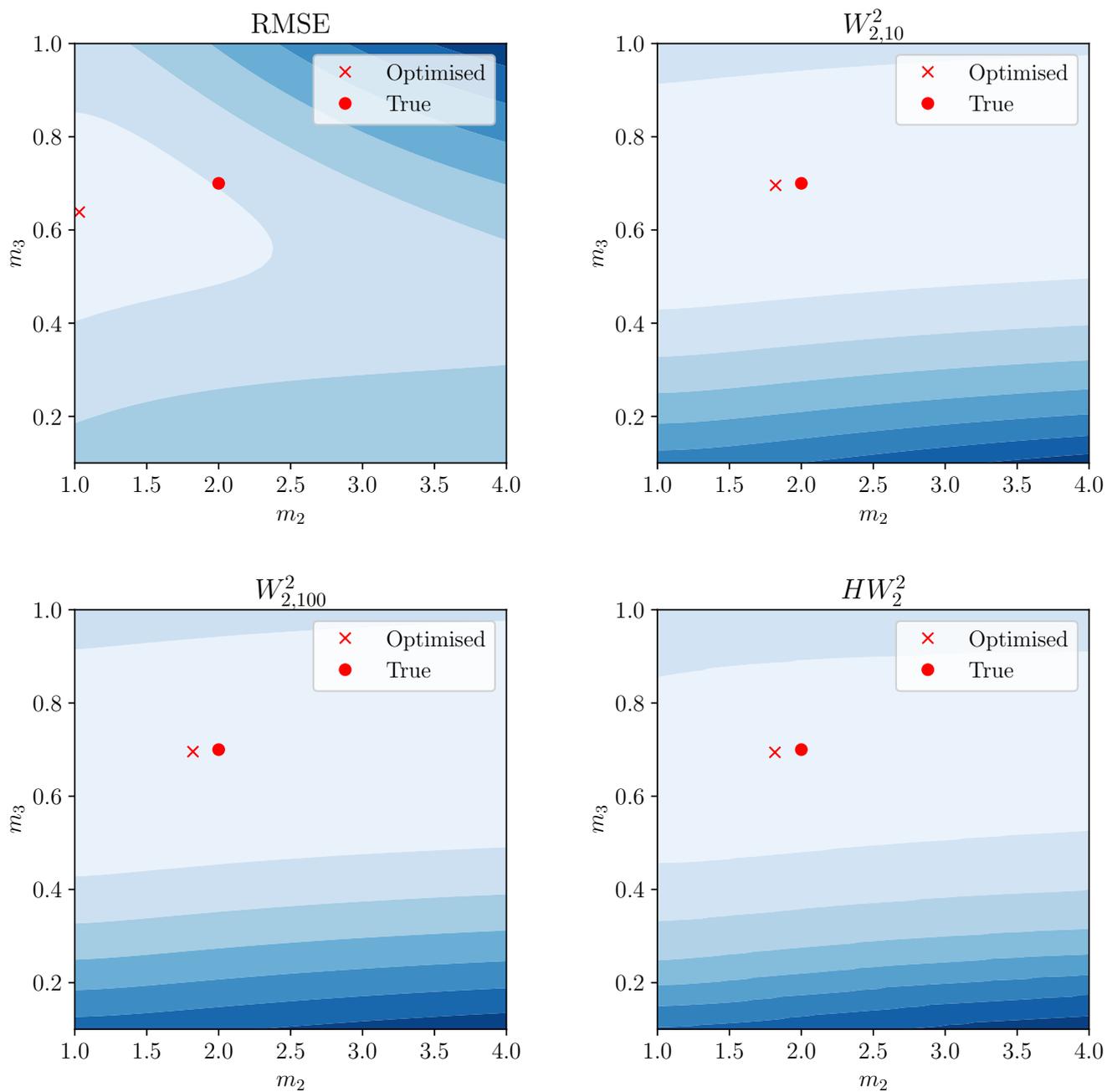
Figure 13. The same display is Fig. 11, but considering model parameters m_2 and m_3. Red circle still denotes true parameters and red cross the minimum of this surface.
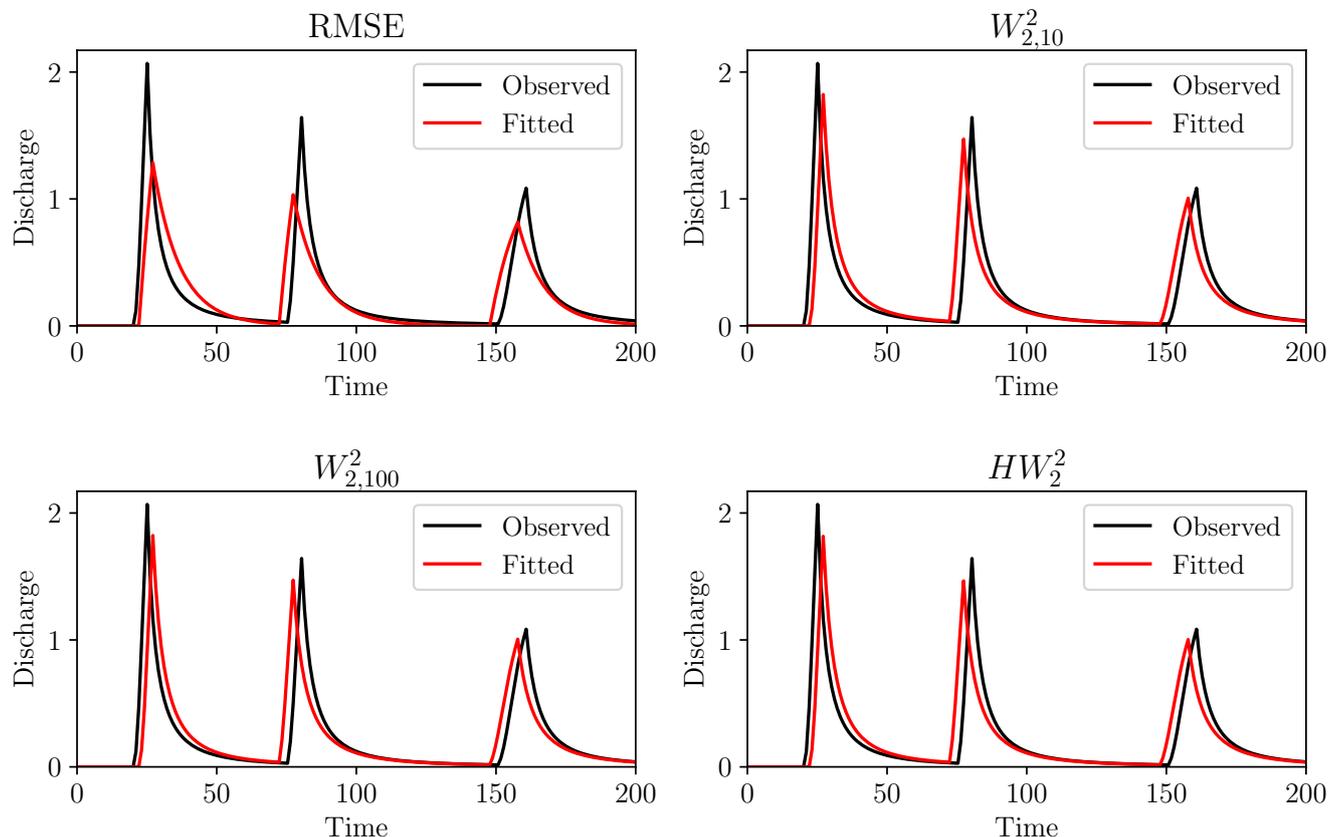
Figure 14. Output of rainfall-runoff model for calibrated parameters under each objective function. Observations displayed in black, and simulation in red.

outperformed the hydrograph-Wasserstein distance for this model. The size of the penalty weighting factor γ had little effect
325 on the location of the optimum for the Wasserstein distance, but altered the shape of the misfit surface (see Figs. 11,12 and 13).

The better performance of the penalised Wasserstein distance compared to the hydrograph-Wasserstein distance may be due to the form of this model. As m_1 and m_2 purely control hydrograph shape and m_3 purely controls total output mass, there is a complete separation in which model parameters affect which aspect of misfit; that is, no model parameter simultaneously influences both the Wasserstein distance *and* the mass balance penalty term unless the change in m_1 or m_2 moves mass outside
330 of the measured time window.

Subsequently, there is no 'trade-off' between lowering the Wasserstein distance or penalty term; both are always possible as the model parameters only influence both if mass is pushed outside the time window (hence there being *any* variability in the m_3 panel of Fig. 15).
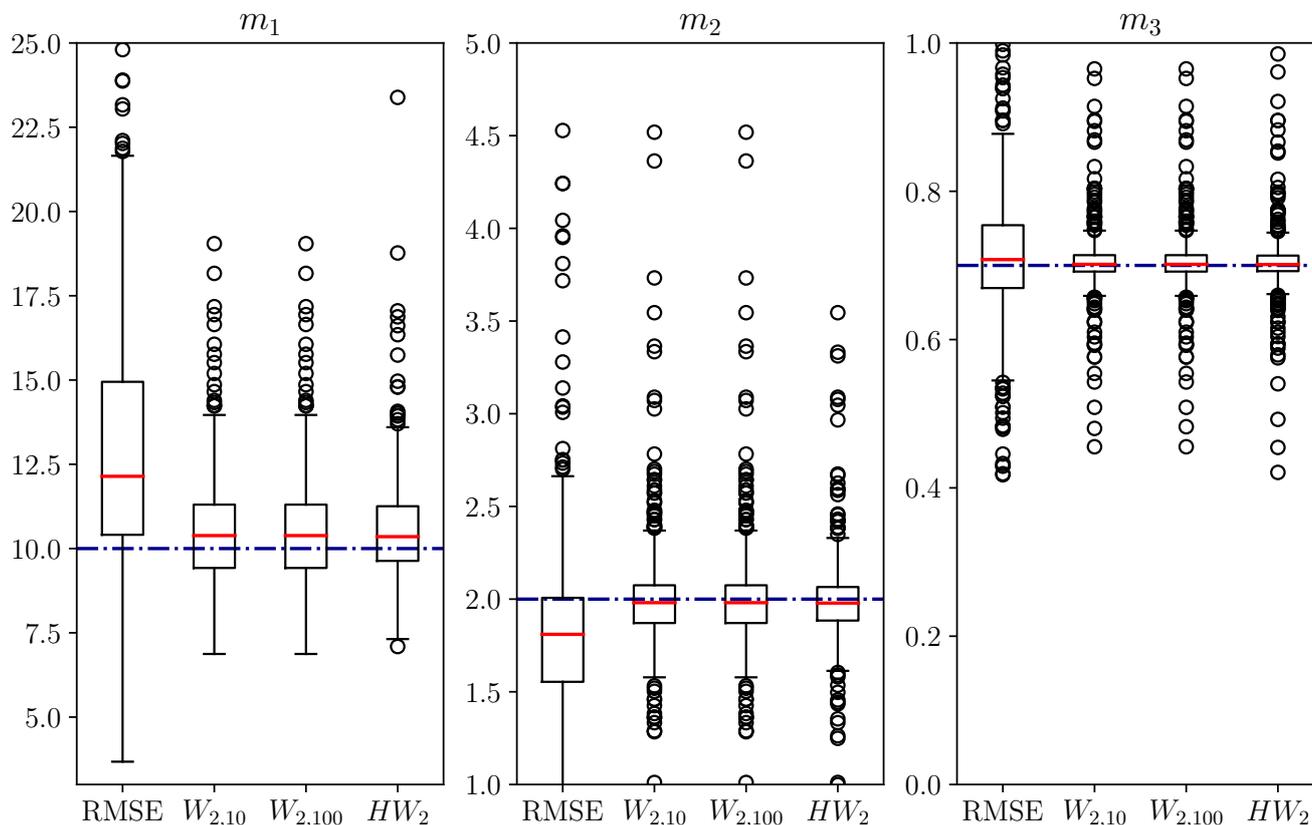
Figure 15. Calibrated model parameters for 500 rainfall events generated randomly using method in Appendix A under the influence of timing errors. Optimisation used Nelder-Mead algorithm initialised at the true model parameters of $(10, 2, 0.7)$. Red solid lines shows median calibrated parameter, blue dash-dot line is true model parameters. Optimisation used RMSE, Wasserstein distance with $\gamma = 10, 100$ penalty terms, and hydrograph-Wasserstein distance.

If there were model parameters that influenced both shape *and* total mass, this separation of aspects would not be present

335 and the magnitude of γ may have increased importance due to the trade-off between fitting the shape and balancing the mass of the hydrographs when tuning a model parameter.

3.2 Instantaneous unit hydrograph model

Section 3.1 showed that the Wasserstein distance gives a better account for timing errors than can be made from point-wise metrics. The effect of delay times from within the hydrological model itself is now considered. A simple unit hydrograph model

340 was used to generate such delays between rainfall and the associated streamflow. The instantaneous unit hydrograph (IUH) is defined as the streamflow response for a unit impulse of effective rainfall over the catchment. Under the assumption of a linear
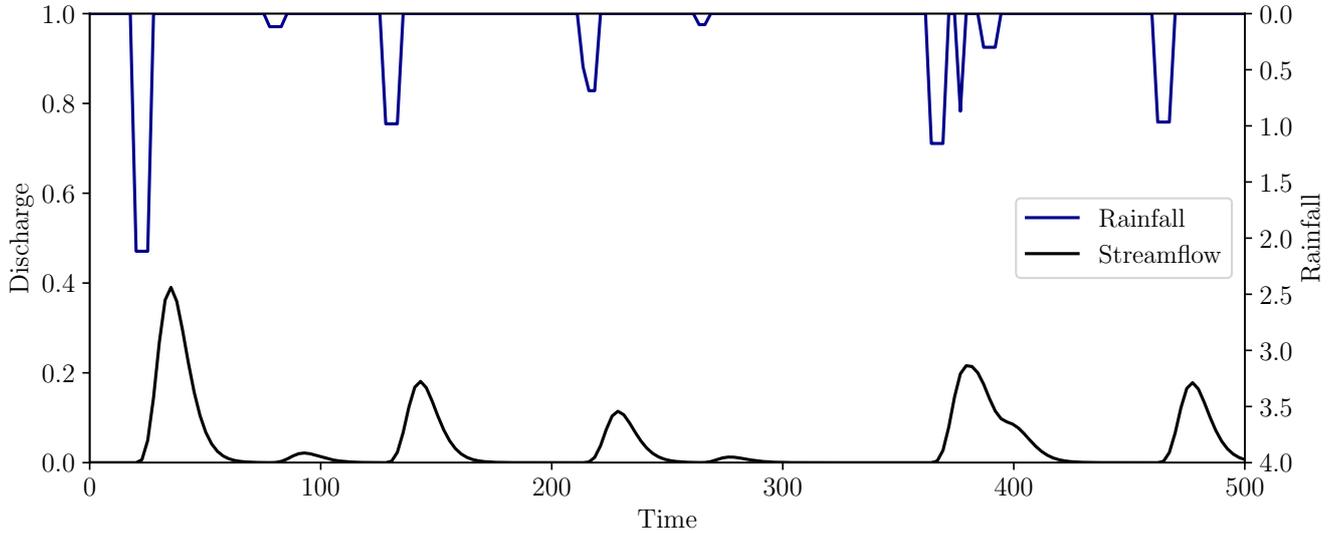
Figure 16. Effective rainfall (top, blue) and observed streamflow (bottom, black) for the test problem. Rainfall was generated using method described in Appendix A with 10 rainfall events of maximum length 5 time units. The streamflow has generated using the unit hydrograph model Eq. (24) with model parameters $(3, 5)$.

response, the stream hydrograph can be modelled via a convolution of the rainfall and IUH, giving the model Eq. (23).

$$q(t) = \int_{-\infty}^{\infty} u(t-s)r(s)\,ds = u(t) * r(t) \tag{23}$$

In reality, we can only measure the streamflow and rainfall, leaving us to infer the IUH from such measurements. To enforce

345 some set of desired characteristics on the IUH, such as smoothness, positivity, and unit mass, it can be restricted to belonging to a parametric set, each of which is of the correct form. A commonly used synthetic unit hydrograph is the Nash hydrograph (Nash, 1957). This assumes the watershed is composed of a series of linear reservoirs, giving a unit hydrograph of the form Eq. (24), where Γ is the Gamma function.

$$u(t) = \frac{1}{\Gamma(k)\theta^k} t^{k-1} e^{-t/\theta} \tag{24}$$

350 This leaves us with two parameters, θ and k, to calibrate. Although there are multiple existing ways of solving this problem, so the OT approach is not necessarily warranted, this application exhibits some of the key features we would like to explore more generally in rainfall-runoff model calibration whilst keeping the model structure simple.

The effective rainfall time-series was generated using the method described in Appendix A with 10 rainfall events, giving the output displayed in Fig. 16, where the true model parameters of the IUH are $\theta = 3, k = 5$.

355 The behaviour of both the Wasserstein distance and RMSE were explored here with respect to the time delays induced by the IUH model. The misfit as a function of the two model parameters is shown in the left panels of Fig. 17.
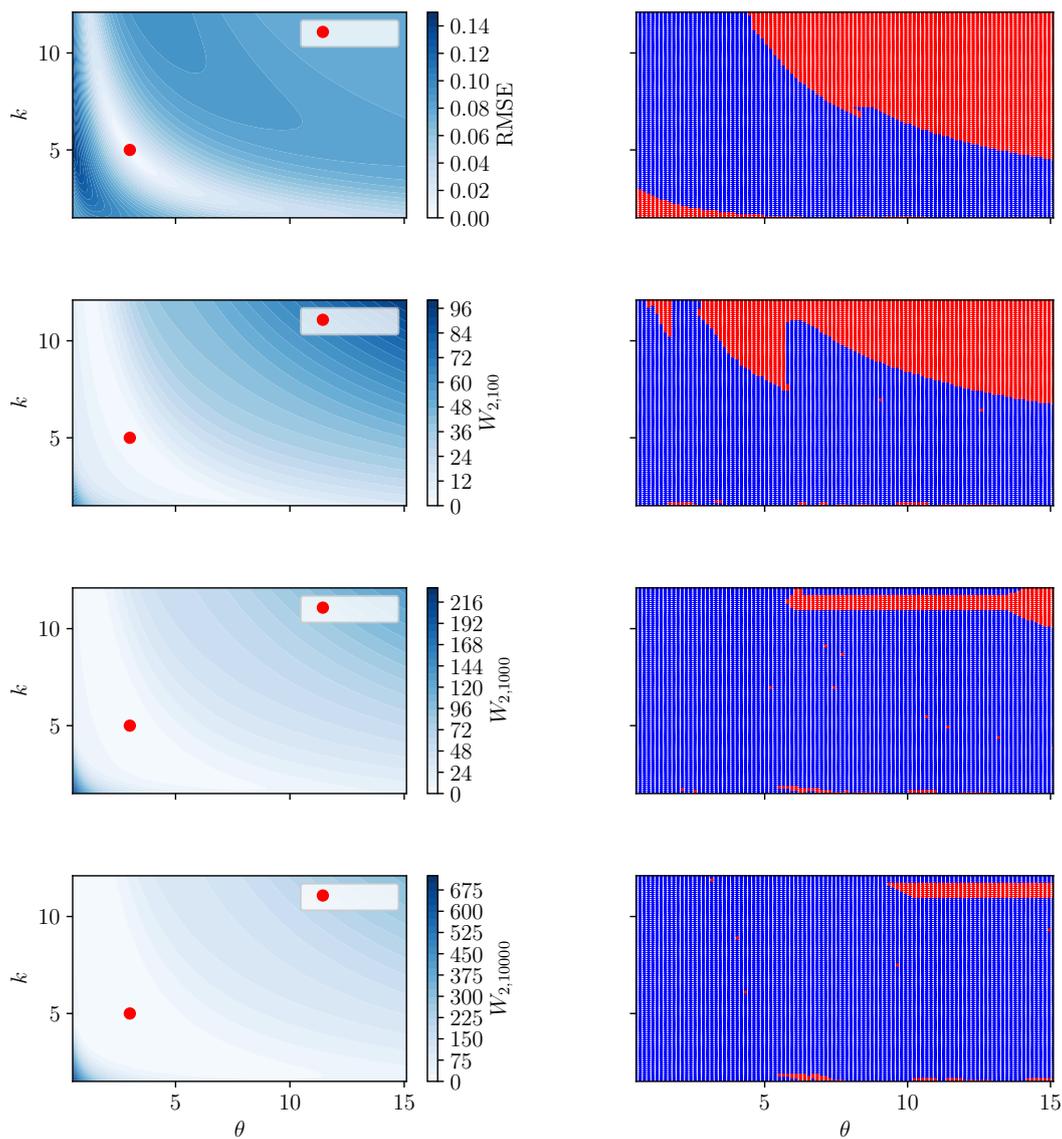
Figure 17. Root mean square error and 2-Wasserstein (with $\gamma = 100, 1000, 10000$) misfit surfaces for rainfall event in Fig. 16 and unit hydrograph model (24) (left panels). Red marker shows true solution. The RMSE surface displays local minima in the lower left and upper right corners where poor convergence is observed. Right panels show if convergence criteria was met (blue) or not (red) for each model initialisation.

The key difference between the misfit surfaces in Fig. 17 is that local optimisation methods will converge for a greater number of starting locations towards the true optimal point (marked red in left panels) when using the Wasserstein-based distance. While many initial estimates with the RMSE will also converge to this same point, there also is alternative minimal points to which the optimisation scheme may also converge. A grid of starting estimates of the model parameters was used to examine this more concretely in the right panels of Fig. 17. These panels show the convergence properties according to the initial model parameter estimate for the RMSE and the Wasserstein distance for three choices of γ. Convergence is defined as the optimised model $\boldsymbol{m}^*$ satisfying $\|\boldsymbol{m}^* - \boldsymbol{m}_{\text{true}}\|_2 < 0.001$, where $\boldsymbol{m}_{\text{true}}$ is the true solution, denoted as the red point in the left panels of Fig. 17.

From Fig. 17, there is a clearly defined 'basin of attraction' about the optimal solution for RMSE, beyond which local optimisation algorithms will not converge. This matches well with the corresponding misfit surface in the left panel. The Wasserstein distance converged for a greater number of initial estimates, with improvement seen for a larger choice of γ. This is likely due to the low sensitivity of the misfit surface within the 'valley' about the optimal solution (visualised as white region in Fig. 17). The larger value of γ increases the sensitivity (gradient of the misfit surface), and potentially prevents early termination of the optimisation algorithm.

We do acknowledge that it takes a relatively poor initial estimate of the model parameters to become trapped in a local minimum for this IUH model. Indeed, with only two model parameters, a brute force global optimisation method is certainly feasible. Furthermore, a good initial estimate may be generated using the method of moments for the Nash IUH. However, these results show the ability of the Wasserstein distance to recognise the error in improperly aligned peak flows.

The power of the Wasserstein distance may come to the fore when a more complex watershed and thus model is used, perhaps producing multiple pulses per rainfall event and yielding multiple peaks in the IUH. These results also capture the general behaviour of models possessing significant delay times, allowing the simulated hydrograph to 'shift' in time across the observations, in a similar manner to Fig. 3.

3.3 Hydrograph barycenters

Beyond the use of the Wasserstein distance purely as an objective function, attention will now be turned to the application of Wasserstein barycenters to hydrology, using hydrographs as the object of study. The power of the Wasserstein barycenter is that it gives a notion of 'average' when features have been displaced.

This means that an ensemble of hydrographs describing the same event, perhaps with different climate or hydrological models, can be 'averaged' into a single hydrograph that carries characteristics of each ensemble member. Note that this will only be true if the main source of difference between ensemble members is timing and peak shape differences, rather than peak volume differences.

To test this premise, two different IUH models were applied to the same synthetic rainfall event, giving the differing hydrographs shown in the top panel of Fig. 18. As they were from the same rainfall event and a unit hydrograph model was used, the volume of the peaks in each hydrograph were approximately the same, with only the shape and timing differing, so suitable for the Wasserstein approach. The bottom panel of Fig. 18 compares the Wasserstein barycenter with the mean of the hydrographs.
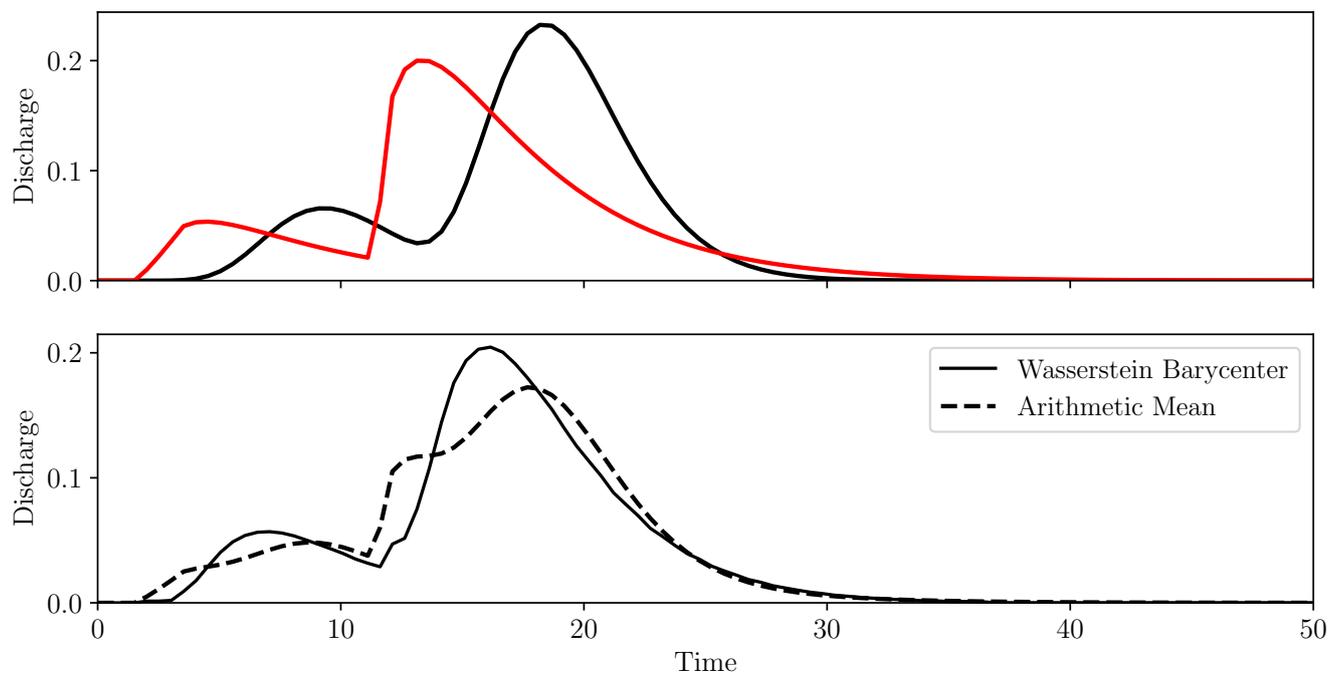
Figure 18. Two hydrographs for the same rainfall event but with different hydrological models (top panel, black and red solid lines). Bottom panels shows the Wasserstein barycenter in solid black and mean in dashed black.

Unlike the ordinary mean, the Wasserstein distance captured the correct number and general characteristics of the peak flows. Again, this is built on the assumption that the peaks of each hydrograph mainly differ in shape and timing, but not volume. If they differed in volume, the barycenter could exhibit peaks in unusual locations, as the mass is being 'viewed' halfway through transport across the domain.

Consider the two hydrographs shown in the top panel of Fig. 19. They are the same in timing, but differ in the volume of water. In this case, the regular mean gave a much better notion of average, as the difference between the hydrographs is better explained in terms of amplitude, rather than displacement as was the case in Fig. 18.

We therefore see that while the Wasserstein barycenter is well suited for an ensemble of hydrographs with differing peak flow timings, it is not well suited for amplitude differences. This is to be expected from its definition. We therefore suggest that the conventional mean is best suited for an ensemble with differing amplitudes but consistent timings, but the Wasserstein barycenter when timing is variable but volumes in peak flows are consistent.

4 Conclusions

Quantifying the similarity of temporal or spatial distributions of water occupies an important role within hydrology. The way in which a 'good' fit is defined directly influences the character of the results. Many commonly used misfit functions, such as root
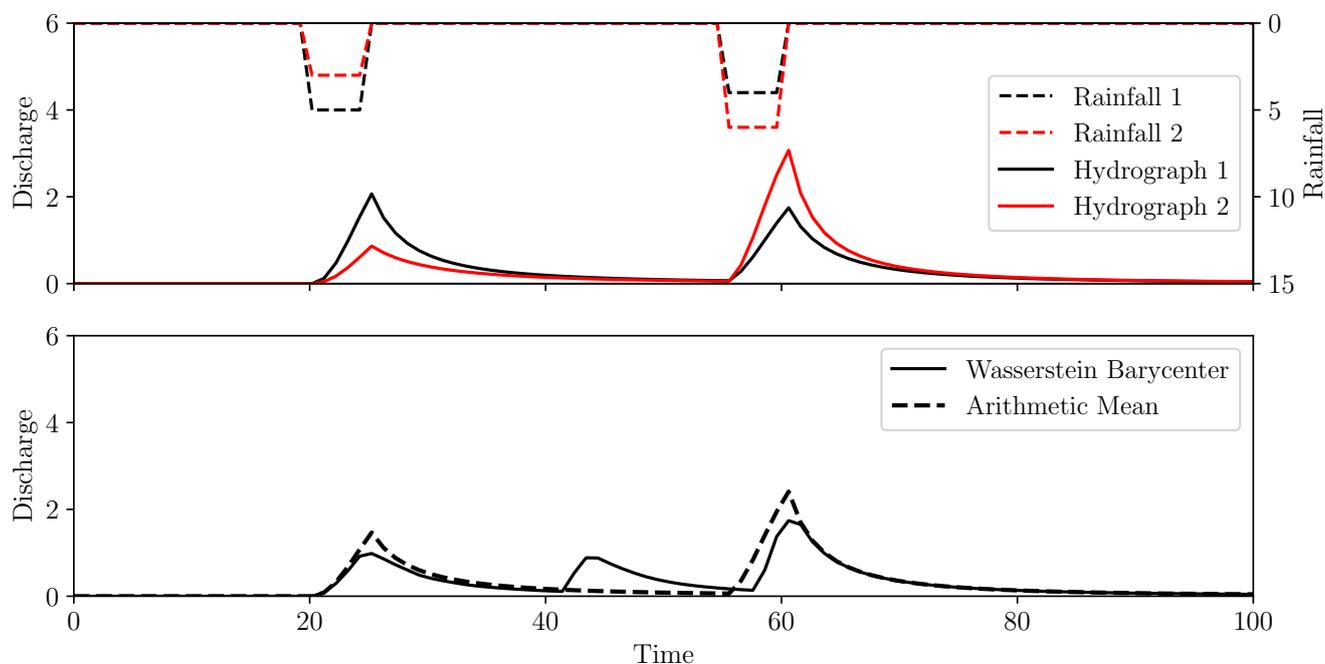
Figure 19. The same hydrological model applied to rainfall events differing in amplitude (dashed black and red, inverted plots in top panel), giving hydrograph differences purely in amplitude (solid black and red, top panel). Bottom panel shows mean (blue) and Wasserstein barycenter (black).

405   mean square error, and Nash-Sutcliffe efficiency, quantify fit by considering differences in amplitude. While this is perfectly acceptable when errors are restricted to amplitude, these measures of misfit do not well quantify displacement errors.

In this work, we have suggested the Wasserstein distance, derived from optimal transport, as a candidate misfit function for applications where displacement or timing errors are prominent. This quantifies difference in terms of the effort required to transform one mass distribution into another. A modification of traditional OT and the associated 'hydrograph-Wasserstein

410   distance' was also developed.

While the Wasserstein-based measures certainly gave better calibration results than RMSE, there was still a slight bias towards peaks of reduced amplitudes, although to a much lesser extent. Some type of multi-objective method may circumvent this, by using a measure comparing the amplitude of peak flows. Although the Wasserstein distance with a penalty term outperformed the hydrograph-Wasserstein distance for these synthetic tests, there may be broader implications for this second

415   misfit function. Very little interest has been given thus far to the modification of OT for particular applications, with most preferring to force data into the density function mould. Other ways of modifying, whilst still capturing the essence of OT are therefore a key point of further research, as they may allow the simultaneous capture of displacement and total mass errors without the need for a user-defined weighting term between the aspects.

As proposed by Ehret and Zehe (2011), a major drawback of the Wasserstein distance for hydrographs is the mapping of
420   water masses between functionally different parts of the hydrographs. For example, if the modelled volume of a particular
storm event exceeds the observations, this excess water is mapped to the next (and unrelated) storm event. If the storm events
are separated by long dry periods, a large transport cost would be incurred by moving the mass between the events. Therefore,
if fitting multiple events interspersed with dry spells, it may be more appropriate to segment the hydrographs into each discrete
event, and compute the Wasserstein distance individually within each time window to prevent this mapping between events.
425   The objective function would then become the summation of the Wasserstein distances for each time segment. This would not
be possible for wet periods with overlapping rainfall events, but the penalty for transporting mass between events would be
much smaller in this case as they occur closer together in time.

It is also important to remember that for more complex models and data, we cannot expect the Wasserstein distance to
capture *all* aspects of fit. Much like the amplitude-based metrics, it prioritises particular aspects of fit and has the potential to
430   yield large amplitude errors at specific points, as long as mass has not been shifted far from the target location. We are therefore
led back to our introductory question of *what is a good 'fit'?* If we know, or expect, timing or spatial displacement errors in
our measurements or model output, a transport-based definition of fit would seem appropriate. However, this may not be the
case when other types of errors are at play. A multi-objective approach may therefore still be a necessary, and even encouraged
technique. The results here do however show that the Wasserstein distance is suitable for quantifying displacement errors, so a
435   valuable addition to the toolbox of misfit functions in operation in hydrology.

Code availability. Code is available at https://github.com/JMagyar15/wasserstein_hydro

Appendix A: Random rainfall generation

For the synthetic experiments, a method for generating random rainfall events was required. This is done by first setting a time
window length. Let this window be the interval $[0, T]$, where any units can be used.
440   A number of rainfall events N is then specified by the user. The start time of each of these rainfall events is chosen from a
random uniform distribution within the time window. That is, if the start time of the i-th event is denoted s_i, then

$$s_i \sim U(0, T)$$

The length of each of these rainfall events is then chosen from a uniform distribution of specified maximum length t_{max}, so
the length of the i-th event is given by

445   $$l_i \sim U(0, t_{max})$$

Together, these give a start and end time for each storm, where the end time e_i is given by

$$e_i = s_i + l_i$$

The intensity of the rainfall r_i for each storm is chosen independently from an exponential distribution with rate parameter λ.

$$r_i \sim \text{Exponential}(\lambda)$$

450   In this study, we set $\lambda = 1$, as the rainfall units are arbitrary.

To corrupt the measurements with timing errors, the timing of each storm is shifted independently by a number δ_i drawn from a uniform distribution ranging between a positive and negative maximum time shift, δ_{max}, so

$$\delta_i \sim U(-\delta_{max}, \delta_{max})$$

This gives the altered start and end times for each storm as

455   $$s_i' = s_i + \delta_i$$
$$e_i' = e_i + \delta_i$$

Between the start and end time of each rainfall event, constant rainfall of the randomly chosen intensity is applied. If rainfall events overlap, the intensities are added for this cross-over period.

Author contributions. J.M. wrote the manuscript and code for this paper. J.M. and M.S. both developed the theory of the work. M.S. assisted
460   in editing and preparation of the manuscript, and supervised the project.

Competing interests. The authors declare no competing interests.

Acknowledgements. J.M. completed this work under the support of an InLab honours scholarship funded from CSIRO Future Science
platform for Deep Earth Imaging. The authors would also like to acknowledge helpful discussions with Andrew Valentine on this topic.

References

Ambrosio, L., Gigli, N., and Savaré, G.: Gradient flows: in metric spaces and in the space of probability measures, Springer Science & Business Media, 2008.

Benamou, J.-D., Froese, B. D., and Oberman, A. M.: Numerical solution of the Optimal Transportation problem using the Monge–Ampère equation, Journal of Computational Physics, 260, 107–126, https://doi.org/https://doi.org/10.1016/j.jcp.2013.12.015, 2014.

Bonneel, N., Rabin, J., Peyré, G., and Pfister, H.: Sliced and Radon Wasserstein Barycenters of Measures, Journal of Mathematical Imaging and Vision, 51, 22–45, https://doi.org/10.1007/s10851-014-0506-3, 2015.

Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions, Communications on Pure and Applied Mathematics, 44, 375–417, https://doi.org/https://doi.org/10.1002/cpa.3160440402, 1991.

Cuturi, M.: Sinkhorn Distances: Lightspeed Computation of Optimal Transport, in: Advances in Neural Information Processing Systems, vol. 26, 2013.

Ehret, U. and Zehe, E.: Series distance - An intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events, Hydrology and Earth System Sciences, 15, 877–896, https://doi.org/10.5194/hess-15-877-2011, 2011.

Engquist, B. and Froese, B. D.: Application of the Wasserstein Metric to Seismic Signals, Communications in Mathematical Sciences, 12, 979–988, https://doi.org/10.4310/CMS.2014.v12.n5.a7, 2014.

Engquist, B., Hamfeldt, B., and Yang, Y.: Optimal Transport for Seismic Full Waveform Inversion, Communications in Mathematical Sciences, 14, https://doi.org/10.4310/CMS.2016.v14.n8.a9, 2016.

Ewen, J.: Hydrograph matching method for measuring model performance, Journal of Hydrology, 408, 178–187, https://doi.org/10.1016/j.jhydrol.2011.07.038, 2011.

Farchi, A., Bocquet, M., Roustan, Y., Mathieu, A., and Quérel, A.: Using the Wasserstein distance to compare fields of pollutants: Application to the radionuclide atmospheric dispersion of the Fukushima-Daiichi accident, Tellus, Series B: Chemical and Physical Meteorology, 68, https://doi.org/10.3402/tellusb.v68.31682, 2016.

Feyeux, N., Vidard, A., and Nodet, M.: Optimal transport for variational data assimilation, Nonlinear Processes in Geophysics, 25, 55–66, https://doi.org/10.5194/npg-25-55-2018, 2018.

Genevay, A., Cuturi, M., Peyré, G., and Bach, F.: Stochastic Optimization for Large-Scale Optimal Transport, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, p. 3440–3448, 2016.

Kantorovich, L. V.: On the translocation of masses, Dokl. Akad. Nauk. USSR (NS), 37, 199–201, 1942.

Lerat, J. and Anderssen, R. S.: Calibration of hydrological models allowing for timing offsets, in: 21st International Congress on Modelling and Simulation, MODSIM 2015, pp. 126–132, https://doi.org/10.36334/modsim.2015.a2.lerat, 2015.

Li, L., Vidard, A., Le Dimet, F. X., and Ma, J.: Topological data assimilation using Wasserstein distance, Inverse Problems, 35, https://doi.org/10.1088/1361-6420/aae993, 2019.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J.: Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, Geophysical Journal International, 205, 345–377, https://doi.org/10.1093/gji/ggw014, 2016.

Monge, G.: Mémoire sur la théorie des déblais et des remblais., Histoire de l'Académie Royale des Sciences de Paris, 1781.

Nash, J. E.: The form of the Instantaneous Unit Hydrograph, International Association of Hydrological Sciences, 3, 114–121, 1957.

500   Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nelder, J. A. and Mead, R.: A Simplex Method for Function Minimization, The Computer Journal, 7, 308–313, https://doi.org/10.1093/comjnl/7.4.308, 1965.

Ouyang, R., Ren, L., Cheng, W., and Zhou, C.: Similarity search and pattern discovery in hydrological time series data mining, Hydrological
505   Processes, 24, 1198–1210, https://doi.org/10.1002/hyp.7583, 2010.

Peyré, G. and Cuturi, M.: Computational Optimal Transport: With Applications to Data Science, Foundations and Trends® in Machine Learning, 11, 355–607, https://doi.org/10.1561/2200000073, 2019.

Rabin, J., Peyré, G., Delon, J., and Bernot, M.: Wasserstein Barycenter and Its Application to Texture Mixing, in: International Conference on Scale Space and Variational Methods in Computer Vision, pp. 435–446, 2011.

510   Read, A. L.: Linear interpolation of histograms, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 425, 357–360, https://doi.org/10.1016/S0168-9002(98)01347-3, 1999.

Roberts, N. M. and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Monthly Weather Review, 136, 78–97, https://doi.org/10.1175/2007MWR2123.1, 2008.

Sambridge, M., Jackson, A., and Valentine, A. P.: Geophysical inversion and optimal transport, Geophysical Journal International, 231,
515   172–198, https://doi.org/10.1093/gji/ggac151, 2022.

Seguy, V., Damodaran, B. B., Flamary, R., Courty, N., Rolet, A., and Blondel, M.: Large-Scale Optimal Transport and Mapping Estimation, in: International Conference on Learning Representations, pp. 1–15, 2018.

Solomon, J., de Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., and Guibas, L.: Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains, ACM Trans. Graph., 34, https://doi.org/10.1145/2766963, 2015.

520   Villani, C.: Optimal Transport: Old and New, Springer Verlag, 2009.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, , Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner,
525   A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G. L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk,
530   O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Wernli, H., Paulat, M., Hagen, M., and Frei, C.: SAL - A novel quality measure for the verification of quantitative precipitation forecasts,
535   Monthly Weather Review, 136, 4470–4487, https://doi.org/10.1175/2008MWR2415.1, 2008.

Yang, Y., Engquist, B., Sun, J., and Hamfeldt, B. F.: Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, GEOPHYSICS, 83, R43–R62, https://doi.org/10.1190/geo2016-0663.1, 2018.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, Journal of Hydrology, 204, 83–97, https://doi.org/10.1016/S0022-1694(97)00107-8, 1998.