Towards inverse modeling of landscapes using the Wasserstein distance

Matthew J Morris[1], Alex G Lipp[2], and Gareth G Roberts[1]

[1]Imperial College London
[2]Merton College, Oxford University

July 8, 2023

Towards inverse modelling of landscapes using the Wasserstein distance

M. J. Morris[1], A. G. Lipp[2], G. G. Roberts[1]

[1]Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK
[2]Merton College, University of Oxford, Oxford, OX1 4JD, UK

Key Points:

- The use of the Wasserstein distance for identifying optimal landscape evolution models is demonstrated.
- This approach can produce simple objective functions, simplifying the search for models that minimise data misfit.
- Accurate amplitudes and locations of uplift can be retrieved from synthetic landscapes generated using different initial conditions.

Corresponding author: M. J. Morris, `matthew.morris15@imperial.ac.uk`

Abstract

Extricating histories of uplift and erosion from landscapes is crucial for many branches of the Earth sciences. An objective way to calculate such histories is to identify calibrated models that minimise misfit between observations (e.g. topography) and predictions (e.g. synthetic landscapes). In the presence of natural or computational noise, widely used Euclidean measures of similarity can have complicated objective functions, obscuring the search for optimal models. Instead, we introduce the Wasserstein distance as a means to measure misfit between observed and theoretical landscapes. Our results come in two parts. First, we show that this approach can generate much smoother objective functions than Euclidean measures, simplifying the search for optimal models. Second, we show how locations and amplitudes of uplift can be accurately recovered from synthetic landscapes even when seeded with different noisy initial conditions. We suggest that this approach holds promise for inverting real landscapes for their histories.

Plain Language Summary

The shapes of Earth's landscapes tell us about how they were formed by processes like tectonic uplift and erosion. Mathematical models are used to predict how landscapes change over time due to these processes. However, identifying models that produce theoretical landscapes that resemble reality can be challenging. One way to do so is by comparing model predictions to actual landscapes we observe. To make this comparison, we need a way to measure how similar or different predicted and observed landscapes are. One common approach is to compare heights of land from both cases. However, this method can struggle because a small shift in the position of a theoretical valley, say, can dramatically change the outcome of a comparison. In this paper, we introduce an alternative approach that uses a metric called the Wasserstein distance from the field of 'Optimal Transport.' The Wasserstein distance is a measure of how different two probability distributions are from each other by considering how much 'work' is needed to transform one distribution into the other. We show that this metric is effective for finding models to understand how landscapes were shaped by uplift over time.

1 Introduction

Planetary surface topography is shaped by geologic and geomorphic processes operating on a broad range of spatial and temporal scales (e.g. Davis, 1899; Bishop, 2007; Anderson & Anderson, 2010; Wapenhans et al., 2021). A general goal is to identify geologic and geomorphic models that can accurately predict observed landscapes. For example, a suite of inverse models have been developed to identify uplift rate histories that yield low residual misfits to longitudinal river profiles (e.g. Pritchard et al., 2009; Roberts & White, 2010; Goren et al., 2014; Gallen & Fernández-Blanco, 2021). Despite increased computational expense, forward and inverse modeling of two dimensional landscapes can incorporate geomorphic information that is not captured by river profiles (e.g. Croissant & Braun, 2014; Barnhart, Tucker, et al., 2020; O'Malley et al., 2021). They can also be used to relax assumptions about drainage planform stability and include other erosional processes (e.g. hillslope erosion).

An important development in this field has been the application of methodologies to efficiently search the parameter space of landscape evolution models for those that yield low residual misfit, e.g. conjugate direction, linear least squares, neighbourhood algorithm and Bayesian minimisations (e.g. Roberts & White, 2010; Fox et al., 2014; Croissant & Braun, 2014; Rudge et al., 2015; Glotzbach, 2015). Fundamental to these approaches is deciding how observed and synthetic landscapes should be compared. An obvious and commonly used approach is to measure similarities using Euclidean distances such as root mean squared misfit (e.g. Roberts & White, 2010; Croissant & Braun, 2014). In special cases the objective functions used to search for optimal models can vary smoothly and have single minima when calculated using Euclidean distances. For example, if the landscapes being compared vary smoothly (e.g. approximating Gaussian domes). However, most landscapes have complicated shapes (e.g. ridges and valleys, drainage networks), which can produce complex objective functions with many local minima. It is challenging to automate the search for optimal landscape evolution models when objective functions are complex and contain local minima.

As is well known, a related and important issue is that, the location of drainage networks in landscape evolution models are sensitive to inserted noise (see e.g. Lipp & Roberts, 2021). Noise is often inserted into the starting conditions of landscape evolution models to initiate channelisation. Models with slightly different distributions of noise can produce landscapes with ridges and valleys in different locations, even when inserted uplift histories and erosional parameters are the same (e.g. Kwang & Parker, 2019). Consequently, Euclidean measures of misfit, which typically calculate pixel-wise differences in elevation, are sensitive to noise. As such, automating the search for optimal ('true') uplift or erosional histories again becomes challenging.

Here, we introduce a methodology using the Wasserstein distance, a metric from the field of Optimal Transport, which resolves many of these issues. We first demonstrate the calculation of Wasserstein distances using one dimensional topographic transects. We then demonstrate how it can be used to identify optimal two dimensional synthetic landscapes even in the presence of added noise.

2 Methodology

2.1 Wasserstein Distance

2.1.1 Introduction

The Wasserstein distance, also known as the Earth-mover's distance, is a statistical measure of the work required to map or transform one probability distribution to another given a specified metric space (e.g. Euclidean space). Intuitively, the Wasserstein distance is considered in terms of moving one pile of material (e.g. sand) into another pile, with no loss or gain of material. The optimal transport plan is the one which transports the material from one pile to another in the least amount of distance. The Wasserstein distance is the total cost associated with this optimal transport plan. Originating from the work of Monge (1781) on Optimal Transport, calculation of Wasserstein distances have been developed by Kantorovich (1942) and Villani (2003). The Wasserstein distance has been applied to a range of problems in computing (Arjovsky et al., 2017), chemistry (Seifert et al., 2022), oceanography (Hyun et al., 2022; Nooteboom et al., 2020), climate science (Chang et al., 2015; Vissio et al., 2020), geophysics (Engquist & Froese, 2014; Métivier et al., 2016a, 2016b; Sambridge et al., 2022), sedimentology (Lipp & Vermeesch, 2023), and hydrology (Magyar & Sambridge, 2022). As far as we are aware it has not been used to invert landscapes for their properties (e.g. uplift histories, erosion rates, sedimentary fluxes).

For one dimensional distributions along the real line, the Wasserstein distance has an analytical solution. We consider two 1D distributions f and g where $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} g(x)\mathrm{d}x = 1$. The p^{th} Wasserstein distance between f and g is given by

$$W_p(f,g) = \left[\int_0^1 |F^{-1} - G^{-1}|^p \mathrm{d}t\right]^{1/p}, \tag{1}$$

where F and G represent cumulative density functions (CDFs) of $f(x)$ and $g(x)$; $F(x) = \int_{-\infty}^{x} f \ \mathrm{d}x$ and $G(x) = \int_{-\infty}^{x} g \ \mathrm{d}x$. F^{-1} and G^{-1} are therefore quantile functions. In one dimension, when $p = 1$, Equation 1 decomposes to represent the area between the two inverse CDFs.

2.1.2 Wasserstein distances applied to landscapes

The Wasserstein distance is, in general, applicable to distributions of N dimensions. As a result, the Wasserstein distance between two landscapes, for example two dimensional rectangular arrays of elevations, $z(x, y)$, could be calculated directly by comparing their two dimensional density functions (see Supporting Information). Unlike the Wasserstein distance between one dimensional distributions this problem has no known analytical solution and is solved via linear programming (e.g. Peyré & Cuturi, 2019). Whilst feasible for the problems we address in this manuscript, this approach can be computationally expensive, which is undesirable for inverse problems. Whilst we recognise that algorithms to efficiently calculate approximate multidimensional Wasserstein distances have been developed (e.g. using entropic regularisation; Cuturi, 2013) we opt instead to simply compare the one dimensional marginal sums of the two dimensional landscapes, following Sambridge et al. (2022)'s demonstration of a similar approach for inverse modeling of seismogram 'fingerprints'.

Two marginal profiles are generated for each landscape by summing elevations along the x and y axes. These are transformed into probability distributions by normalisation. If we consider a marginal elevation profile $z(x)$ evaluated on N pixels each with width Δx, we thus calculate $z^*(x)$, the normalised elevation profile through

$$z^*(x) = \frac{z_i(x)}{\Delta x \sum_i z_i(x)}, \tag{2}$$

where $i \in \{1, 2 \dots N\}$. A consequence of this normalisation is that information about absolute elevation is lost. Fortunately, additional misfit term(s), which incorporate information about absolute elevations, can be straightforwardly incorporated into the objective function. Choosing an appropriate term will likely depend on the specific problem, for example inverse modeling of two dimensional landscapes may benefit from incorporating hypsometry, which has the added benefit of also being a cumulative density function. We note that there are special cases in which identical marginal profiles can be produced from different two dimensional distributions, which could be problematical for identifying optimal landscape evolution models. For instance, a target synthetic 'landscape' in a square domain with elevation equal to zero everywhere except along a diagonal band that extends from one corner to another (e.g., along $y = x$) will have the same marginal profiles as its mirror image (diagonal non-zero band along $y = -x$). However, the presence of noise and non-trivial uplift and erosion of (observed and synthetic) landscapes indicates that encountering such marginal distributions is unlikely. Nonetheless, straightforward solutions could be implemented to overcome such scenarios if necessary (e.g. the inclusion of non-orthogonal marginal profiles in the calculation of the objective functions; additional penalty functions), with a small increase in computational expense.

In this study, we test a simple scheme in which squared and scaled differences in mean elevation are incorporated as a penalty function, $P(z_1, z_2) = \bar{z}_1 - \bar{z}_2$, where $\bar{z}_1$ and $\bar{z}_2$ are mean elevations of two landscapes (e.g. Figure 1). Combining the above steps we define a misfit function, H, utilising the Wasserstein distance, between a 'target' landscape, t, and a 'source' landscape s,

$$H(t, s) = W_2^x(t, s)^2 + W_2^y(t, s)^2 + \mu P(t, s)^2, \tag{3}$$

where $W_2^x(t, s)$ and $W_2^y(t, s)$ are the 2$^{\text{nd}}$ one dimensional Wasserstein distances ($p = 2$; Equation 1) between the normalised elevation marginals calculated by summing along the x and y axes, respectively. The scaling factor μ is a hyper-parameter which is adjusted systematically to test its impact on calculated misfit values. A `python` script which implements Equation 3 is provided at `github.com/MatthewJMorris/landscape-wasserstein`.

In this study we demonstrate the above approach using two simple examples. The first demonstrates calculation of the Wasserstein distance by comparing a target, one dimensional, topographic transect, $z(x)$, to systematically translated transects (sources). In the second example, we demonstrate a search for an optimal two dimensional, noisy, theoretical landscape, $z(x, y)$, using the misfit function defined above (Equation 3).

2.2 Euclidean distance

We compare objective functions generated using Wasserstein and more widely used Euclidean distances. These include root mean squared (rms) misfit and the L_2 norm,

$$\left[\frac{1}{N}\sum_{i=1}^{N}\left(z_i^t - z_i^s\right)^2\right]^{1/2} \quad \text{and} \quad \left[\sum_{i=1}^{N}\left(z_i^t - z_i^s\right)^2\right]^{1/2}, \tag{4}$$

respectively, where z_i^t and z_i^s are the respective elevations of the target and source landscapes. N is the number of measurements of elevation, e.g. along a transect.

2.3 Landscape evolution models

We demonstrate the use of Wasserstein distances for inverse modeling of two dimensional synthetic landscapes. The calculated landscapes are produced using the surface process computing package Landlab (Hobley et al., 2017; Barnhart, Hutton, et al., 2020). Landscape geometry is governed by the history of uplift, the erosional model, inserted additional noise and time. We assume an advective-diffusive formulation of erosion, such that,

$$\frac{\partial z}{\partial t} = -vA^m \nabla z^n + \kappa \nabla^2 z + U(x, y, t) + \eta(x, y, t), \tag{5}$$

where z is elevation, t is time, v, m, n and κ are erosional constants. A is upstream drainage area, x is distance upstream and U is uplift rate. The model is parameterized using a uniform grid with dimensions of 300 × 100 km, and cell size of 1 × 1 km. Each starting condition is generated by the following three steps. First, a central rectangular block is assigned an initial uniform elevation, u. Secondly, uniform (white) noise, η, with amplitudes 0–4% of initial elevation is added. Finally, sink-filling is performed once to permit continuous flowlines to the model boundaries (Barnes et al., 2014). We set m and $n = 0.5$ and 1, respectively, $v = 10^{-3}$ kyr^{-1}, erosional 'diffusivity' $\kappa = 100$ m^2/kyr. The FastScape erosion scheme is used to solve Equation 5 (Braun & Willett, 2013). Flow-routing is performed with the 'D8' algorithm (O'Callaghan & Mark, 1984). The landscape model is run forwards in time for 10 Myr, with flow-routing, and advective and diffusive erosion calculated at each timestep. In all models $U = 0$, i.e., no uplift is added beyond the initial elevation. The timestep is set to 8 kyr, such that the Courant-Friedrichs-Lewy condition for numerical stability is satisfied.

3 Results and Discussion

3.1 One dimensional topographic transect

A simple demonstration of how Wasserstein distances can be calculated and used to identify an optimal topographic transect is shown in Figure 1. This transect was chosen to demonstrate how an optimal source (e.g. theoretical landscape) can be identified even for quasi-periodic topography, analogous to 'cycle-skipping' problems in seismology. The transect was extracted from the SRTM 1 arc second dataset across the Appalachian mountains, USA (Figure 1a: A—A′; horizontal resolution ∼ 30 m). The resultant target profile is shown in Figure 1b (black polygon). Source transects are generated by translating the target transect (e.g. red polygon in Figure 1b). Initially the source transect does not overlap with the target, it is progressively shifted until it fully overlaps, before

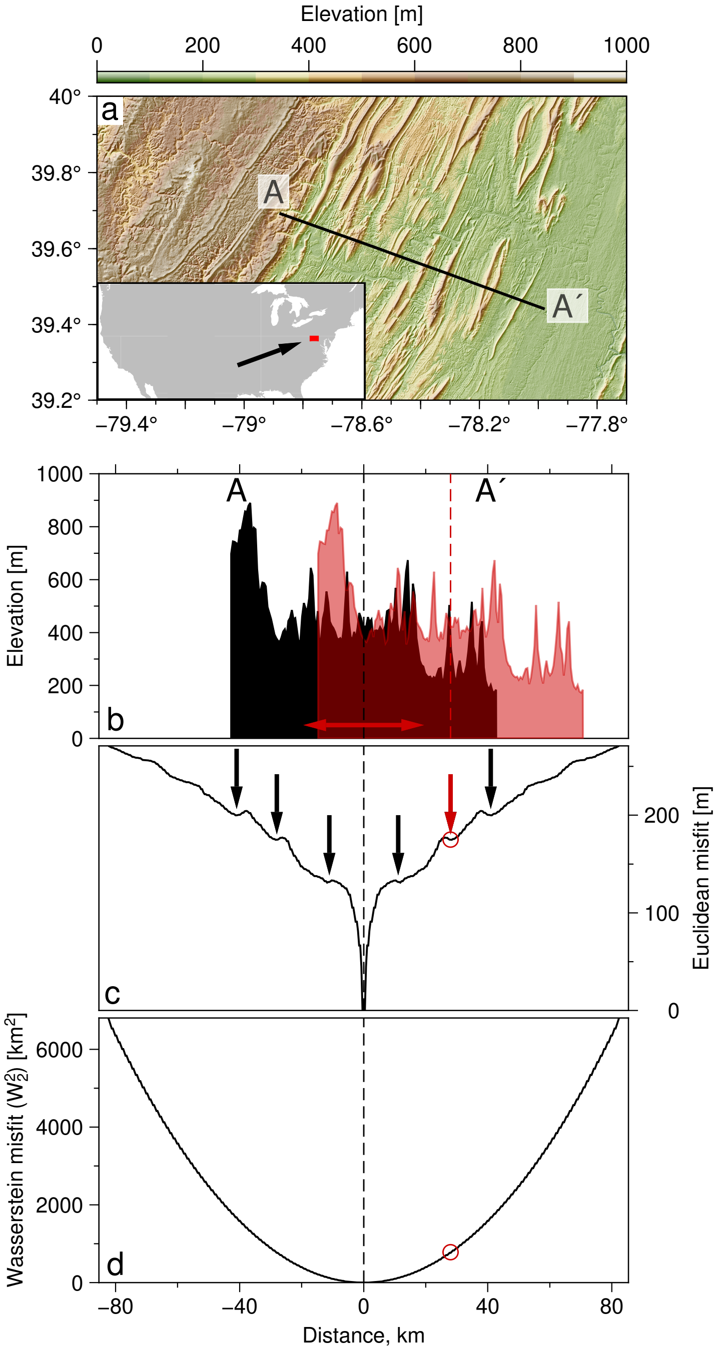

Figure 1. Wasserstein and Euclidean (*rms*) misfit between observed (target) and theoretical (source) topographic transects. (a) Transect A—A′ across the Appalachian mountains, USA. Red box on inset map indicates region shown in panel (a). Topography extracted from SRTM 1 arc second dataset. (b) Black polygon shows observed elevation along transect A—A′; dashed black line = profile mid-point. Theoretical transects were generated by translating the observed transect left and right (red arrows); red dashed line = mid-point of example theoretical transect (red translucent polygon). (c) Euclidean (*rms*) misfit between observed and theoretical transects. Distance = distance between mid-points of observed and theoretical transects. Arrows = local minima. Black dashed line = global minima; centre of observed transect. Red arrow and circle = misfit of theoretical transect shown in panel (a). (d) Wasserstein misfit (Equation 1) between observed and translated transects. Red circle = mid-point of theoretical transect shown in panel (b). Black dashed line = global minima at mid-point of observed transect shown in panel (a).

continuing translating until, again, there is no overlap. Euclidean misfit and Wasserstein distance are calculated for each translated (source) transect. These transects have identical absolute elevations and therefore there is no need to include an additional penalty term. Thus, in this scenario we define the Wasserstein misfit simply as the squared one dimensional Wasserstein distance between the two normalised elevation profiles (Equation 1, $p = 2$).

The Euclidean misfit function is complicated and contains several local minima (Figure 1c: arrows). However, the Wasserstein misfit function is smooth and quadratic, with a single minimum located where the two transects are aligned (Figure 1b-d: dashed lines). This simple one dimensional example suggests that the Wasserstein distance holds promise for comparing real topographic data. In particular, that automated techniques for efficient locating of global minima (e.g. Brent's method) are more likely to be successful if similarity between topography is measured using Wasserstein, rather than Euclidean, statistics.

3.2 Two dimensional landscapes

We now extend the problem to two dimensions. In this simple example we first seek the location of uplift that was used to generate a synthetic (target) landscape. We wish to identify optimal theoretical landscapes even when additional noise is inserted into the model, as is common in landscape evolution models when channelisation is required. Figure 2b shows the target landscape generated following the procedure described in Section 2.3 (Equation 5). The marginals are shown in Figure 2a and 2c. A separate source landscape is generated by changing the distribution of added noise, η, at the initial condition, all other parameters are held constant. This source landscape is translated along the x axis of the model domain (midpoints from 50 to 250 km; 200 translations), illustrated by the red arrows in Figure 2d-e. An example of a translated landscape and its marginals are shown in Figure 2d-f. Figure 2g shows the difference in cell elevations between the target and source landscapes when fully aligned. This map shows the changes in landscape geometry (including elevations and fluvial planform) that are governed by changing the initial noise inserted into the landscapes at their inception. Figure 2h illustrates that despite different initial noise conditions generating different drainage networks (Figure 2g), the Wasserstein distance successfully identifies a global minimum misfit where the landscapes are fully aligned. In comparison, the Euclidean misfit contains local minima and the position of the global minimum is offset from the target. This marginal experiment took 0.1 seconds on an Intel i7-6700 desktop computer with 64 gigabytes of RAM. Calculating Euclidean distances took 0.03 seconds. In comparison, calculating Wasserstein distances using two dimensional density functions for 7 translations of the source landscape took approximately 45 minutes (see Supporting Information for the results of this experiment).

We now extend this test by changing the amplitude of initial uplift, u, used to generate the target landscape. Wasserstein and Euclidean statistics are calculated for each source landscape. We systematically varied amplitude of uplift by $\Delta u = 25$ m between 25 to 250 m. For each of these ten landscapes, 100 random distributions of white noise with amplitudes 0–4% of uplift were added prior to the initial sink-filling and flow routing steps, generating the starting condition for each model. As in the previous example, these source landscapes are translated along the x axis. This Monte Carlo style experiment is designed to test the impact of noisy initial conditions on identifying optimal theoretical landscape evolution models. In short, we perform 100 brute force inversions to identify two parameters: the location and magnitude of initial uplift, each with a different initial distribution of noise in the starting condition.

Figure 3a shows Euclidean (rms) misfit as a function of initial elevation (uplift) and position for a more general version of the test shown in Figure 2. In Figures 3a and

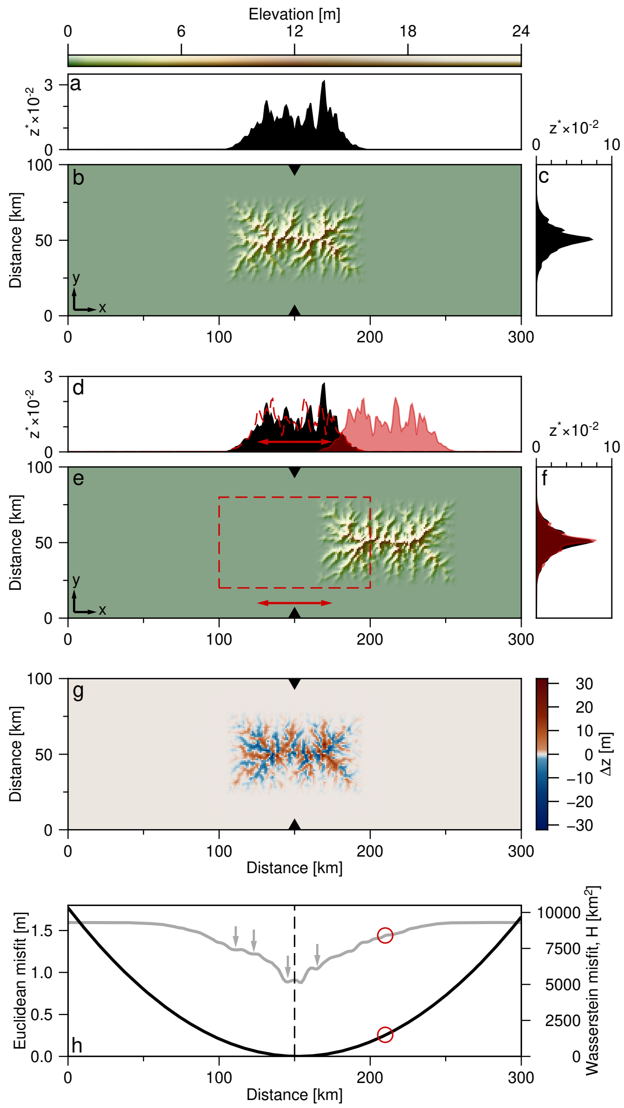

Figure 2. Misfit between two dimensional, $z(x,y)$, theoretical landscapes. (a) Marginal elevations (z_x^*) of the landscape shown in panel (b). (b) Synthetic, 'target', landscape (see body text for construction). Filled black triangles point to centre of the domain. (c) Marginal elevations (z_y^*) of the landscape shown in panel (b). (d) Black profile is as per panel (a). Red dashed profile = marginal generated from synthetic landscape that has different initial noise but same distribution of uplift used to generate landscape shown in panel (b): red-dashed rectangle shown in panel (e). Red translucent marginal was generated from the translated, 'source', landscape shown in panel (e). (e) Red arrows indicate translation directions. Topography = example translated landscape. (f) As per panel (c); red translucent profile = marginal from landscape shown in panel (e). (g) Elevation differences, Δz, between the target and source landscapes when all conditions bar initial noise are constant. (h) Grey curve = *rms* misfit (Equation 4) between synthetic landscape shown in panel (b) and translated landscapes (e.g. panel e). Grey arrows = local minima. Red circle = misfit between target and source landscapes shown in panels (b) and (e), respectively. Black solid curve = Wasserstein misfit (Equation 3, $p = 2$). Black dashed line = target and source landscapes with common mid-points indicated by black triangles on panels (b), (e) and (g).

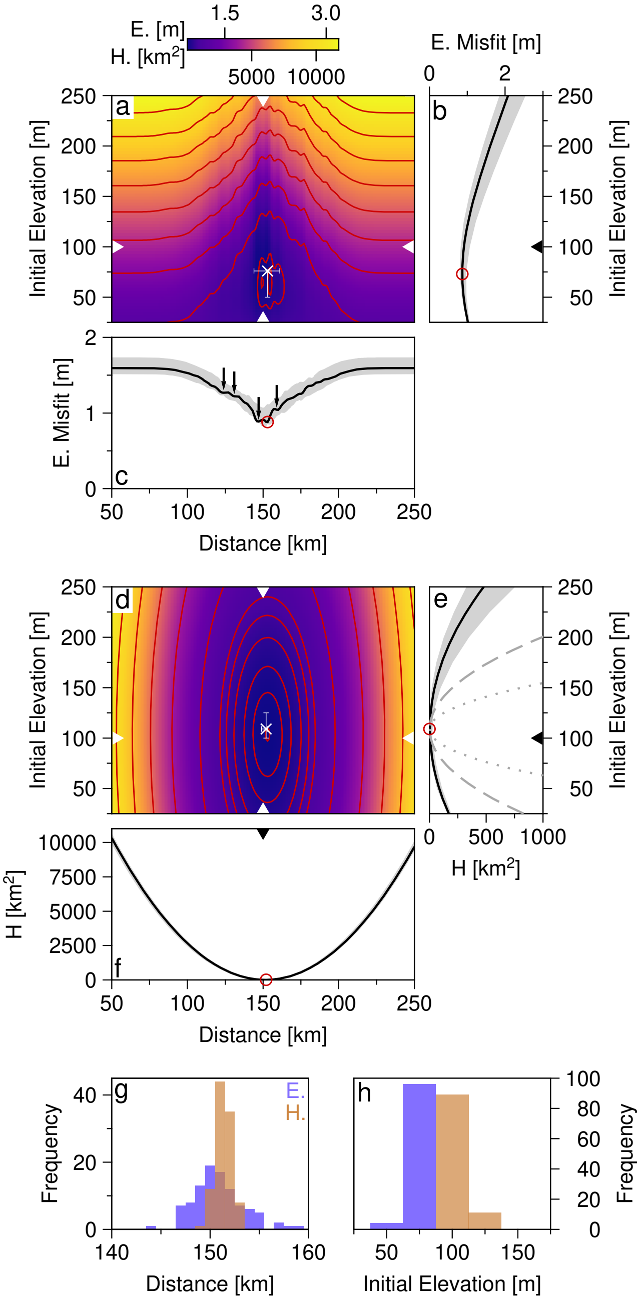

Figure 3. Inverse modeling of noisy synthetic landscapes: Euclidean and Wasserstein statistics. (a) Euclidean (*rms*) misfit (Equation 4) between target landscape (Figure 2b) and source landscapes as a function of noise and uplift. Uplift was translated as demonstrated in Figure 2; initial elevations (amplitude of uplift) were systematically varied between 25–250 m. Amplitudes of initial added white noise were scaled to be 0–4% of initial elevation. White cross = minimum misfit; error bars = range of misfit minima for 100 different parameterizations of noise. Outer white triangles point to initial elevation and mid-point of target landscape (100 m, 150 km). (b-c) Black curves = transects through misfit function between white triangles in panel (a). Shaded grey regions = minimum and maximum misfit for the 100 different distributions of noise. Black arrows = local minima of the 1D transect; circles = global minimum. (d-f) Wasserstein (H) misfit (Equation 3). (e) Black curve and grey envelope: $\mu = 10^4$; dashed and dotted grey curves: $\mu = 2.5 \times 10^5$ and 10^6, respectively. (g-h) Loci of misfit minima for the 100 different distributions of initial noise; E/H = Euclidean/Wasserstein misfit.

b, $\Delta u = 3$ m, instead of 25 m, in order to produce a more granular misfit space. The black curves in Figure 3b-c show slices through the Euclidean misfit function centred on the target landscape parameterization (indicated by the white arrowheads in panel a). Local minima are indicated by the black arrows in Figure 3c. The error bars in Figure 3a show the range of global minima for the models examined. These results are summarised in the histograms shown in Figure 3g-h. The grey envelope in Figure 3b-c indicates the range of misfit values for all initial noise conditions. The global minima from all models is shown by the red circles. The Wasserstein-based misfit functions for the same models are shown in Figure 3d-h (again Δu =3 m for d and e, instead of 25 m). The shape of the objective function with respect to initial elevation is governed by the value of the scaling parameter, μ. Increasing μ increases the contribution of the penalty term, P. A value of $\mu = 10^4$ was used to generate the results shown in Figure 3d-h (e.g. thin black curve and grey envelope in panel e). For comparison, the dashed and dotted grey curves in panel e show the shape of the objective function as a function of initial elevation when $\mu = 2.5 \times 10^5$ and 10^6, respectively.

Euclidean minima are broadly distributed around a point offset from the expected parameter values (Figure 3a). In contrast, minima calculated using Wasserstein distances have a smaller spread around the expected value (Figure 3d). These results indicate that Wasserstein distances provide means to identify optimal landscape evolution models even in the presence of noisy initial conditions that hinder Euclidean-based approaches. We note that if source landscapes are, simply, translated versions of the target (i.e. they have exactly the same elevations; same inserted noise), the Wasserstein misfit function, H, is symmetric about the mid-point of the target landscape.

4 Conclusions

Landscape geometries are determined by histories of geologic and geomorphologic processes. A corollary is that landscape form contains information about driving processes (e.g. uplift, erosion, climate). A general goal is to identity theoretical landscape evolution models that accurately predict landscape form and its history. Here we demonstrate the use of the Wasserstein distance for identifying such models. These distances are used to quantify similarity between observed and theoretical landscapes. They are shown to generate simple, quadratic, objective functions with single (global) minima. Even in the presence of noise (which could be computational or real) Wasserstein statistics can identify optimal landscape evolution models when equivalent Euclidean statistics (rms, L_2 norm) are complex, with global minima offset from the expected value. We suggest that Wasserstein statistics show promise for inverse modeling of landscapes to identify the processes that drive their evolution, for example, uplift histories.

Acknowledgments

We thank C. O'Malley and F. Richards for their help and two anonymous reviewers for helping us to clarify our thesis.

5 Open Research

We use the Python library `POT` to calculate Wasserstein distances (Flamary et al., 2021). `SRTM` data can be downloaded from `https://earthexplorer.usgs.gov/`. Figures were generated using `GMT 6.3.0` (Wessel et al., 2019). Accompanying code is archived on Zenodo at `https://doi.org/10.5281/zenodo.7602208`.

References

Anderson, R. S., & Anderson, S. P. (2010). *Geomorphology: The Mechan-*

ics and Chemistry of Landscapes. Cambridge University Press. doi: 10.1017/CBO9780511794827

Arjovsky, M., Chintala, S., & Bottou, L. (2017). *Wasserstein GAN.* arXiv. doi: 10.48550/ARXIV.1701.07875

Barnes, R., Lehman, C., & Mulla, D. (2014). Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, *62*, 117-127.

Barnhart, K. R., Hutton, E. W., Tucker, G. E., Gasparini, N. M., Istanbulluoglu, E., Hobley, D. E., … others (2020). Landlab v2. 0: A software package for Earth surface dynamics. *Earth Surface Dynamics*, *8*(2), 379–397. doi: 10.5194/esurf-8-379-2020

Barnhart, K. R., Tucker, G. E., Doty, S. G., Shobe, C. M., Glade, R. C., Rossi, M. W., & Hill, M. C. (2020). Inverting Topography for Landscape Evolution Model Process Representation: 1. Conceptualization and Sensitivity Analysis. *Journal of Geophysical Research: Earth Surface*, *125*(7). doi: 10.1029/2018JF004961

Bishop, P. (2007). Long-term landscape evolution: Linking tectonics and surface processes. *Earth Surface Processes and Landforms*, *32*(3), 329–365. doi: 10.1002/esp.1493

Braun, J., & Willett, S. D. (2013). A very efficient O(n), implicit and parallel method to solve the stream power equation governing fluvial incision and landscape evolution. *Geomorphology*, *180–181*, 170–179. doi: 10.1016/j.geomorph.2012.10.008

Chang, C.-P., Ghil, M., Latif, M., & Wallace, J. M. (2015). *A Mathematical Theory of Climate Sensitivity or, How to Deal With Both Anthropogenic Forcing and Natural Variability?* (Vol. 6). WORLD SCIENTIFIC.

Croissant, T., & Braun, J. (2014). Constraining the stream power law: A novel approach combining a landscape evolution model and an inversion method. *Earth Surface Dynamics*, *2*(1), 155–166. doi: 10.5194/esurf-2-155-2014

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26). Curran Associates, Inc.

Davis, W. M. (1899). The geographical cycle. *Geographical Journal*, *14*, 481-504.

Engquist, B., & Froese, B. (2014). Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Sciences*, *12*(5), 979–988. doi: 10.4310/CMS.2014.v12.n5.a7

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., … Vayer, T. (2021). POT: Python Optimal Transport. *Journal of Machine Learning Research*, *22*(78), 1–8.

Fox, M., Goren, L., May, D. A., & Willett, S. D. (2014). Inversion of fluvial channels for paleorock uplift rates in Taiwan. *Journal of Geophysical Research: Earth Surface*, *119*(9), 1853–1875. doi: 10.1002/2014JF003196

Gallen, S. F., & Fernández-Blanco, D. (2021). A New Data-Driven Bayesian Inversion of Fluvial Topography Clarifies the Tectonic History of the Corinth Rift and Reveals a Channel Steepness Threshold. *Journal of Geophysical Research: Earth Surface*, *126*(3), e2020JF005651. doi: 10.1029/2020JF005651

Glotzbach, C. (2015, 6). Deriving rock uplift histories from data-driven inversion of river profiles. *Geology*, *43*(6), 467–470. doi: 10.1130/G36702.1

Goren, L., Fox, M., & Willett, S. D. (2014). Tectonics from fluvial topography using formal linear inversion: Theory and applications to the Inyo Mountains, California. *Journal of Geophysical Research: Earth Surface*, *119*(8), 1651-1681. doi: 10.1002/2014JF003079

Hobley, D. E., Adams, J. M., Nudurupati, S. S., Hutton, E. W., Gasparini, N. M., Istanbulluoglu, E., & Tucker, G. E. (2017). Creative computing with Landlab:

an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics . *Earth Surface Dynamics*, *5*(1), 21–46. doi: 10.5194/esurf-5-21-2017

Hyun, S., Mishra, A., Follett, C. L., Jonsson, B., Kulk, G., Forget, G., ... Bien, J. (2022). Ocean mover's distance: Using optimal transport for analysing oceanographic data. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *478*(2262). doi: 10.1098/rspa.2021.0875

Kantorovich, L. V. (1942). On the Translocation of Masses. *Dokl. Acad. Nauk. USSR*, *37*, 7–8, 227–229.

Kwang, J. S., & Parker, G. (2019). Extreme Memory of Initial Conditions in Numerical Landscape Evolution Models. *Geophysical Research Letters*, *46*(12), 6563-6573. doi: 10.1029/2019GL083305

Lipp, A. G., & Roberts, G. G. (2021). Scale-Dependent Flow Directions of Rivers and the Importance of Subplate Support. *Geophysical Research Letters*, *48*(1), e2020GL091107. doi: 10.1029/2020GL091107

Lipp, A. G., & Vermeesch, P. (2023). Comparing detrital age spectra, and other geological distributions, using the wasserstein distance. *Geochronology, in review*. doi: 10.31223/X5TM02

Magyar, J. C., & Sambridge, M. S. (2022). The Wasserstein distance as a hydrological objective function. *EGUsphere*, *2022*, 1–32. doi: 10.5194/egusphere-2022-1117

Monge, G. (1781). Mémoire sur la théorie des déblais et des remplais. *De l'Imprimerie Royale*.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J. (2016a, 02). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, *205*(1), 345–377. doi: 10.1093/gji/ggw014

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J. (2016b, sep). An optimal transport approach for seismic tomography: Application to 3D full waveform inversion. *Inverse Problems*, *32*(11), 115008. doi: 10.1088/0266-5611/32/11/115008

Nooteboom, P. D., Delandmeter, P., van Sebille, E., Bijl, P. K., Dijkstra, H. A., & von der Heydt, A. S. (2020). Resolution dependency of sinking Lagrangian particles in ocean general circulation models. *PLOS ONE*, *15*(9), 1–16. doi: 10.1371/journal.pone.0238650

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, *28*(3), 323-344. doi: 10.1016/S0734-189X(84)80011-0

O'Malley, C. P. B., White, N. J., Stephenson, S. N., & Roberts, G. G. (2021). Large-Scale Tectonic Forcing of the African Landscape. *Journal of Geophysical Research: Earth Surface*, *126*(12), e2021JF006345. doi: 10.1029/2021JF006345

Peyré, G., & Cuturi, M. (2019). Computational Optimal Transport. *Foundations and Trends in Machine Learning*, *11*, 355–607. doi: 10.48550/ARXIV.1803.00567

Pritchard, D., Roberts, G. G., White, N. J., & Richardson, C. N. (2009). Uplift histories from river profiles. *Geophysical Research Letters*, *36*(24). doi: 10.1029/2009GL040928

Roberts, G. G., & White, N. (2010). Estimating uplift rate histories from river profiles using African examples. *Journal of Geophysical Research: Solid Earth*, *115*(B2). doi: 10.1029/2009JB006692

Rudge, J. F., Roberts, G. G., White, N. J., & Richardson, C. N. (2015). Uplift histories of Africa and Australia from linear inverse modeling of drainage inventories. *Journal of Geophysical Research: Earth Surface*, *120*(5), 894–914. doi: 10.1002/2014JF003297

Sambridge, M., Jackson, A., & Valentine, A. P. (2022). Geophysical Inversion and Optimal Transport. *Geophysical Journal International*, *231*(1), 172–198. doi: 10.1093/gji/ggac151

Seifert, N. A., Prozument, K., & Davis, M. J. (2022). Computational optimal transport for molecular spectra: {The} semi-discrete case. *The Journal of Chemical Physics*, *156*(13), 134117. doi: 10.1063/5.0087385

Villani, C. (2003). *Topics in Optimal Transportation.* American Mathematical Society.

Vissio, G., Lembo, V., Lucarini, V., & Ghil, M. (2020). Evaluating the Performance of Climate Models Based on Wasserstein Distance. *Geophysical Research Letters*, *47*(21). doi: 10.1029/2020GL089385

Wapenhans, I., Fernandes, V. M., O'Malley, C., White, N., & Roberts, G. G. (2021). Scale-Dependent Contributors to River Profile Geometry. *Journal of Geophysical Research: Earth Surface*, *126*(6). doi: 10.1029/2020JF005879

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. *Geochemistry, Geophysics, Geosystems*, *20*(11), 5556-5564. doi: 10.1029/2019GC008515

Supporting Information for "Towards inverse modelling of landscapes using the Wasserstein distance"

M. J. Morris[1], A. G. Lipp[2], G. G. Roberts[1]

[1]Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK

[2]Merton College, University of Oxford, Oxford, OX1 4JD, UK

Contents of this file

1. Calculation of two dimensional Wasserstein distances for synthetic landscapes
2. Figure S1: Two dimensional Wasserstein distance for a synthetic landscape

Two Dimensional Wasserstein

As stated in the main text, the Wasserstein distance can be applied to distributions of N dimensions. We explain here the approach to calculate a Wasserstein distance for $N = 2$, using a pair of two dimensional density functions, $f(x, y)$ and $g(x, y)$. First, a cost matrix is calculated, representing the distance from each coordinate point in the source function (f) to every coordinate point in the target function (g). A choice must be made about the method used to define the distance between these points, known as the cost function. Several approaches exist including the squared Euclidean distance and Manhattan distance. The cost c_{ij} to transport point i to point j using a squared Euclidean distance is given by

$$c_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2, \tag{1}$$

where x_i and y_j are the x and y coordinates of the ith and jth discrete points in the distributions respectively. A map of transport cost, known as the cost matrix may be constructed by computing the cost function for all coordinates.

There are a number of ways to transport f to g, however some are more efficient than others. Any chosen method of transport is given by a transport plan, π_{ij}, defined as the mapping of the source distribution onto the target distribution. i.e., entry i, j corresponds to how much of the source distribution f is translated from point i onto point j. Thus the total cost, C, of any given transport plan is

$$C = \sum_{i=1,j=1}^{n_f,n_g} \pi_{ij} c_{ij}. \tag{2}$$

We seek the most efficient, or optimal, transport plan, which results in the least cost to transport f to g (i.e. minimises C). This transport plan can be given generally, as per Kantorovich (1942), by the Linear Programming problem

$$W_p^p = \min_{\pi_{ij}} \sum_{i,j} \pi_{ij} c_{ij}, \tag{3}$$

where p is dependent on the exponent of the cost function. In our case for a squared euclidean distance, $p = 2$. Equation 3 is subject to three constraints. First, and secondly, the row and column sums of the transport plan are equal to the number of elements in f and

g respectively, and thirdly, that the optimal transport plan is greater than or equal to zero.

Examples of Wasserstein distances calculated using two dimensional density functions (Equation 3) are shown in Figure S1. This figure shows the results of an experiment similar to that shown in Figure 2 of the main manuscript. Due to the computational expense involved with this approach, only 7 translations of the source landscape are presented, rather than the 200 translations used for the marginals approach. Nonetheless, Figure S1c shows that the calculated misfit function correctly identifies the optimal location of the source landscape. We provide a `python` script which calculates the Wasserstein distance between two synthetic landscapes using the methodology described above at `github.com/MatthewJMorris/landscape-wasserstein`.

References

Kantorovich, L. V. (1942). On the Translocation of Masses. *Dokl. Acad. Nauk. USSR*, *37*, 7–8, 227–229.

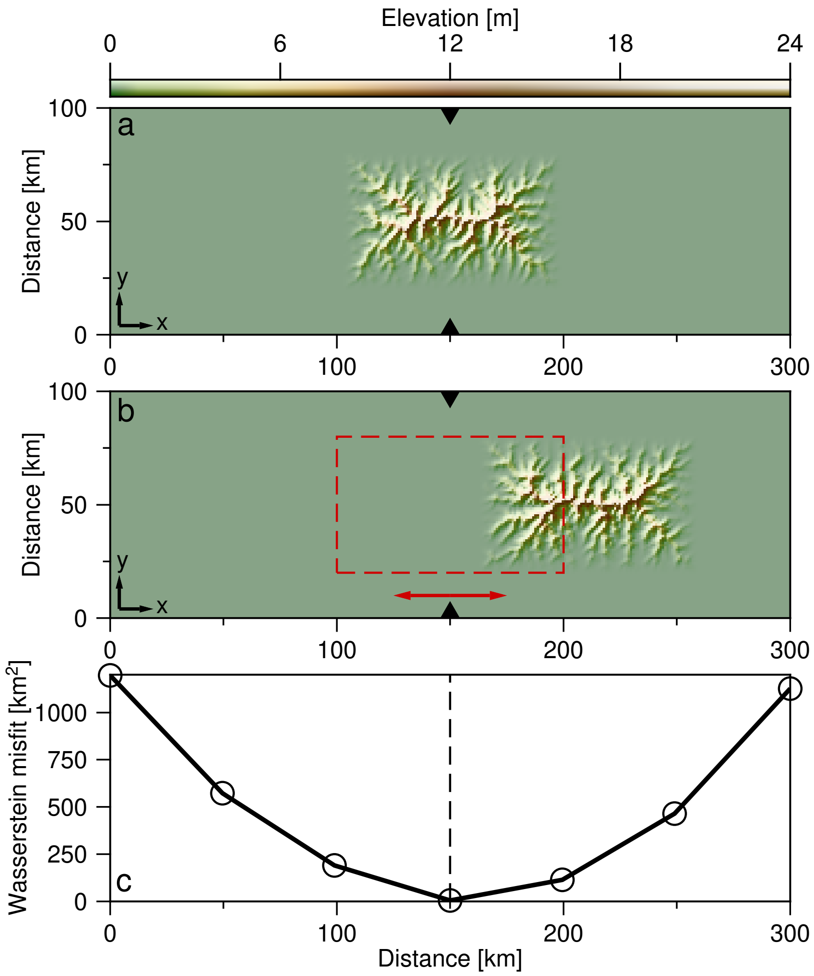

Figure S1. Two-dimensional Wasserstein distances for synthetic landscapes. (a) Target landscape, as per Figure 2b of the main manuscript. (b) Source landscape, as per Figure 2e of the main manuscript. (c) Two dimensional Wasserstein distance calculated between the target and the source landscapes for 7 translations (circles) of the source landscape across the domain using the two dimensional approach (Equation 3).