Hierarchies and Ranks for Persistence Pairs

Bastian Rieck[1] Heike Leitte[1] Filip Sadlo[2]

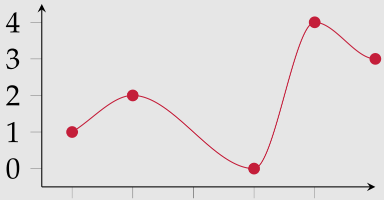
[1]TU Kaiserslautern, Germany

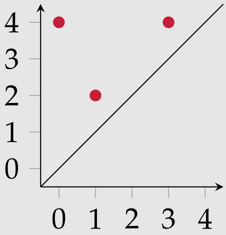
[2]Heidelberg University, Germany

28 February 2017

Motivation

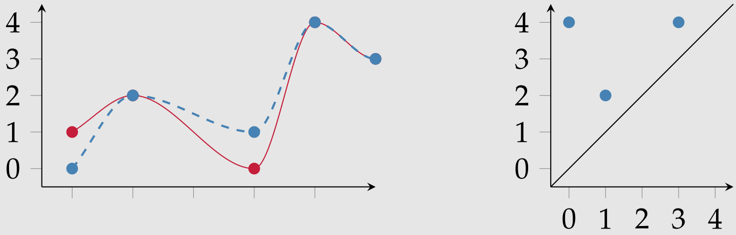
Different functions may have identical persistence diagrams





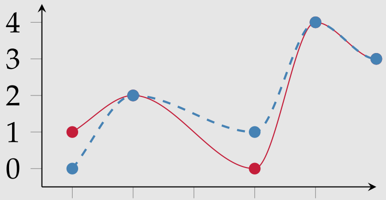
Motivation

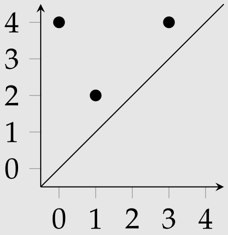
Different functions may have identical persistence diagrams



Motivation

Different functions may have identical persistence diagrams





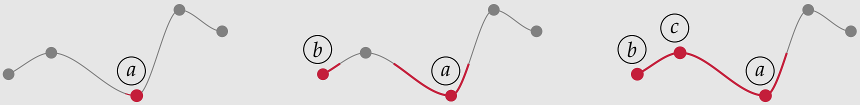
Motivation, continued

Identical persistence diagrams

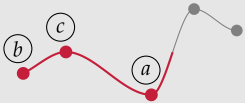
- Generic issue: occurs both in sublevel set and superlevel set calculations
- Solution: add additional (geometrical) information, e.g. merge trees

Assumptions

- Pairing of connected components (zero-dimensional persistent homology)
- Pairing uses “elder rule”: The “older” connected component persists, i.e. the one with the smaller index with respect to the filtration
- In the example below, component (a) persists, but component (b) is destroyed by the merge at (c)



Regular persistence hierarchy



Add $b \to a$ to the hierarchy. Notice that the hierarchy uses *directed* edges.

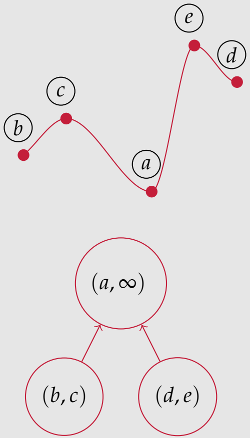
Require: A domain $\mathbb{D}$
Require: A function $f\colon \mathbb{D} \to \mathbb{R}$
  $\mathrm{U} \leftarrow \emptyset$
  Sort the function values of f in ascending order
  for function value y of f **do**
    if y is a local minimum **then**
      Create a new connected component in U
    else if y is a local maximum or a saddle **then**
      Use U to merge the two connected components
      Let y' refer to the creator of the older component
      Create the edge (y', y) in the hierarchy
    else
      Use U to add y to the current connected component
    end if
  end for

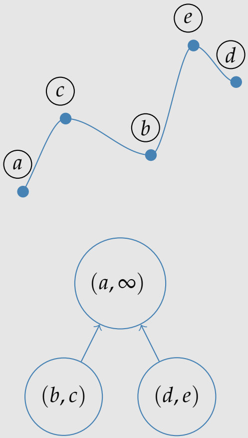
Regular persistence hierarchy, continued

- Introduced by Bauer, 2011, “Persistence in discrete Morse theory”
- By definition, the hierarchy forms a directed acyclic graph
- Original motivation: determining cancellation sequences of Morse functions

Problem

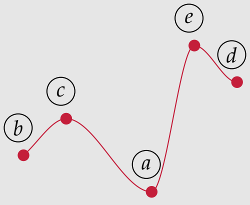
Lack of expressiveness

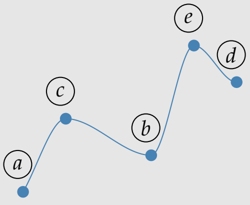




Problem

Lack of expressiveness



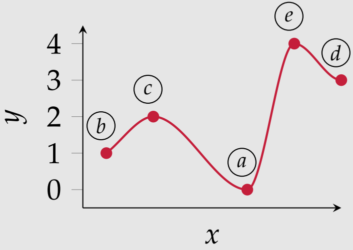


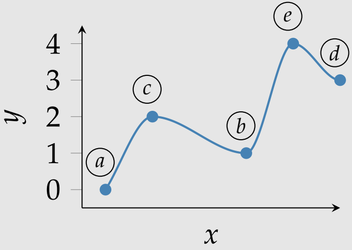
Key observation

- Not all merges in the sublevel sets are equal!
- Take connectivity with respect to other critical points into account.

Sublevel set connectivity

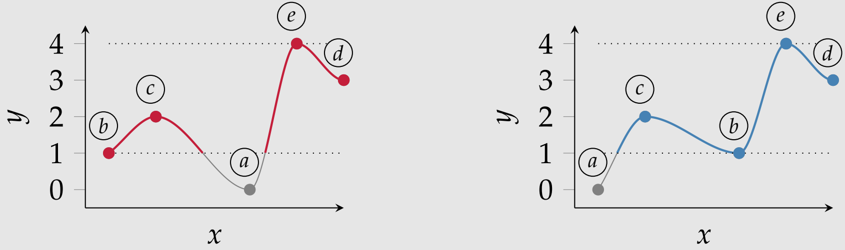
Use interlevel set





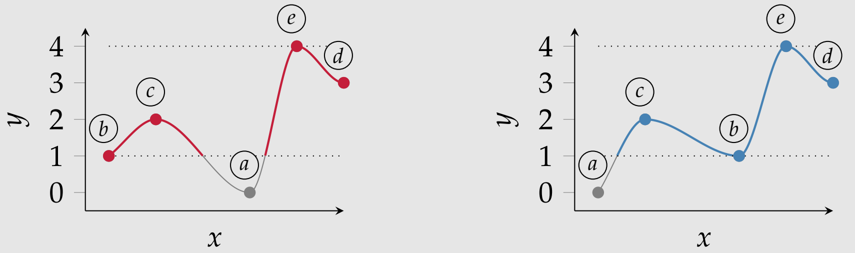
Sublevel set connectivity

Use interlevel set



Sublevel set connectivity

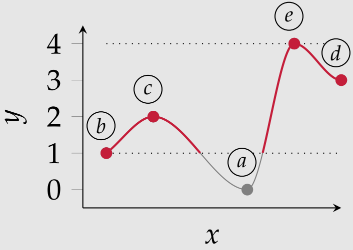
Use interlevel set

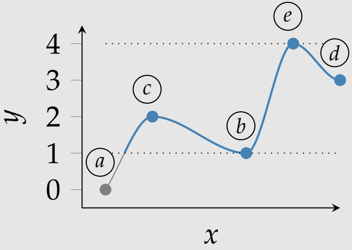


$$\mathcal{L}_{l,u}(f) := \mathcal{L}_u^-(f) \setminus \mathcal{L}_l^-(f) = \{x \in \mathbb{D} \mid l \leq f(x) \leq u\}$$

Sublevel set connectivity

Use interlevel set



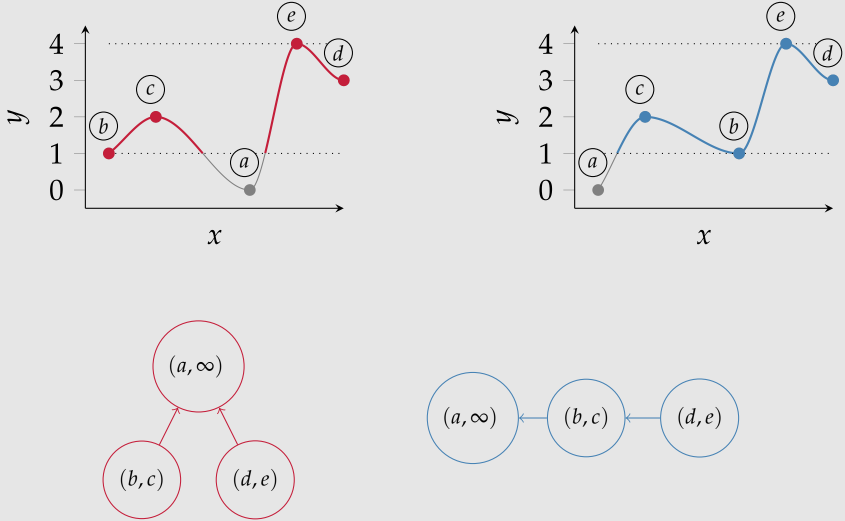


$$\mathcal{L}_{l,u}(f) := \mathcal{L}_u^-(f) \setminus \mathcal{L}_l^-(f) = \{x \in \mathbb{D} \mid l \leq f(x) \leq u\}$$

- $\mathcal{L}_{b,e}(f)$ has two connected components for the ● function, but only one for the ● function
- Hence: use the same level for the ● function, but insert pair on lower level for the ● function

Sublevel set connectivity

Use interlevel set

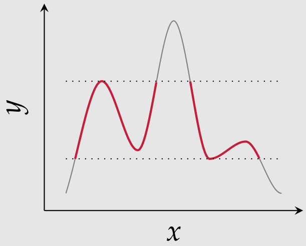


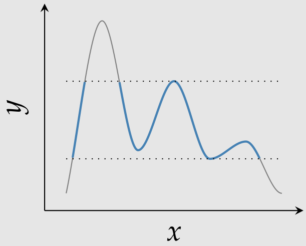
Algorithm

Excerpt; shortened notation

for function value y of f **do**
  if y is a local maximum **then**
    Use U to merge the two connected components
    Let C_1 and C_2 be the two components at y (w.l.o.g. let C_1 be the older one)
    if both components have a trivial critical value **then**
      Create the edge (C_1, C_2) in the hierarchy
    else
      Let c_1, c_2 be the critical values of C_1, C_2
      Create the interlevel set $L := \mathcal{L}_{c_2,y}(f)$
      if shortest path between c_1, c_2 in L contains no other critical points **then**
        Create edge (c_1, y) in the hierarchy
      end if
    end if
  end if
end for

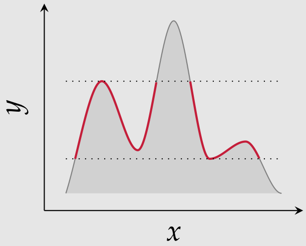
Necessity of the connectivity check

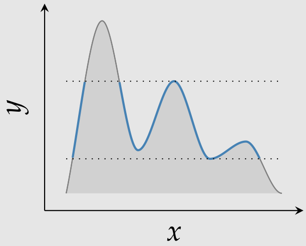




In one dimension (segments), a simple connectivity check is sufficient. In two dimensions (isolines), *both* interlevel sets are connected, though!

Necessity of the connectivity check





In one dimension (segments), a simple connectivity check is sufficient. In two dimensions (isolines), *both* interlevel sets are connected, though!

Implications

- Extended persistence hierarchy usually has more levels than the regular one
- The calculation incorporates a modicum of geometrical information

Open questions

- Is this connectivity check sufficiently distinctive?
- What is the relation to “basins of attraction” in discrete Morse theory?

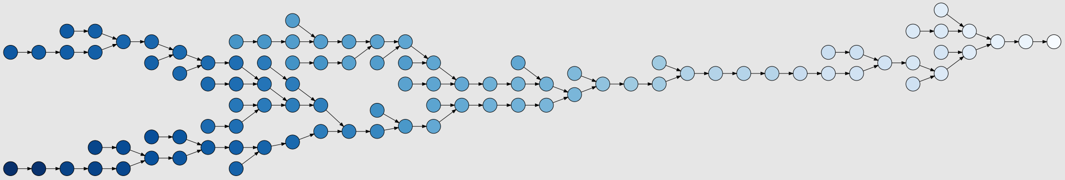
Comparison with other tree-based concepts

In the paper

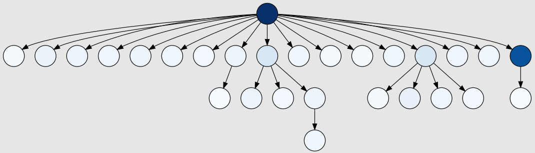
- Regular persistence hierarchy can be obtained via branch decomposition
- Merge trees are discriminative, but their branch decomposition may still coincide for different functions
- Hence, extended persistence hierarchy cannot be derived that way

Robustness

Merge tree vs. extended persistence hierarchy, colored by persistence



Merge tree



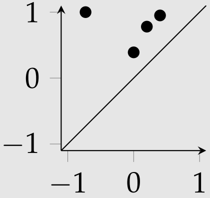
Extended persistence hierarchy

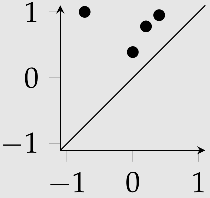
Application

Ranks

How many nodes can be reached from a given node u in the (extended) persistence hierarchy $\mathcal{H}$?

$$\operatorname{rank}(u) := \operatorname{card}\{v \in \mathcal{H} \mid u \sim v\}$$



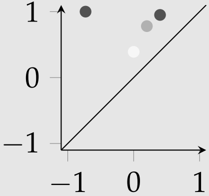


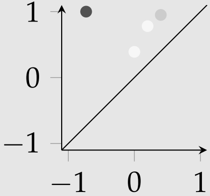
Application

Ranks

How many nodes can be reached from a given node u in the (extended) persistence hierarchy $\mathcal{H}$?

$$\mathrm{rank}(u) := \mathrm{card}\,\{v \in \mathcal{H} \mid u \sim v\}$$





Application

Stability measure

Overarching question

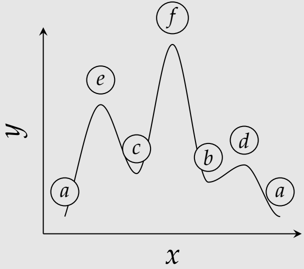
How stable is the *location* of a critical point? Persistence pairs are a continuous function of the input data, but their location is not.

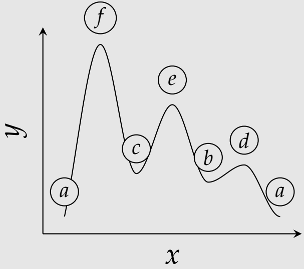
Previous work

Bendich & Bubenik, 2015, "Stabilizing the output of persistent homology computations".

Stability measure

Example (superlevel sets)





Critical points:

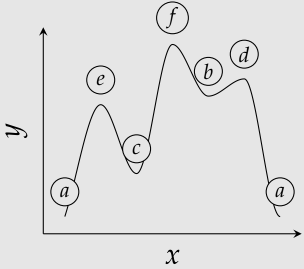
- $(f, -\infty)$
- (e, c)
- (d, b)

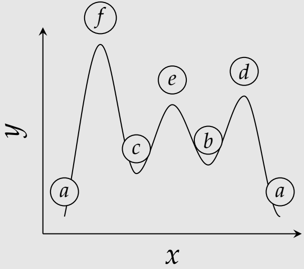
Critical points:

- $(f, -\infty)$
- (e, c)
- (d, b)

Stability measure

Example, perturbed





Critical points:

- $(f, -\infty)$
- (e, c)
- (d, b)

Critical points:

- $(f, -\infty)$
- (d, c)
- (e, b)

Stability measure
Formal definition

For an edge $e := \{(\sigma, \tau), (\sigma', \tau')\}$ in the hierarchy $\mathcal{H}$:

$$\mathrm{stab}(e) := \max\left\{|f(\sigma) - f(\sigma')|, |f(\tau) - f(\tau')|\right\} \tag{1}$$

For a vertex v:

$$\mathrm{stab}(v) := \min\left\{\min_{e=(v,w)\in\mathcal{H}} \mathrm{stab}(e), \mathrm{pers}(v)\right\} \tag{2}$$

Here: $\mathrm{stab}\left(\textcircled{e}\right) \ll \mathrm{pers}\left(\textcircled{e}\right)$ for the second hierarchy.

Using the *minimum* of all stability values is an extremely conservative worst-case assumption!

Implications

Another criterion for distinguishing between functions with equal persistence diagrams, based on worst-case location stability of creators of critical pairs.

Open questions

- How useful is this assumption?
- Does it characterize *all* perturbations of critical points?

Application

Dissimilarity measure

Use existing *tree edit distance* algorithms. Cost function for relabeling a node:

$$\text{cost}_1 = \max\left(|c_1 - c_2|, |d_1 - d_2|\right) \tag{3}$$

Cost function for deleting or inserting a node:

$$\text{cost}_2 = \text{pers}(c, d) = |d - c|, \tag{4}$$

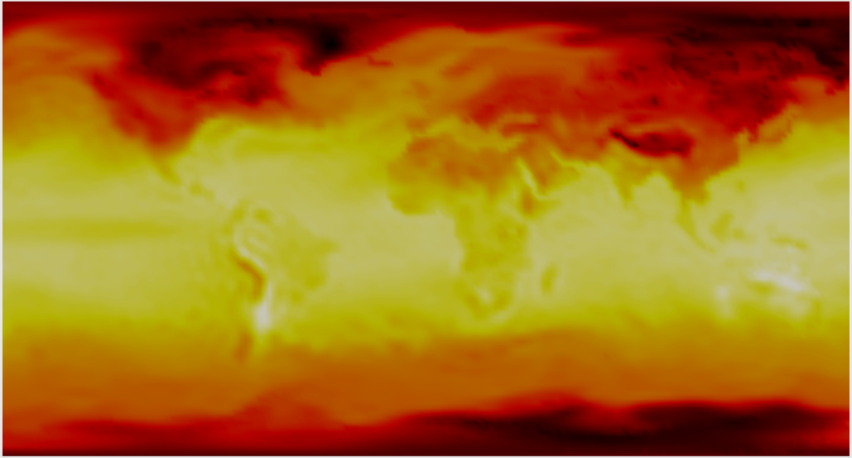
The choice of these costs is somewhat "natural" as the L_∞-distance is used for bottleneck distance calculations, for example.

Advantages of this dissimilarity measure

- Complexity of $\mathcal{O}\left(n^2 m \log m\right)$, where n is number of nodes in smaller hierarchy.
- Bottleneck distance
  - $\mathcal{O}\left((n+m)^3\right)$ (naïve)
  - $\mathcal{O}\left((n+m)^{1.5} \log(n+m)\right)$ (Kerber et al., "Geometry helps to compare persistence diagrams")

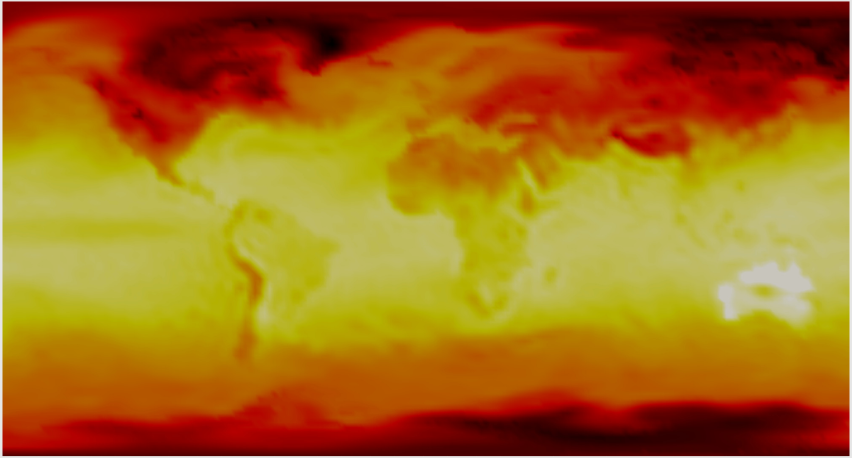
Results

Time-varying scalar field (climate model simulation)



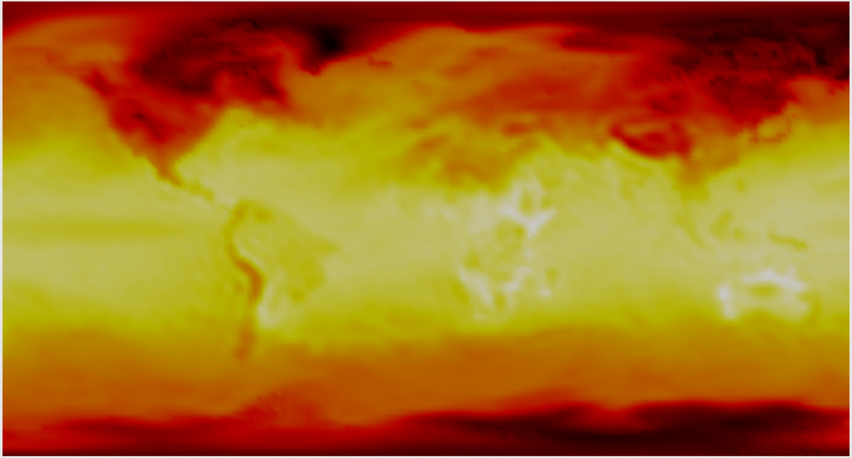
Results

Time-varying scalar field (climate model simulation)



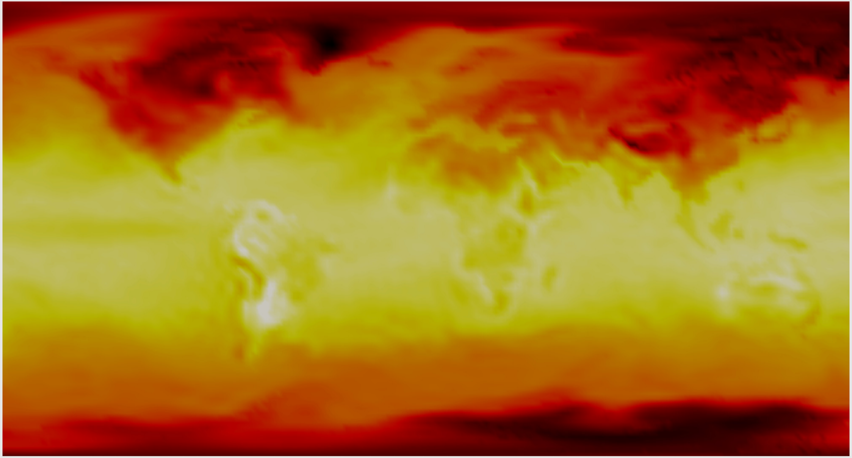
Results

Time-varying scalar field (climate model simulation)



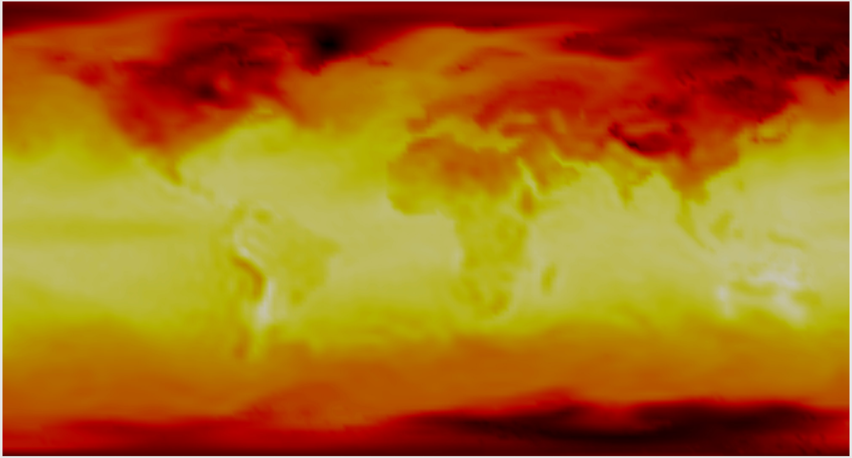
Results

Time-varying scalar field (climate model simulation)



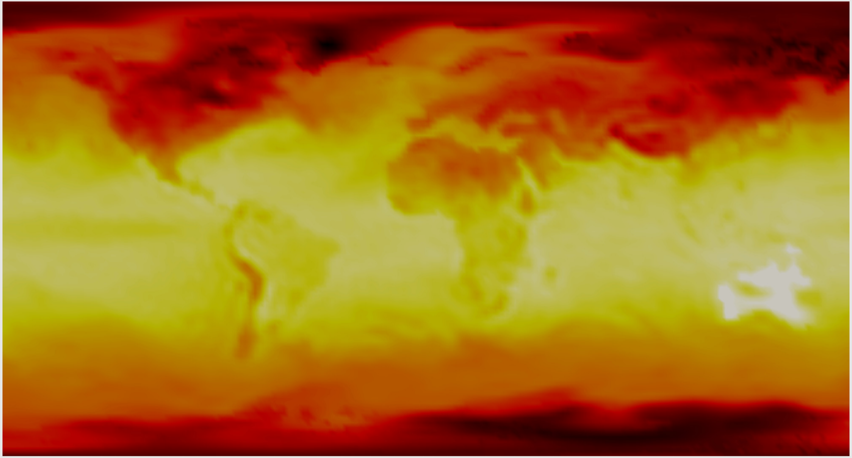
Results

Time-varying scalar field (climate model simulation)



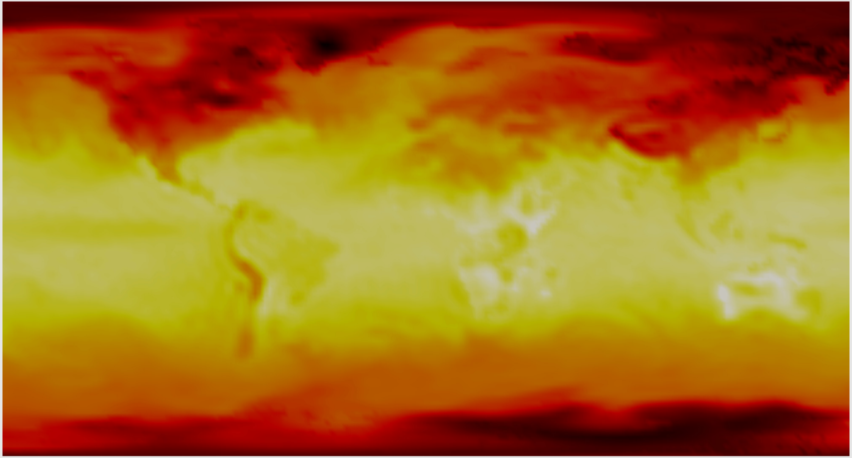
Results

Time-varying scalar field (climate model simulation)



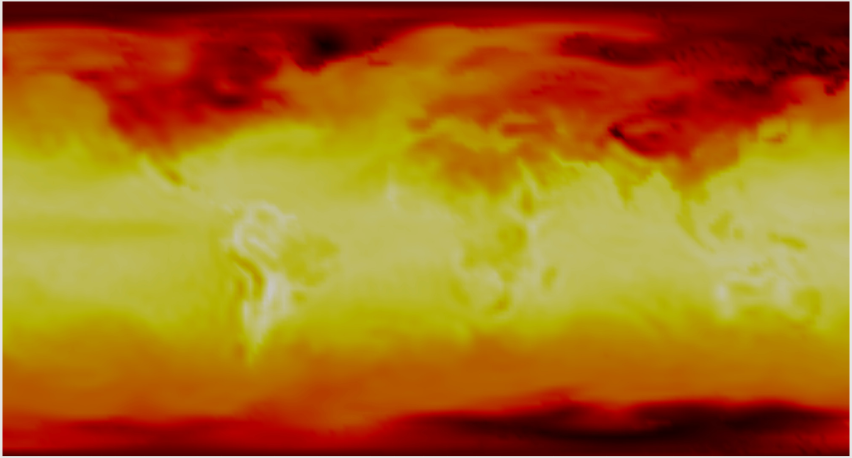
Results

Time-varying scalar field (climate model simulation)



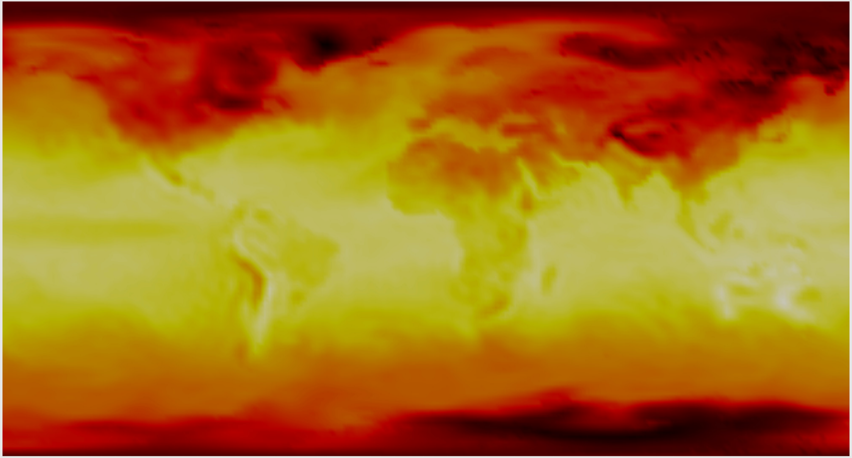
Results

Time-varying scalar field (climate model simulation)



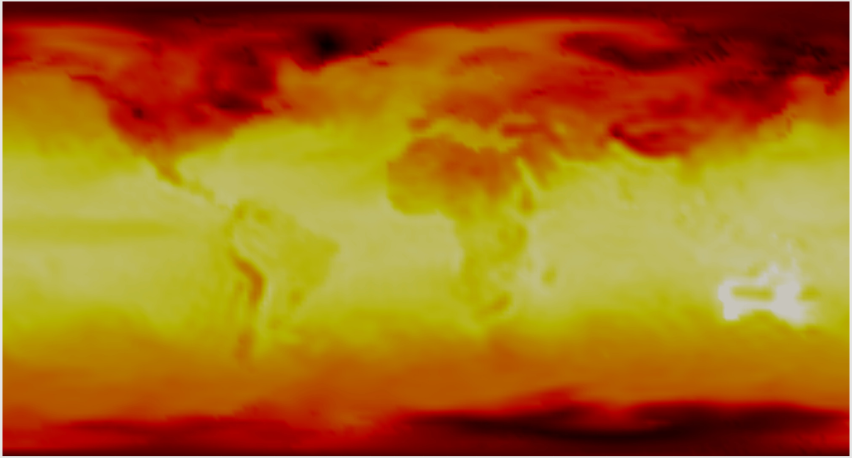
Results

Time-varying scalar field (climate model simulation)



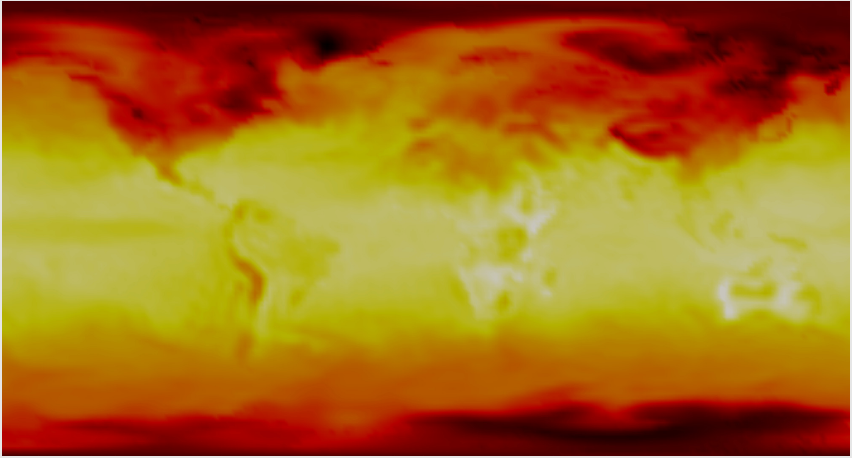
Results

Time-varying scalar field (climate model simulation)



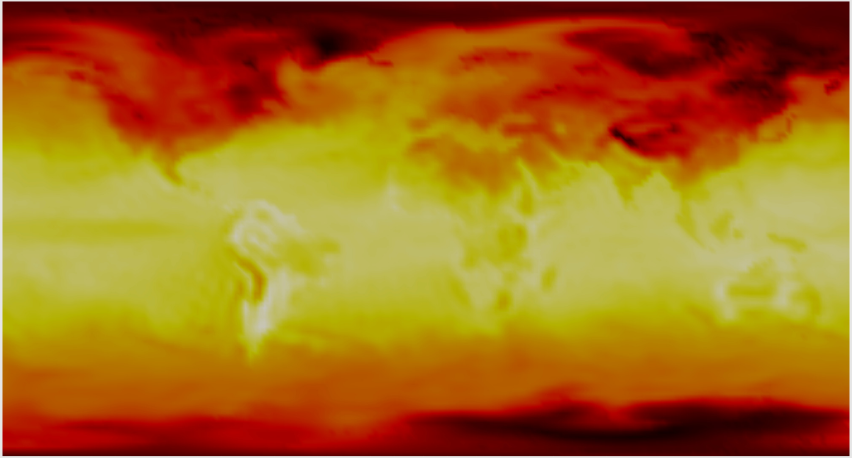
Results

Time-varying scalar field (climate model simulation)



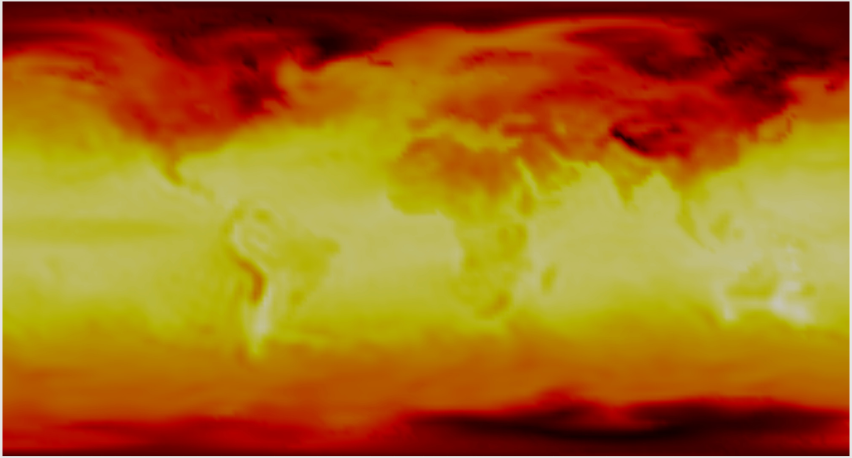
Results

Time-varying scalar field (climate model simulation)



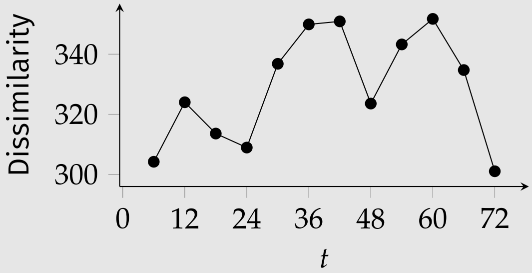
Results

Time-varying scalar field (climate model simulation)

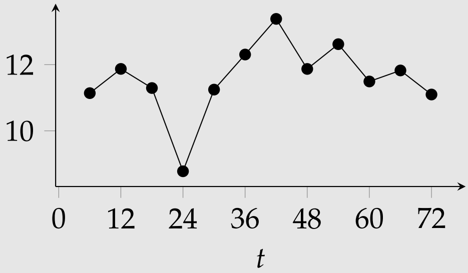


Results

Dissimilarity measure in comparison with second Wasserstein distance



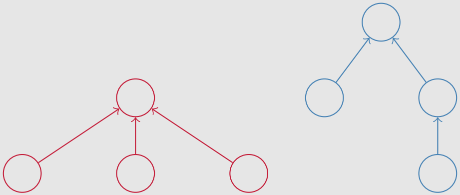
Tree edit distance



Second Wasserstein distance

Conclusion

An expressive novel hierarchy for relating zero-dimensional persistence pairs.



Open questions

- How to formalize the properties of the extended persistence hierarchy?
- Is it possible to extend the hierarchy to higher-dimensional topological features?