# Weighted Moving Window Score: a flexible spatial verification scoring method

Romain Boutelet[1], Julie Bessac[2]

[1]Michigan State University, [2]National Renewable Energy Laboratory

## Abstract

The verification (or scoring) of weather forecast fields presents many issue due to its inherent spatially correlated structure. As stressed by Gilleland et. al. [1], when the interest is to capture how well the prediction field match spatially coherent structures from the observation field, the use of straight-forward summary statistics such as the RMSE or MAE become irrelevant. Furthermore, the notion of "matching" spatially coherent structures has many aspects that need to be taken into account such as the scale, shape and amplitude of the spatial features.

## Motivating example

To highlight the need for a flexible scoring method, we use temperature anomaly fields during September 7, 2022 and September 8, 2022 in the Northern Great Plains. Although both fields have a similar sized "hotspot", its shape and direction are not similar, leading to stark differences when comparing the two fields point-by-point.

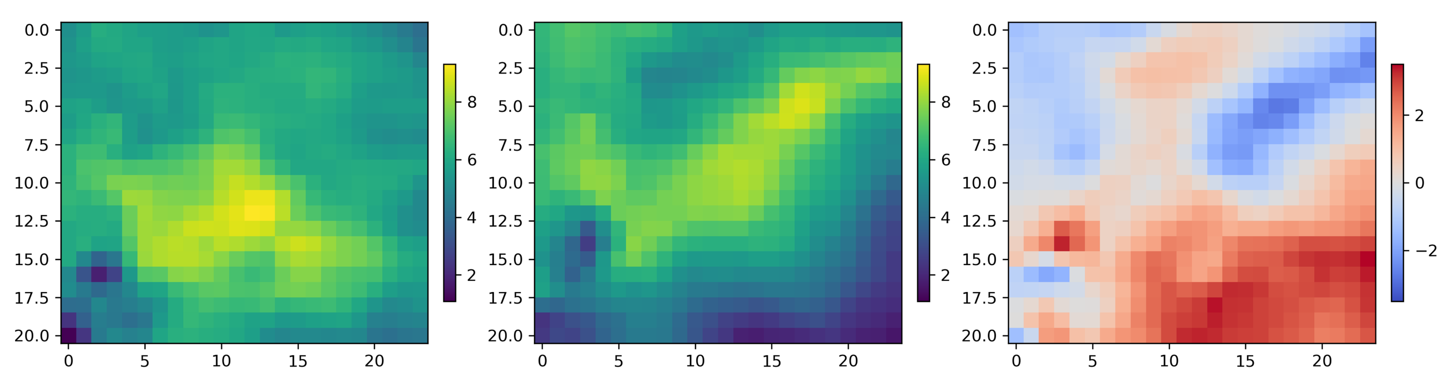

Figure 1. Two temperature anomaly fields (left and center) and their difference (right).

A more flexible scoring method should be able to capture this small spatial displacement with similar surroundings, and obtain a corresponding score comparatively lower than a point-by-point metric such as RMSE.

## Using Optimal Transport for Spatial Verification

Optimal Transport refers to the method that optimizes the transport of mass between a source distribution and a target distribution. It has been especially in recent years as a way of measuring the similarity of distributions. Recently, Nishizawa et. al. [2] have use an extension of optimal transport as a similarity metric in the context of spatial verification of precipitation field.

In more details, let $\mathbf{x}, \mathbf{y}$ be two vector fields of same total mass, ie. $\sum_{i,j} x_{i,j} = \sum_{i,j} y_{i,j}$. The optimal transport $\gamma$ between $\mathbf{x}$ and $\mathbf{y}$ under Euclidian distance transportation cost is defined as :

$$\gamma = \text{argmin}_\gamma \sum_{i,j,k,l} \gamma_{i,j,k,l} \|x_{i,j} - y_{k,l}\|_2^2$$

The total transportation cost $C$ is then the sum of all the individual transportation costs $\gamma_{i,j,k,l}$.

Since spatial fields do not generally have the same mass, we will use an extension of this optimal transport, where the vector fields have an absolute mass difference $m$. This Unbalanced Optimal Transport (UOT) is defined as the sum of the total transportation cost between the balanced vector fields $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, and the absolute mass difference between the two fields $m$:

$$\text{UOT}(\mathbf{x}, \mathbf{y}) = \min_\gamma \sum_{i,j,k,l} \gamma_{i,j,k,l} \|\tilde{x}_{i,j} - \tilde{y}_{k,l}\|_2^2 + |\sum_{i,j} x_{i,j} - \sum_{i,j} y_{i,j}|$$

## Defining a more flexible Spatial Verification Score

To address this issue, we introduce the **Weighted Moving Window Score** (WMWS), which compares the results from neighbouring windows from the observation field, and calculates a weighting average. The score can then be defined on each point $(i, j)$ as

$$\text{WMWS}(\hat{z}_{i,j}, z_{i,j}) = \sum_{(k',l') \in \mathcal{N}_s(k,l)} w(k,l,k',l') |\hat{z}_{i,j} - z_{i',j'}|$$

where $w(k,l,k',l') = \dfrac{w_{dist}(B_{i,j}, B_{i',j'}) w_{sim}(\hat{y}(B_{i,j}), y(B_{i',j'}))}{\sum_{(k'',l'') \in \mathcal{N}_s(k,l)} w_{dist}(B_{i,j}, B_{i'',j''}) w_{sim}(\hat{y}(B_{i,j}), y(B_{i'',j''}))}$, with:

- $w_{dist}(B_{i,j}, B_{i',j'}) = \dfrac{1}{\alpha + \|b_{i,j} - b_{i',j'}\|_2^2}$ is the measure of the distance between windows $B_{i,j}$ and $B_{i',j'}$. The parameter $\alpha$ can be tuned to give more or less weight to distant windows.
- $w_{sim}(\hat{y}(B_{i,j}), y(B_{i',j'})) = \dfrac{1}{\beta + \text{UOT}(\hat{y}(B_{i,j}), y(B_{i',j'}))}$ is the measure of the similarity between windows $B_{i,j}$ and $B_{i',j'}$. The parameter $\beta$ can be tuned to give more or less weight to spatially similar windows, in term of optimal transport.

The final score is the average of the score at every point: $\text{WMWS}(\hat{\mathbf{z}}, \mathbf{z}) = \sum i, j \text{WMWS}(\hat{z}_{i,j}, z_{i,j})$

## A simple example

To showcase the abilities of our method, we perform it on a simple example, where the two fields have the same continuous pattern, but with a diagonal displacement.

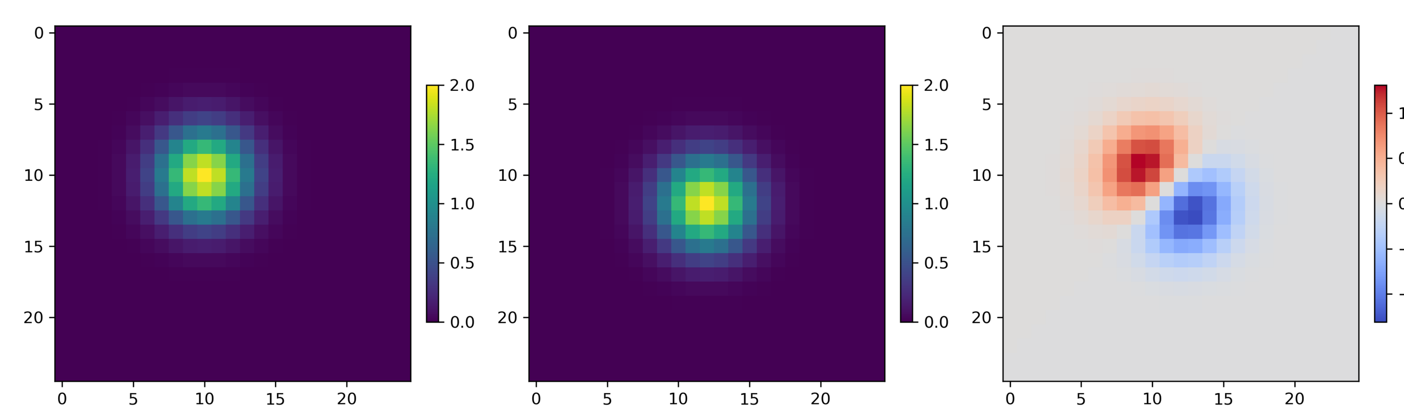

Figure 2. The two vector fields (left and center), and their difference (right).

We observe from Figure 3 that the spatial displacement is well captured by our method, and the resulting errors are reduced from a $\pm 1.5$ observed difference to a maximum of $0.2$ on the WMWS field.

An important observation from this example is the asymmetric behavior of our scoring method, more specifically that performing $\text{WMWS}(\mathbf{z1}, \mathbf{z2})$ will return a different score field than $\text{WMWS}(\mathbf{z2}, \mathbf{z1})$.

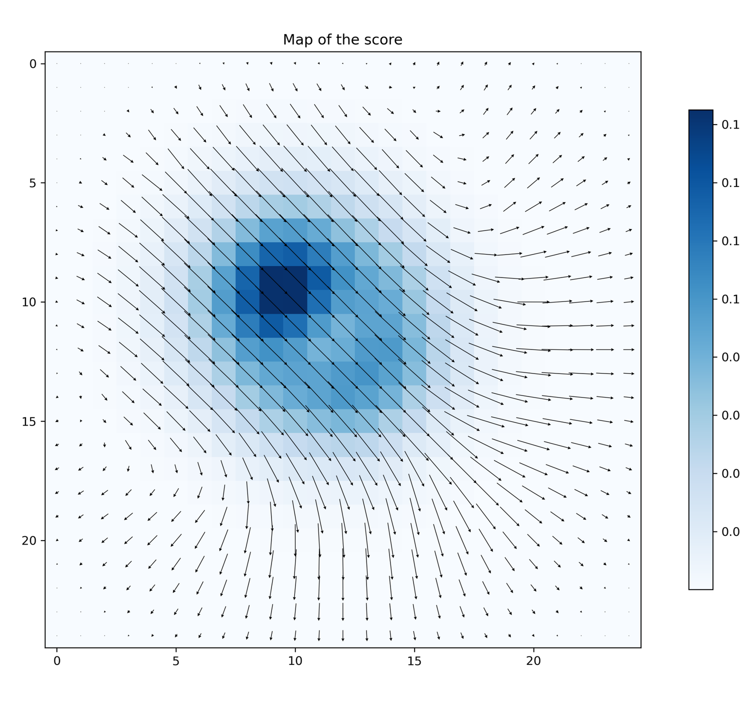


Figure 3. The score on the whole spatial field, and the directions given by the weights.

## Comparing the score to point-by-point metrics

We now want to compare the results of our method to the results from simple point-by-point metrics such as RMSE and MAE. To do so, we present first the results from two identical fields with white noise added to one of them as a benchmark to show that the order of magnitude is the same.

We then compare the results from the three metrics using the same example as previously but with varying spatial displacements.

| Deformation | WMWS | MAE | RMSE |
|---|---|---|---|
| White noise | 0.0733 | 0.0754 | 0.0958 |
| Diagonal displacement of 1 | 0.0115 | 0.0321 | 0.1115 |
| Diagonal displacement of 2 | 0.0156 | 0.0613 | 0.2076 |
| Diagonal displacement of 4 | 0.1142 | 0.1033 | 0.3230 |
| Diagonal displacement of 8 | 0.1503 | 0.1291 | 0.3613 |

Table 1. Comparison of score results across different deformation cases.

## Performing the WMWS on the motivating example

We come back to the motivating example that we introduced previously and show the results of out method on this example.

In Figure 4, we can see that our method substantially reduced the score along the "hotspot" and the most biggest relative share of the total score is due to the differences in amplitudes in the small "coldspots" in the bottom left corner.

In addition, we can compare our WMWS score to the other metrics.
The WMWS show a score of $0.9614$, while the MAE is $1.3542$ and the RMSE is $1.7062$.

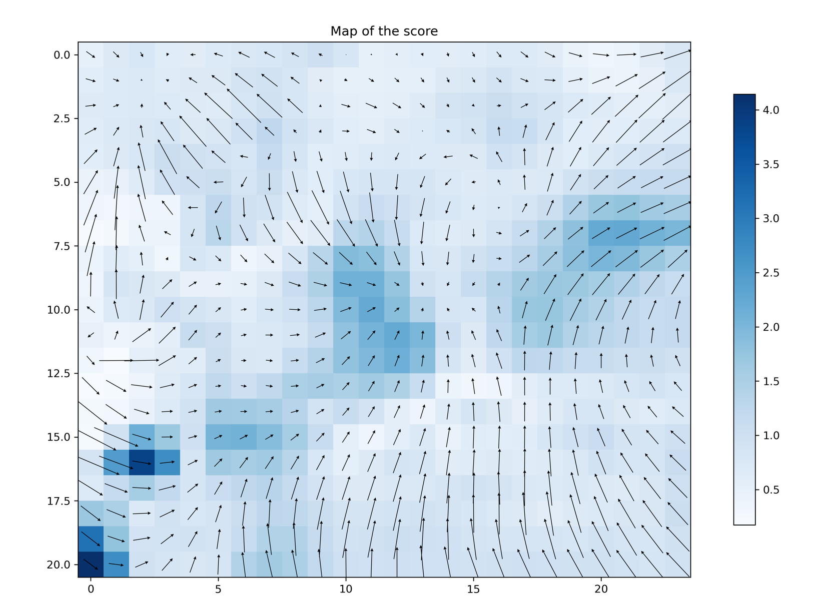


Figure 4. WMWS score on the whole fields, and the directions given by the weight.

## Conclusion

We constructed a scoring method that takes advantage of the flexibility of optimal transport methods in order to more accurately assess the forecast skill when spatial features are locally displaced.
More work needs to be done to accurately describe the interpretation of these score results, and on the potential additional information due to the non-symmetry of this scoring method.

## References

[1] Eric Gilleland, David A. Ahijevych, Barbara G. Brown, and Elizabeth E. Ebert.
Verifying forecasts spatially.
*Bulletin of the American Meteorological Society*, 91(10):1365 – 1376, 2010.

[2] S. Nishizawa.
Extracting latent variables from forecast ensembles and advancements in similarity metric utilizing optimal transport.
*Journal of Geophysical Research: Machine Learning and Computation*, 1(2):e2023JH000112, 2024.



boutelet@msu.edu