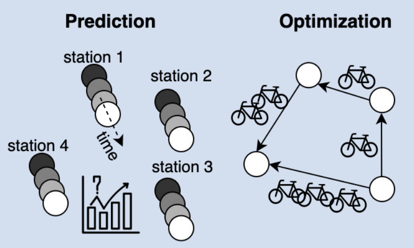
On the Potential of Optimal Transport in Geospatial Data Science

Nina Wiedemann[1], Martin Raubal[1]
[1]Mobility Information Engineering Lab, ETH Zurich
ICLR 2024 Workshop: Tackling Climate Change with Machine Learning

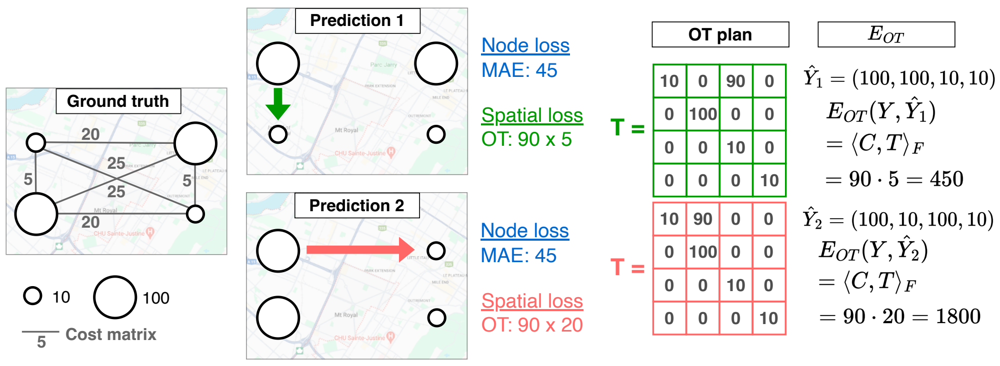
1 Introduction

- On-demand transportation services promise to reduce CO_2 emissions in the transport sector
- Usually, there are two steps towards improving efficiency: Prediction and optimization, e.g.
  - Predicting car sharing demand & optimal relocation
  - Predicting travel demand & optimal routing of on-demand buses
  - Predicting EV charging station occupancy & navigation
- ML research oftentimes ignores the use of predictions in downstream tasks
- How to measure operational costs of prediction errors?



2 An Optimal Transport based evaluation framework

- Idea: Measure *spatial displacement / misallocation* of predictions, i.e., the divergence between the predicted and true spatial distribution
- Optimal Transport: Mathematical framework for computing the difference between two distributions in terms of the *minimal cost* to align them



- Given locations $l_1, \dots, l_n$
- C_{ij} : 2D matrix with cost to transport mass from l_i to l_j
- Y^t: Spatiotemporal time series with ground truth values y_i^t for l_i at time t
- $\widehat{Y^t}$: Predictions $\hat{y}_i^t$ for l_i at time t
- Goal: Find the optimal transportation plan T (where T_{ij} is the mass that is moved from from l_i to l_j) to match predictions to ground truth
- Solved with a Linear Program

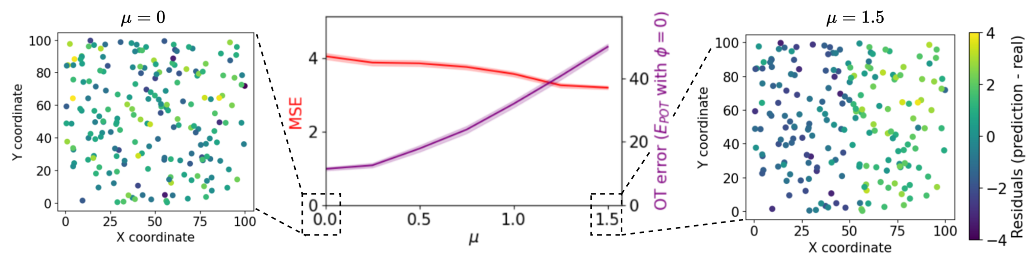
$$\min_T \sum_i \sum_j T_{ij} C_{ij} \ \ subject\ to\ \ T_{ij} \geq 0, \sum_j T_{ij} = \hat{y}_i, \sum_i T_{ij} = y_j$$

- OT error: $E_{OT} = \sum_i \sum_j T^*_{ij} C_{ij}$ where T^* is the solution of the LP
- **Partial OT:** allows for unbalanced data with cost for mass import / export

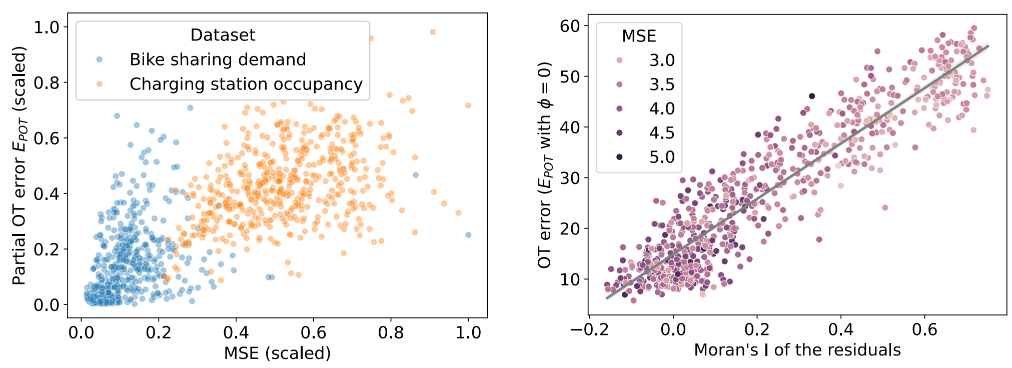
Example:

- $Y^t = (10, 100, 100, 10)$ is the demand for bike sharing at 4 stations
- A model predicts $\widehat{Y^t} = (100, 100, 10, 10)$
- A "mass" of 90 must be transported from location 1 to 3 to align the prediction with the ground truth → $E_{OT} = 90 \cdot C_{13}$

3 Results – Synthetic data



- Simulating error distributions
  → Intuition: If the demand is overestimated on one side of the city and underestimated on the other side, the spatial cost is larger than for a random error distribution
- Experiment:
  - Residuals drawn from $N(\mu, \sigma)$ for the left side (x<50) and from $N(-\mu, \sigma)$ for the right side
  - The larger μ, the higher the transportation cost
  - This is not reflected in the MSE
- For real data, the MSE correlates weakly with the OT error (see bottom left)
- Strong relation to spatial autocorrelation, measured with the Moran's I indicator (see bottom right)



4 Results – Training with a Sinkhorn loss

- Sinkhorn algorithm computes entropy-regularized OT [1] → differentiable
- Experiments on bike sharing & charging station occupancy datasets show that training with Sinkhorn loss helps to decrease the OT error

Application	Loss function	MSE	E_{OT}	E_{POT} (ϕ low)	E_{POT} (ϕ high)
Bike sharing demand	OT (Sinkhorn) loss	1.26	135.7	195.7	1733.8
	MSE loss	1.24	161.5	242.2	2406.1
Charging station occupancy	OT (Sinkhorn) loss	0.35	30.7	30.8	87.0
	MSE loss	0.34	32.7	30.7	81.1

5 Conclusion

Evaluating spatiotemporal predictions with Optimal Transport is

- *Interpretable*: quantifying operational costs or CO_2 emissions
- *Flexible:* applicable with arbitrary cost matrices
- *Spatially-explicit:* related to core GIS concepts such as autocorrelation

With the Sinkhorn loss, spatial costs can be minimized within model training

References:
[1] Cuturi, Marco. "Sinkhorn distances: Lightspeed computation of optimal transport." *Advances in neural information processing systems* 26 (2013).
Acknowledgements: We thank Thomas Klein and Jannis Born for their valuable comments