On the potential of Optimal Transport in geospatial data science

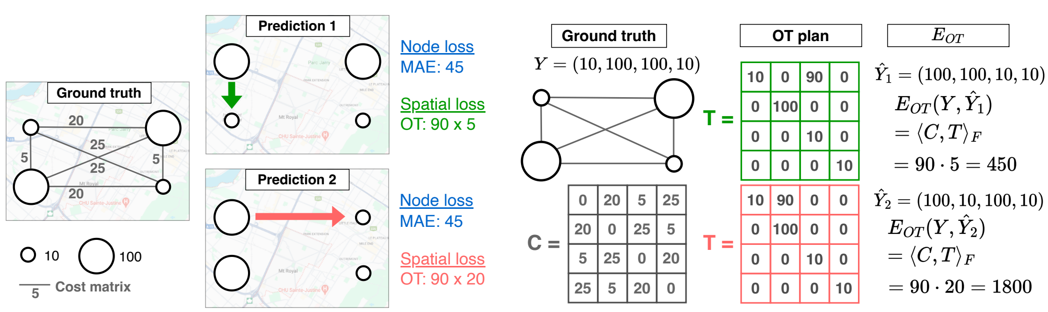
Nina Wiedemann
Chair of Geoinformation Engineering
ETH Zurich
Switzerland
nwiedemann@ethz.ch

Martin Raubal
Chair of Geoinformation Engineering
ETH Zurich
Switzerland

Abstract

Prediction problems in geographic information science and transportation are often motivated by the possibility to enhance operational efficiency and thereby reduce emissions. Examples range from predicting car sharing demand for relocation planning to forecasting traffic congestion for navigation purposes. However, conventional accuracy metrics ignore the spatial distribution of the errors, despite its relevance for operations. Here, we put forward a spatially aware evaluation metric and loss function based on Optimal Transport (OT). Our framework leverages partial OT and can minimize relocation costs in any spatial prediction problem. We showcase the advantages of OT-based evaluation over conventional metrics and further demonstrate the application of an OT loss function for improving forecasts of bike sharing demand and charging station occupancy. Thus, our framework not only aligns with operational considerations, but also signifies a step forward in refining predictions within geospatial applications. All code is available at https://github.com/mie-lab/geospatial_optimal_transport.

1 Introduction



(a) Standard metrics neglect spatial costs

(b) Computation of the OT error

Figure 1: Optimal transport for evaluating spatiotemporal predictions. In contrast to other metrics, our OT framework takes into account spatial distances and the related costs or emissions (a). In detail, the costs of prediction errors are quantified in terms of the required resource reallocation, based on the optimal transportation plan mapping from predictions to ground truth (b).

The transport sector accounts for 20% of CO_2-emissions worldwide (Statista) and thus plays a key role in climate action. One possible avenue to reducing emissions is the adoption of on-demand services such as (autonomous) car sharing, which was shown to effectively reduce car ownership (Mishra et al., 2015; Martin and Shaheen, 2011; Liao et al., 2020). There are two main research avenues to improving on-demand transport services: Prediction (Nguyen et al., 2018), e.g., estimating the number of shared cars/bicycles that will be picked up in the next hour, and optimization, e.g., computing the most efficient way to re-distribute bikes/cars. Importantly, good predictions only lead to a reduction of emissions if the system is optimized with respect to the predicted demand.

Meanwhile, machine learning (ML) research in geographic information sciences (GIS) or transportation usually treats prediction as a standalone problem, ignoring its role in downstream tasks (Yan and Wang, 2022). Consider the example of forecasting bike sharing demand per hour and per station. Usually, a time series prediction model such as an LSTM is trained on the data and the prediction quality is evaluated via the mean squared error (MSE) or mean absolute percentage error (MAPE) (Hulot et al., 2018; Brahimi et al., 2022; Shin et al., 2020; Ma and Faye, 2022), since evaluating the resulting CO_2-efficiency or business costs is cumbersome. Crucially, such metrics only quantify the error per time step and station, but ignore the spatial distribution of residuals and their implications in a production setting involving relocation costs, as illustrated in Figure 1a. Critically, these costs depend on the distance between erroneous predictions and can be viewed either as a *resource* relocation or as a *user* relocation that is necessary due to prediction errors.

We propose Optimal Transport (OT) to approximate and minimize relocation costs and thereby the involved emissions. OT provides methods to measure the disparity between two (probability) distributions, which can be leveraged as an evaluation framework comparing the real and predicted spatial distribution in any spatiotemporal prediction task, such as estimating bike sharing demand, traffic congestion or charging station occupancy. Moreover, we demonstrate how the relocation costs can be directly minimized with an OT-based loss function. Our framework is based on partial OT (Guittet, 2002; Piccoli and Rossi, 2014; Maas et al., 2015) and provides important tools to researchers and industry working with spatiotemporal data to achieve actual advances in resource management and operational efficiency with ML methods.

2 Methods

2.1 Optimal transport for evaluating spatiotemporal predictions

Optimal Transport dates back to 1781, when Gaspard Monge pondered on the following question: Given a pile of earth, what is the most efficient way to redistribute it to a desired shape (Monge, 1781)? Originally termed Earth Mover's Distance (EMD) and later generalized to Optimal Transport (OT) theory (Tolstoi, 1930; Cook et al., 1998), this field is concerned with quantifying the divergence between two distributions in terms of the lowest cost necessary to transport the initial to the target distribution. The first efficient solution approach only became available with the introduction of linear programming (LP) (Villani, 2021). OT recently found extensive application in ML research (Peyré et al., 2019; Khamis et al., 2023) due to its ability to measure the divergence between two probability distributions with the *Wasserstein distance* (Vaserstein, 1969; Kantorovich, 1960), as well as due to important speedups in approximating the optimal transportation plan with the Sinkhorn algorithm (Cuturi, 2013).

In this work, we apply OT to evaluate spatiotemporal predictions at discrete locations. This setting allows to define the EMD between *signatures*: Given a fixed set of locations $l_i,\ i \in [1..n]$, let P and Q describe the initial and the desired spatial distribution of mass respectively, where p_i is the initial mass at l_i and q_i the corresponding target mass. For instance, p_i could be the car sharing demand at l_i. With OT, the minimal transportation costs for re-distributing the mass from P to match the distribution in Q is computed, based on a given cost matrix C. We first consider *balanced* OT assuming $\sum_i p_i = \sum_i q_i$. The EMD is computed by solving a linear program (LP) with the following objective and constraints:

$$\text{minimize}_T \sum_i^n \sum_j^n T_{ij} C_{ij} \quad s.t.\ \ T_{ij} \geq 0 \ \ \forall i,j \in [1...n], \ \ \sum_{j=1}^n T_{ij} = p_i \ \ \forall i, \ \ \sum_{i=1}^n T_{ij} = q_j \ \ \forall j$$

where T is the transportation plan to be optimized. The solution of the LP, T^*_{ij}, indicates how much mass should be transported from the i-th to the j-th location, with the constraints ensuring that the marginals of T^* correspond to P and Q respectively. C_{ij} is the cost for one unit of mass to be moved from l_i to l_j. In the simplest case, C is set to the Euclidean distance between the locations defined by two-dimensional geographic coordinates $l_i = (u_i, v_i)$, such that $C_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$. However, C could also express monetary costs, map-matched driving distances, or CO_2-emissions. We define $E_{OT}(P, Q)$ as the minimal total cost for transporting all mass from P to Q based on the optimal transportation plan T^*:

$$E_{OT}(P, Q) = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} T^*_{ij} \tag{1}$$

To apply OT for evaluating spatio-temporal predictions, we compute the minimal transportation cost $E_{OT}(\hat{Y}^t, Y^t)$ between the predicted spatial distribution $\hat{Y}^t = \{\hat{y}^t_1, \ldots, \hat{y}^t_n\}$ and the true distribution $Y^t = \{y^t_1, \ldots, y^t_n\}$ over all locations at time t. For the sake of simplicity, we omit the time indices t in the following. The approach is shown in Figure 1b.

2.2 Partial Optimal Transport to consider the total prediction error

The standard formulation of OT assumes the total mass of P and Q to be equal, which is unlikely if P contains predictions for Q. An alternative is *partial OT* (Guittet, 2002; Piccoli and Rossi, 2014; Maas et al., 2015). In partial OT, so-called "dustbin"(Sarlin et al., 2020; Dang et al., 2020) or "waste vectors" (Guittet, 2002) are added to C and T, allowing mass to vanish or to emerge with a certain cost; here denoted ϕ. Fortunately, the partial OT problem can be reduced to balanced OT by modifying the inputs. Specifically, let $\widetilde{C}(\phi) \in \mathbb{R}^{(n+1)\times(n+1)}$ be the resulting cost matrix when enlarging C by one column and one row both filled with ϕ. Furthermore, let δ denote the total mass difference $\delta = \sum_i q_i - \sum_i p_i$. We define $\widetilde{P} \in \mathbb{R}^{n+1}$ by extending P with $\widetilde{p}_{n+1} = \max(\delta, 0)$ and $\widetilde{Q}$ analogously with $\widetilde{q}_{n+1} = \max(\delta, 0)$. By design, the total masses of $\widetilde{P}$ and $\widetilde{Q}$ are equal. The partial OT error is then simply $E^{\phi}_{POT}(P, Q, C) = E_{OT}(\widetilde{P}, \widetilde{Q}, \widetilde{C}(\phi))$. For a full definition of the variables, see Appendix A.

In our evaluation framework, E_{POT} can be seen as a combination of the total error over all locations and the distributional error due to the required relocation between locations. The smaller ϕ, the more focus lies on the distributional errors.

2.3 OT-based loss functions to improve spatiotemporal predictions

A natural progression for the OT error as an evaluation criterion is its integration into the *training* of neural networks as a spatial loss function. However, computing the EMD requires solving an LP and is not differentiable. Entropy-regularized OT and the Sinkhorn algorithm (Cuturi, 2013) enabled the employment of OT in deep learning applications (Cao et al., 2021; Genevay et al., 2018; Wong et al., 2019). We implement a Sinkhorn loss function for spatiotemporal prediction using the the `geomloss` package (Feydy et al., 2019). The loss function is plugged into an established time series prediction model, N-HiTS (Challu et al., 2022), implemented in the `darts` library (Herzen et al., 2022). While OT has been used for evaluating spatio-temporal forecasts (Roberts et al., 2017), to the best of our knowledge, this is the first attempt to improve forecasts of geospatial data with an OT loss function.

3 Comparison of OT-based evaluation with the MSE

We first design a synthetic scenario to demonstrate the benefits of OT in evaluating spatiotemporal predictions. Synthetic data is generated by sampling locations randomly from a uniform distribution $l \sim U[0, 100]$ and drawing the corresponding labels y from a normal distribution $\mathcal{N}(10, 3)$. The predictions are simulated with varying *spatial imbalance* of the residuals, i.e. over and underestimation of y dependent on their spatial location. Such imbalance is very common in real data due to spatial autocorrelation and spatial heterogeneity (Zhang et al., 2009). As a simple scenario, the residuals are sampled from different distributions dependent on the x-coordinate of the location; namely from $\mathcal{N}(\mu, \sigma)$ for $x < 50$ and $\mathcal{N}(-\mu, \sigma)$ for $x \geq 50$. Figure 2 illustrates our experimental setup with two examples for $\mu = 0$ and $\mu = 1.5$ and presents the results for $\mu \in \{0, 0.5, 1, 1.5\}$. σ was tuned to keep the average absolute value of the residuals constant. In contrast to the MSE that can not reflect spatial imbalance,

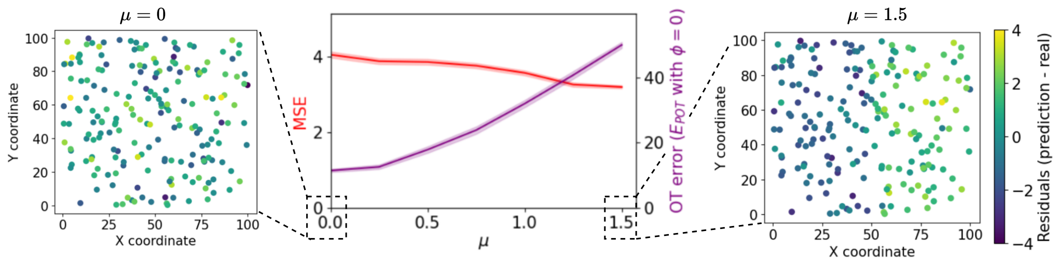


Figure 2: Intuitively, the costs to align ground truth and predictions are higher if the residuals are unevenly distributed in space. We construct synthetic data with increasingly unbalanced distribution ($\mu = 0$: no imbalance - left, $\mu = 1.5$: strong imbalance - right). In contrast to the MSE, the OT error reflects the increasing spatial costs.

the OT distance increases substantially due to the increased transportation cost between predicted and real values. In Appendix B, we further show how the OT error relates to spatial autocorrelation.

4 Minimizing relocation costs with an OT-based loss function

We consider two use cases to demonstrate an OT-based loss function on real data. First, we utilize a public dataset from the BIXI bike sharing service in Montreal. Prediction errors in bike sharing demand lead to relocation costs and emissions, which should be reflected in the loss function. For our experiment, the number of bike pickups at 458 stations is aggregated by hour and by station following Hulot et al. (2018). As a second case study, we regard the task of predicting the occupancy of charging stations for electric vehicles (EVs). In this case, prediction errors lead to relocation costs for drivers of EVs who find the charger occupied. A suitable dataset was published by Amara-Ouali et al. (2023), providing the charging station occupancy for 83 stations in France from July 2020 to February 2021 at a granularity of 15 minutes. In both cases, NHiTS models were trained to predict the demand for the next five time steps, either employing a normal MSE loss or a Sinkhorn loss function. Details on data preprocessing and model training are provided in Appendix C.

Table 1 presents the results on test data in terms of the MSE, the balanced OT cost (scaling Y and $\hat{Y}$ to have equal sum), and E_{POT} with small and large ϕ (see Appendix D for the choice of ϕ and its effect on E_{POT}). The cost matrix C was set to the Euclidean distance between stations in km. For example, relocation effort of around 135.7km is required in total to align the predicted bike sharing demand with the true. As desired, training with an OT-based loss function decreases the OT costs, with minor effect on the station-wise MSE. For the charging station case, the total error δ is minimized better with the MSE, such that only E_{OT} decreases with a Sinkhorn loss, whereas for bike sharing demand prediction, all OT-based metrics are decreased significantly.

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

Table 1: Results when training with an OT-based loss function. At minor increase of the MSE, OT-based metrics can be decreased substantially; e.g., from 161.5 to 135.7km bike relocation cost.

5 Conclusion

This paper proposes to evaluate spatio-temporal predictions with Optimal Transport, highlighting its capacity to reflect reductions in operational costs and emissions within predictive methods. Our experiments on synthetic and real data demonstrate the value of OT for evaluating and training prediction models. The proposed framework is generally applicable to any prediction problem where the spatial distribution of the errors matters. A notable limitation is the computational demand of computing the EMD and the Sinkhorn loss, particularly in cases involving numerous locations. The potential of OT in GIS and transportation extends further, such as its extension to the temporal dimension considering relocation across space *and* time.

Acknowledgements

We would like to thank Jannis Born and Thomas Klein for their helpful comments.

References

Y. Amara-Ouali, Y. Goude, N. Doumèche, P. Veyret, A. Thomas, D. Hebenstreit, T. Wedenig, A. Satouf, A. Jan, Y. Deleuze, et al. Forecasting electric vehicle charging station occupancy: Smarter mobility data challenge. *arXiv preprint arXiv:2306.06142*, 2023.

N. Brahimi, H. Zhang, L. Dai, and J. Zhang. Modelling on car-sharing serial prediction based on machine learning and deep learning. *Complexity*, 2022:1–20, 2022.

T. Cao, A. Bie, A. Vahdat, S. Fidler, and K. Kreis. Don't generate me: Training differentially private generative models with Sinkhorn divergence. *Advances in Neural Information Processing Systems*, 34:12480–12492, 2021.

C. Challu, K. G. Olivares, B. N. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski. N-HiTS: Neural hierarchical interpolation for time series forecasting, 2022. arXiv:2201.12886 [cs].

W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver. Combinatorial optimisation. *Wiley-Interscience Series in Discrete Mathematics and Optimization, USA*, 1:998, 1998.

M. Cuturi. Sinkhorn distances: Lightspeed computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Z. Dang, F. Wang, and M. Salzmann. Learning 3d-3d correspondences for one-shot partial-to-partial registration. *arXiv preprint arXiv:2006.04523*, 2020.

J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouve, and G. Peyré. Interpolating between Optimal Transport and MMD using Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690, 2019.

A. Genevay, G. Peyré, and M. Cuturi. Learning generative models with sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617. PMLR, 2018.

K. Guittet. Extended Kantorovich norms: a tool for optimization. *Dissertation, INRIA*, 2002.

J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. Van Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, et al. Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*, 23(1):5442–5447, 2022.

P. Hulot, D. Aloise, and S. D. Jena. Towards station-level demand prediction for effective rebalancing in bike-sharing systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 378–386, 2018.

L. V. Kantorovich. Mathematical methods of organizing and planning production. *Management science*, 6(4):366–422, 1960.

A. Khamis, R. Tsuchida, M. Tarek, V. Rolland, and L. Petersson. Earth movers in the big data era: A review of Optimal Transport in machine learning. *arXiv preprint arXiv:2305.05080*, 2023.

F. Liao, E. Molin, H. Timmermans, and B. van Wee. Carsharing: the impact of system characteristics on its potential to replace private car trips and reduce car ownership. *Transportation*, 47(2):935–970, 2020.

T.-Y. Ma and S. Faye. Multistep electric vehicle charging station occupancy prediction using hybrid lstm neural networks. *Energy*, 244:123217, 2022.

J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for Optimal Transport of images including dissipation and density modulation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1745–1769, 2015.

E. Martin and S. Shaheen. The impact of carsharing on household vehicle ownership. *Access Magazine*, 1(38):22–27, 2011.

G. S. Mishra, R. R. Clewlow, P. L. Mokhtarian, and K. F. Widaman. The effect of carsharing on vehicle holdings and travel behavior: A propensity score and causal mediation analysis of the san francisco bay area. *Research in Transportation Economics*, 52:46–55, 2015.

G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

H. Nguyen, L.-M. Kieu, T. Wen, and C. Cai. Deep learning methods in transportation domain: a review. *IET Intelligent Transport Systems*, 12(9):998–1004, 2018.

G. Peyré, M. Cuturi, et al. Computational Optimal Transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211:335–358, 2014.

L. Roberts, L. Razoumov, L. Su, and Y. Wang. Gini-regularized Optimal Transport with an application to spatio-temporal forecasting, 2017. arXiv:1712.02512 [cs, stat].

P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.

D.-H. Shin, K. Chung, and R. C. Park. Prediction of traffic congestion based on lstm through correction of missing temporal and spatial data. *IEEE Access*, 8:150784–150796, 2020.

Statista. Transportation emissions worldwide. URL https://www.statista.com/topics/7476/transportation-emissions-worldwide/#topicOverview.

A. Tolstoi. Methods of finding the minimal total kilometrage in cargo transportation planning in space. *TransPress of the National Commissariat of Transportation*, 1:23–55, 1930.

L. N. Vaserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72, 1969.

C. Villani. *Topics in Optimal Transportation*, volume 58. American Mathematical Soc., 2021.

E. Wong, F. Schmidt, and Z. Kolter. Wasserstein adversarial examples via projected Sinkhorn iterations. In *International Conference on Machine Learning*, pages 6808–6817. PMLR, 2019.

R. Yan and S. Wang. Integrating prediction with optimization: Models and applications in transportation management, 2022.

L. Zhang, Z. Ma, and L. Guo. An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *Forest Science*, 55(6):533–548, 2009.

A Partial OT

Partial OT allows to create or remove mass at a certain cost. For the purpose of designing an evaluation metric for spatiotemporal predictions, we argue that *all* available mass should be transported if possible and only the *difference* between predicted and true values is removed or added under penalty. The difference corresponds to $\delta = \sum_i q_i - \sum_j p_j$. Thus, we extend C, P and Q as the following:

$$\widetilde{C}(\phi) = \begin{pmatrix} c_{11} & \dots & c_{1n} & \phi \\ \dots & \ddots & \dots & \dots \\ c_{n1} & \dots & c_{nn} & \phi \\ \phi & \dots & \phi & 0 \end{pmatrix} \tag{2}$$

$$\tilde{p_i} = \begin{cases} p_i & i \leq n \\ \delta & i = n \text{ and } \delta > 0 \\ 0 & else \end{cases} \tag{3}$$

$$\tilde{q_j} = \begin{cases} q_j & j \leq n \\ -\delta & j = n \text{ and } \delta < 0 \\ 0 & else \end{cases} \tag{4}$$

Applying balanced OT on the modified versions of $\widetilde{P}, \widetilde{Q}$ and $\widetilde{C}$ yields a transportation matrix that contains the flow of mass between locations, as well as the outflow or inflow dependent on δ. It is worth noting that the penalty ϕ could easily be defined in a location-dependent manner; i.e., penalizing the import / export to some locations more than to others. For instance, this could be useful when considering predictions of bike sharing demand, where bikes are re-distributed from a central hub location.

B Relation between OT error and Moran's I

As shown in Figure 2, the OT costs are naturally higher if the errors are unevenly distributed in space. This strongly relates to the concept of spatial autocorrelation; one of the most fundamental concepts in GIS. There are several measures to quantify spatial autocorrelation in a dataset, with global Moran's I arguably the most popular one. Moran's I is defined as:

$$I = \frac{n}{\sum_{ij} w_{ij}} \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{5}$$

where x_i is the i-th observation, $\overline{x}$ is the mean of all observed values, n is the number of samples, and w_{ij} is the (distance-based) weighting between to points.

We argue that Moran's I computed on the *residuals* of a prediction strongly relates to the OT cost between prediction and ground truth. Indeed, Figure 3 testifies a pronounced correlation between Moran's I and the OT loss (Pearson $r = 0.93$), that is hardly related to the MSE.

C Data preprocessing and training details

The bike sharing dataset was downloaded from Kaggle[1] and restricted to the period from 15th of April to 15th of November 2014, since the service is closed in winter, leading to large gaps in the time series across years. Only stations with missing coordinates or maintenance stations were removed.

The charging station occupancy dataset was published by Amara-Ouali et al. (2023) in the context of the "Smarter Mobility Data Challenge". Each charging station has three plugs and the challenge is to classify the state of each plug as "available", "charging", "passive" (plug is connected to a car that is fully charged) and "other" (out of order). Here, we frame the task as a regression problem of predicting the fraction of plugs that are occupied, i.e., *charging* or

[1] https://www.kaggle.com/datasets/aubertsigouin/biximtl

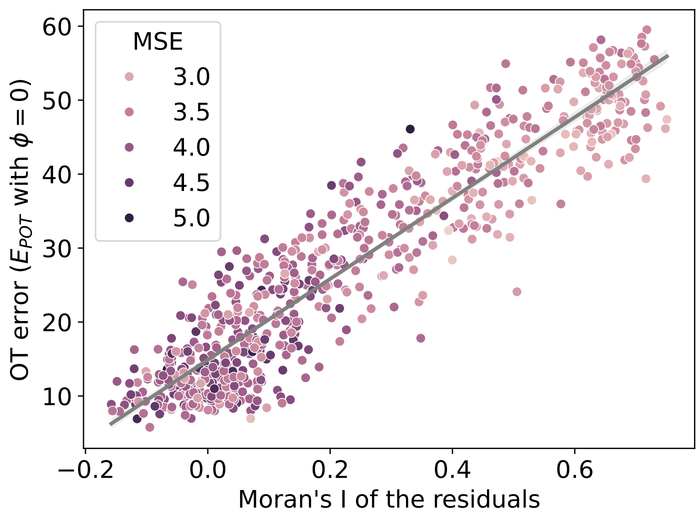


Figure 3: Moran's I correlates with the OT error

passive. The forecasts could help to estimate the energy demand and to facilitate planning of charging stops for owners of electric vehicles.

The data is given at a granularity of 15 minutes from 3rd of July 2020 to 18th of February 2021. The time series is comparably sparse, since a station on average has no plugs in use 61% of the time; one out of three plugs in use by 27%, and only 2.1% where all three plugs are used. From 2020-10-22 onwards, there is also a considerable number of missing data, amounting to 8% of missing information on the number of chargers in use. We execute the preprocessing pipeline[2] of the winning team of the "Smarter Mobility Data Challenge", who employ the exponential moving weighted average to fill missing values. We further removed stations with no charging activity, leaving 83 charging stations. Finally, we scale all values by dividing by 3 for training the model.

The N-HiTS model is trained for 100 epochs with early stopping. The learning rate was set to $1e^{-5}$. The time series was treated as multivariate data with one variable per bike sharing station or charging station. A lag of 24 is used to learn daily patterns, and the hour and weekday are provided as past covariates. The number of stacks in the N-HiTS model was set to 3. The number of output time steps corresponds to our forecast horizon of five time steps. For evaluation, we draw 100 samples from the test data (last 10% of the time series) and predict the next five time steps based on the respectively preceding time series, without re-training the model. For further implementation details, we refer to our source code.

D Partial OT strikes a balance between evaluating distributional and total costs

Intuitively, partial OT strikes a balance between balanced OT (measuring the mismatch between the predicted and true distribution) and the total error δ (mismatch between the sum of predicted and the sum of observed values). The weighting between both depends on ϕ. In Table 1, we reported the results for the partial OT error with low ϕ, specifically setting ϕ to the 0.1-quantile of all pairwise costs C_{ij}, and high ϕ, where ϕ is set to the maximum of the cost matrix $\phi = \max_{ij} C_{ij}$. The reasoning of these parameter settings is illustrated in Figure 4, showing the OT error of the model trained on predicting charging data occupancy with a Sinkhorn loss. In particular, Figure 4 illustrates the dependence of E_{POT} on ϕ. For

[2]Available on GitHub:https://github.com/arthur-75/Smarter-Mobility-Data-Challenge

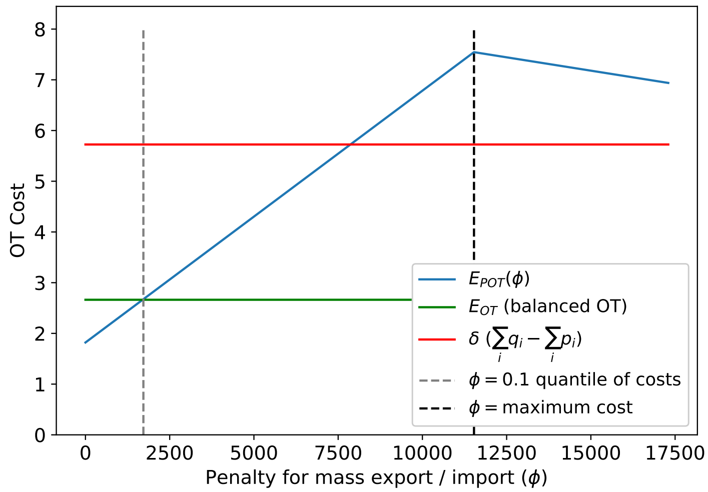


Figure 4: Relation of E_{POT} to ϕ in one synthetic example

ensuring comparability of E_{OT} and E_{POT}, the extended cost matrix $\widetilde{C}$ was normalized by its maximum for this illustration.

We observe that the E_{POT} approximately corresponds to E_{OT} when ϕ is set to the 0.1-quantile of C (intersection of green and blue lines in Figure 4). This observation is consistent for synthetic data as well as the bike sharing dataset. The reason is that with $\phi = 0$, only the spatial distribution would be penalized, but some mass could be imported to / exported from arbitrary locations for free. Thus, $\phi = 0$ leads to lower errors than E_{OT}.

On the other hand, for $\phi \longrightarrow \infty$, all entries of C become zero except for the last row and column which is 1, since all values are divided by ϕ when normalizing by the maximum. Thus, E_{POT} converges to δ for large ϕ (blue line approaching red line). When $\phi = \max_{ij} C_{ij}$, the partial OT error is maximal since E_{OT} is combined with δ.