ML-based Prediction for Train Delays

Towards Uncertainty Quantification

Clément Mantoux

Journée SFdS UQML, 8 Janvier 2025

GROUPE

Thanks to our interns!

Samuel Bazaz

Jawad Chemaou

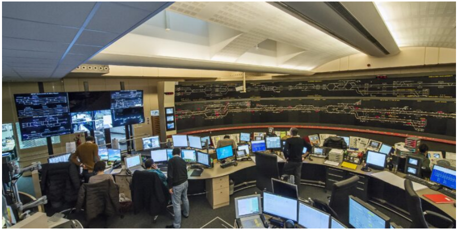
Boubacar Sow

Motivation

Why are we interested in predicting train delays?

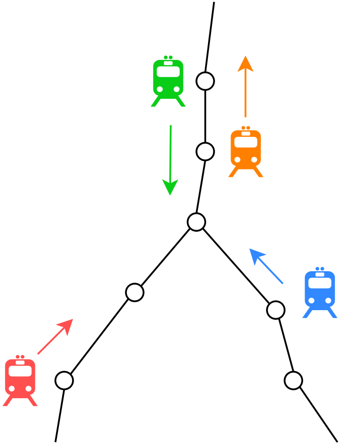
Passenger information

Traffic regulation



Goal

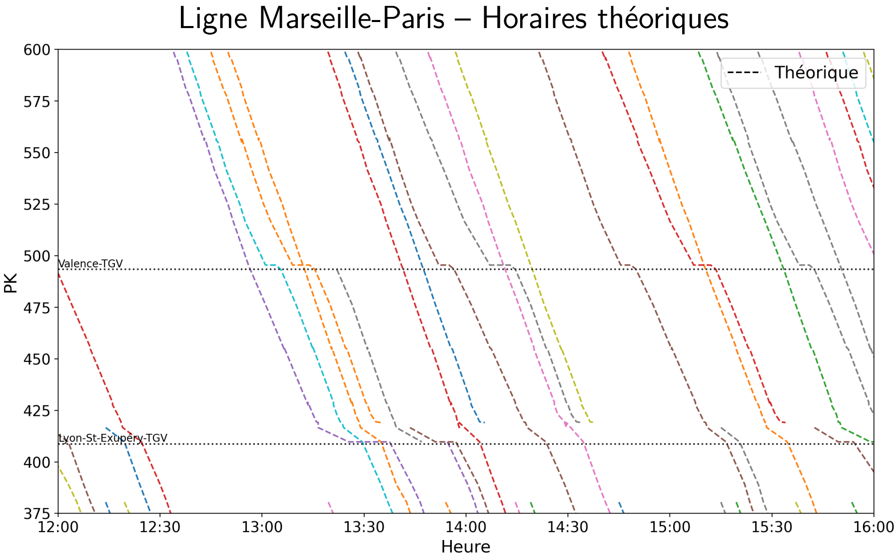
- For each train in the **network**,
- Given their **position** and **schedule**,
- Predict **delay propagation**
- Provide a confidence interval



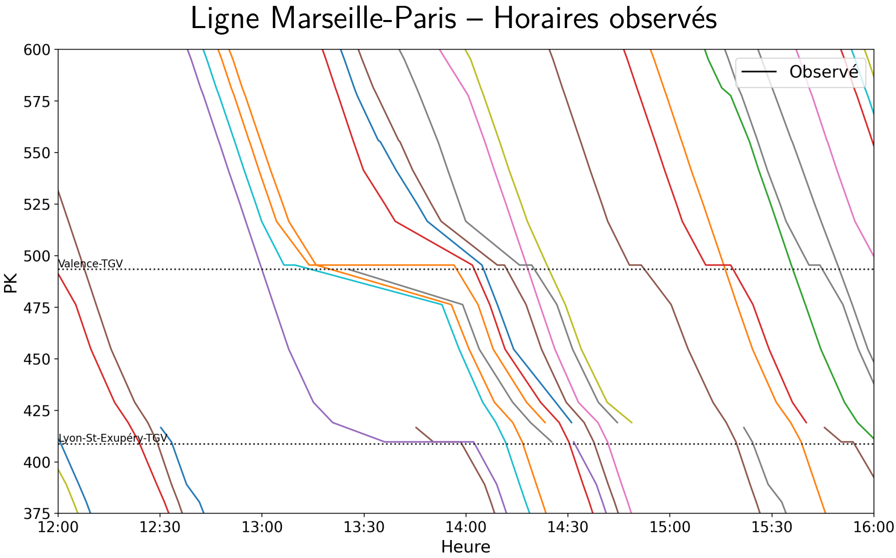
Baseline at SNCF (*translation*):

$$\boxed{\textbf{Future delay} = \textbf{current delay}}$$

Example


Ligne Marseille-Paris – Horaires théoriques
Théorique
Valence-TGV
Lyon-St-Exupery-TGV
PK
600
575
550
525
500
475
450
425
400
375
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Heure

Example

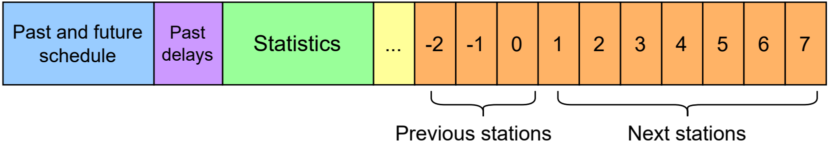

Ligne Marseille-Paris – Horaires observés
Observé
Valence-TGV
Lyon-St-Exupéry-TGV
PK
600
575
550
525
500
475
450
425
400
375
12:00
12:30
13:00
13:30
14:00
14:30
15:00
15:30
16:00
Heure

Model Idea

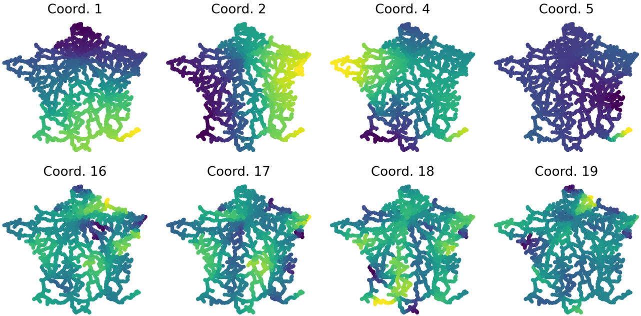
1. Vectorize train features: current delay, schedule, position...
2. Capture train interactions with a dedicated architecture
   $\implies$ self-attention, graph neural networks
3. Process the output to impose constraints

Input vectors

Represent each train i as a vector T_i:

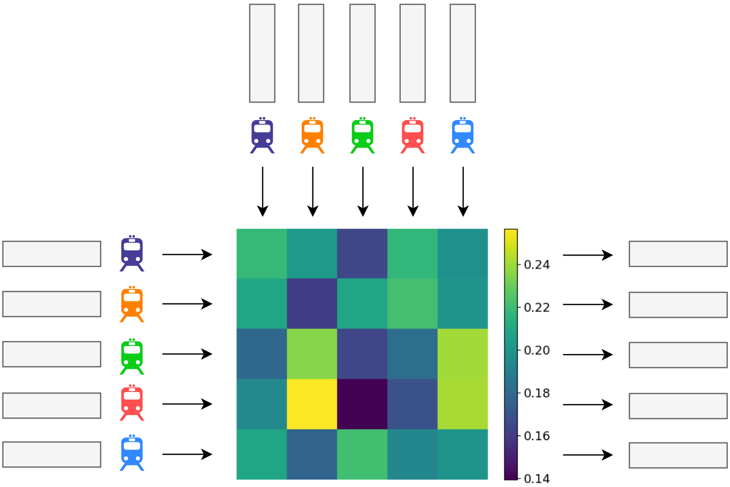


Node embedding: Train station $\mapsto$ vector

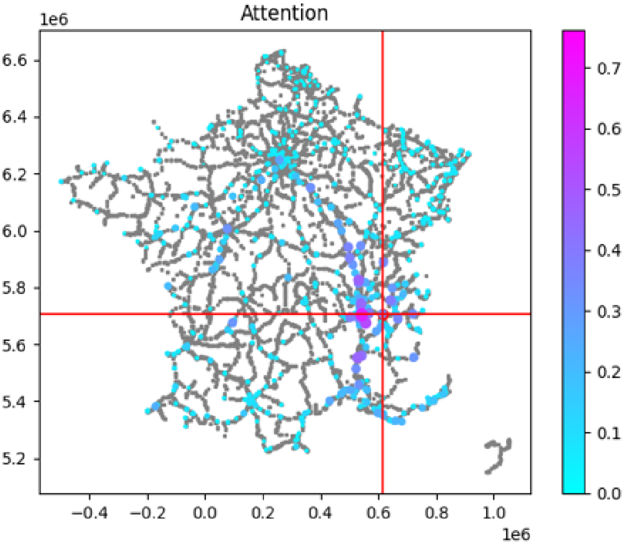


Attention-based Modeling

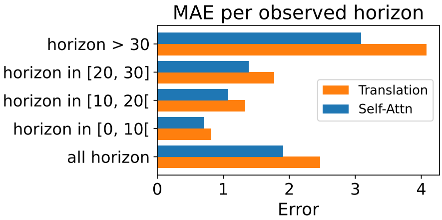
Feed $(T_1, \dots, T_n)$ into self-attention layers:

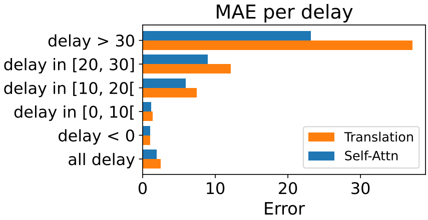


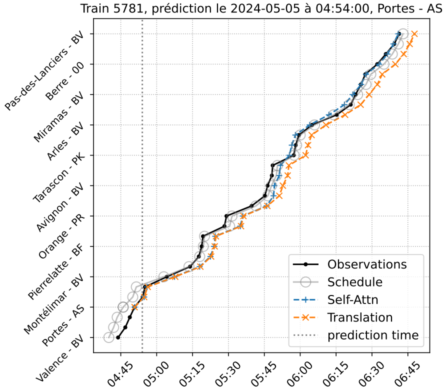
Self-attention between trains



Some results for regression







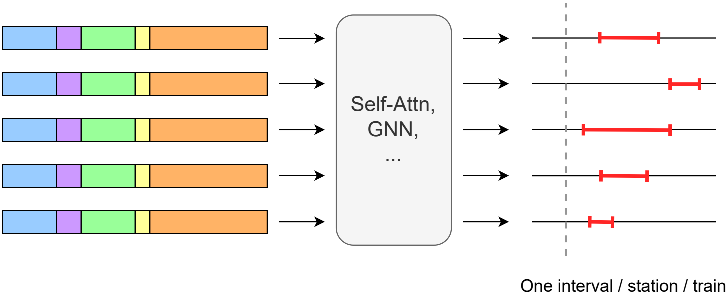
Quantile Regression

Idea: given samples (x_i, y_i), learn a quantile $q_\alpha(x)$ of $(Y \mid X = x)$.

Loss for quantile level α:

$$\hat{L}_\alpha(f) = \sum_{i=1}^{n} \alpha \, \mathbf{1}_{y_i > f(x_i)}(y_i - f(x_i)) + (1-\alpha) \, \mathbf{1}_{y_i \leq f(x_i)}(f(x_i) - y_i)$$

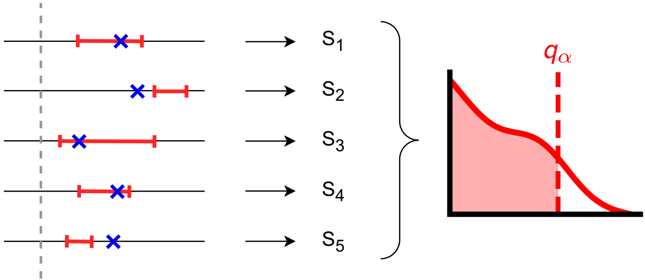
Here: learn conditional quantiles of $((D_{ij}) \mid T)$.



Conformalized Quantile Regression

Define a **nonconformity score**: $S(D_{true}, [a, b]) = \text{dist}(D_{true}, [a, b])$

Compute its quantile over a calibration set:



CQR prediction: $[a, b] \to \{D \mid S(D, [a, b]) < q_\alpha\} = [a - q_\alpha, b + q_\alpha]$

Caveat: work in log domain for better exchangeability

A Matrix Gaussian Model

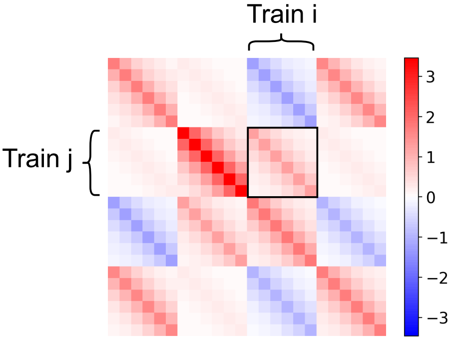
Idea: $(D_{ij}) \sim \mathcal{N}(\mu_\theta(T), \Sigma_\theta(T))$

- $\mu_\theta(T) : (n \times d)$
- $\Sigma_\theta(T) : (n \times d)^2$

$$\Sigma_\theta(T) = A_\theta(T) \otimes B$$

We got:

- $A_\theta(T) : (n \times n)$
- $B : (d \times d)$



To ensure positivity:

$$\begin{cases} A_\theta(T) = U_\theta(T)U_\theta(T)^\top + \mathrm{Diag}(\sigma_\theta^2(T)) \\ B = VV^\top \end{cases}$$

- $U_\theta(T) : (n \times p)$
- $\sigma_\theta(T) : (n)$
- $V : (d \times d)$

Then, learn (θ, V) with MLE.

Perspectives

- Self-attention architectures and GNNs can be used for delay prediction, and do better than translation
- UQ makes the problem more difficult, but standard methods can still be applied
- Future/ongoing directions
  - Joint modeling of spatiotemporal dependencies
  - Asymmetric distributions
  - Simulation-based approaches (ongoing PhD thesis)
  - Account for incidents
  - Predictions coherence
  - ...

Thank you!

Any questions?

Bibliography

- Arthaud, Farid, Guillaume Lecoeur, et Alban Pierre. « Transformers à Grande Vitesse: Massively parallel real-time predictions of train delay propagation ». Journal of Rail Transport Planning & Management 29 (1 mars 2024): 100418.
- Tiong, Kah Yong, Zhenliang Ma, et Carl-William Palmqvist. « A review of data-driven approaches to predict train delays ». Transportation Research Part C: Emerging Technologies 148 (1 mars 2023): 104027.
- Romano, Yaniv, Evan Patterson, et Emmanuel Candes. « Conformalized Quantile Regression ». In Advances in Neural Information Processing Systems, Vol. 32. Curran Associates, Inc., 2019.