

---

---

# Deep Learning Applied to Anomaly Detection in Meteorological Time Series

— Vedúci práce: MSc. Philipp Roberto Miotti —

Študent: Júlia Lichmanová

---

---

# Motivácia

- Význam kontroly kvality (QC) v doménach riadených údajmi.
- Účel QC pri zabezpečovaní štandardu dát.
- Anomálie v dátach sú spôsobené chybným meraním meracích prístrojov.
- Súčasné techniky:
  - štatistické metódy
  - fyzikálne modely
  - klasické metódy strojového učenia
  - manuálna kontrola
  - porovnanie so simulovanými dátami
- Techniky hlbokého učenia aplikované na multidimenzionálne meteorologické dáta neboli v nedávnej literatúre preskúmané.

# Ciel' práce

- Demonštrovať použiteľnosť modelov hlbokého učenia pri detekcii anomálií v časových radoch meteorologických dát.
- Použité modely:
  - Anomaly Transformer
  - TranAD

# Použité metriky

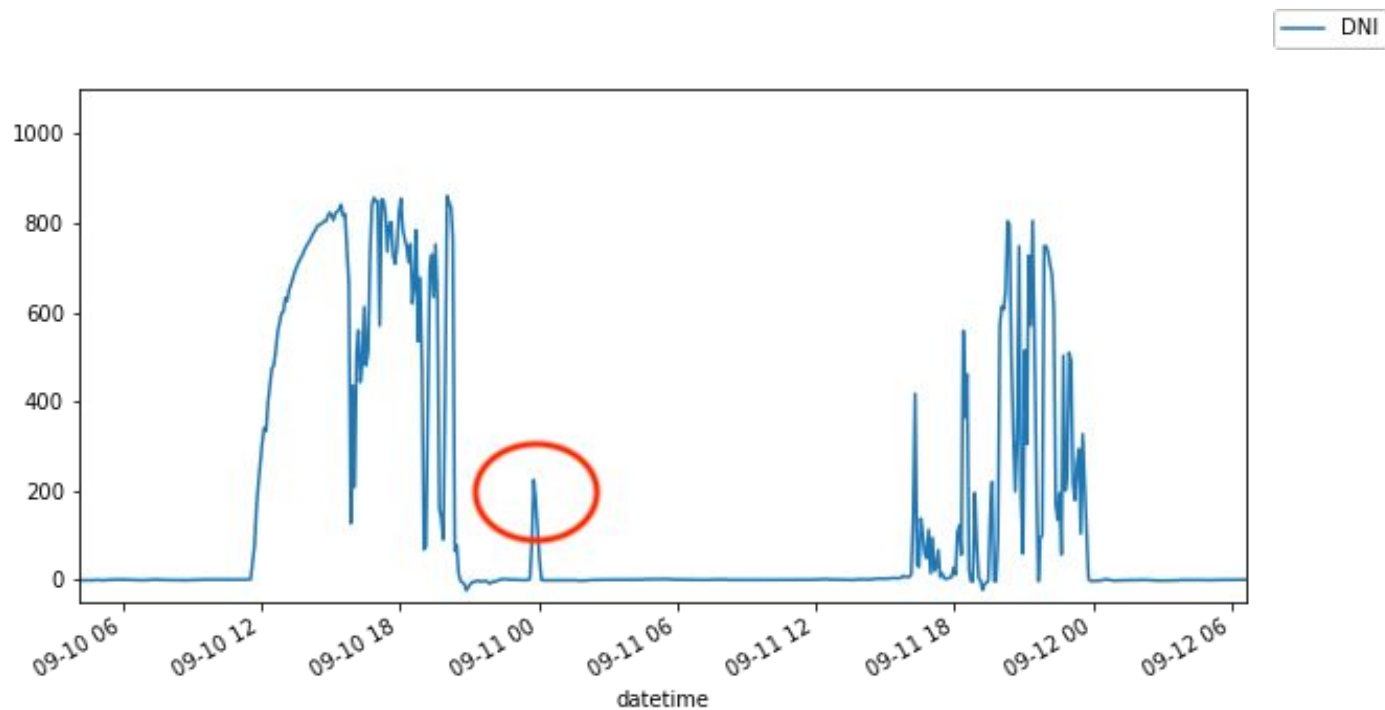
- Precision (True Positive / True Positive + False Positive)
- Recall (True Positive / True Positive + False Negative)
- F1 score (harmonický priemer medzi Precision a Recall)

		Actual	
		Positive	Negative
Predicted	Positive	<b>True Positive</b>	<b>False Positive</b>
	Negative	<b>False Negative</b>	<b>True Negative</b>

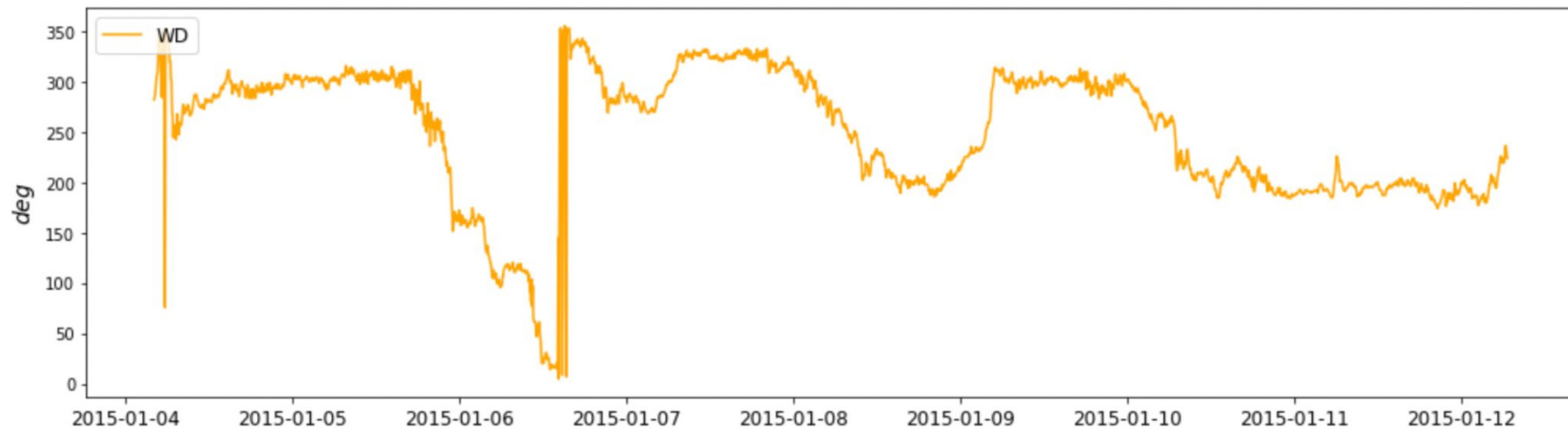
# Meteorologické dáta

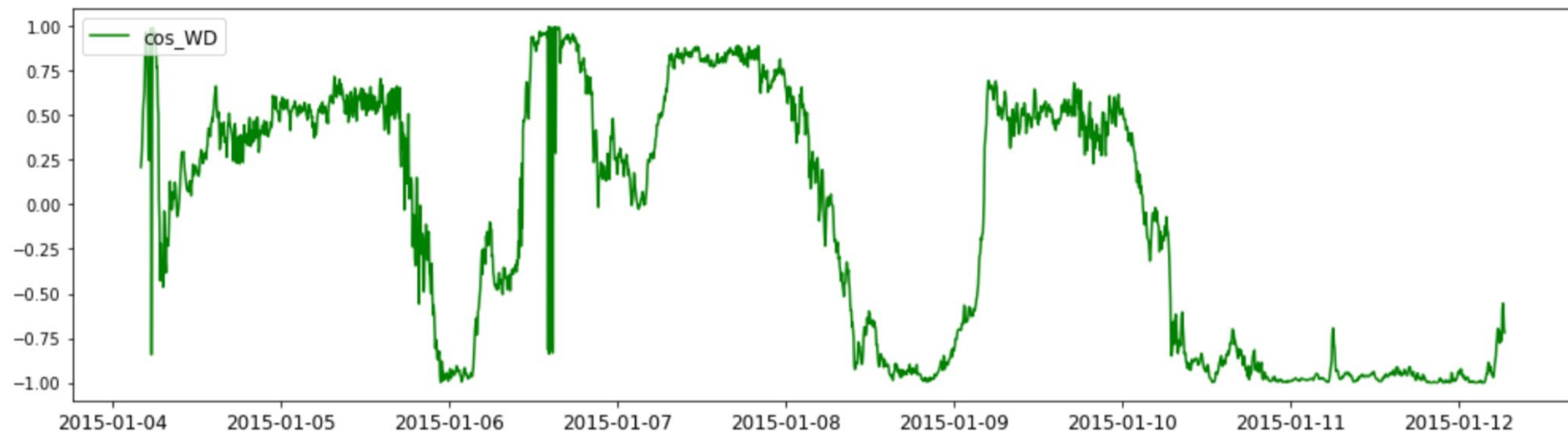
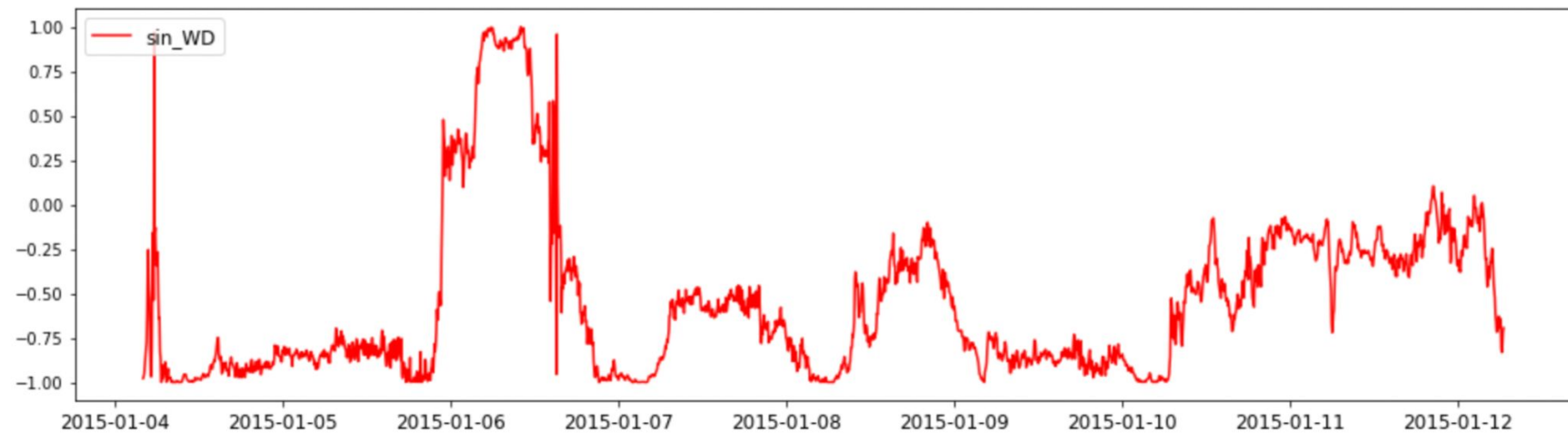
- Merané vo forme časových radov v intervale 5 minút.
- Merané veličiny:
  - globálne horizontálne žiarenie (GHI)
  - priame normálne žiarenie (DNI)
  - teplota (TEMP)
  - rýchlosť vetra (WS)
  - smer vetra (WD)
  - relatívna vlhkosť vzduchu (RH)
  - atmosférický tlak (AP)
- Merané 12 meteorologickými stanicami z USA, južná Afrika, Vietnam.

# Anomálie



# Transformácia smeru vetra

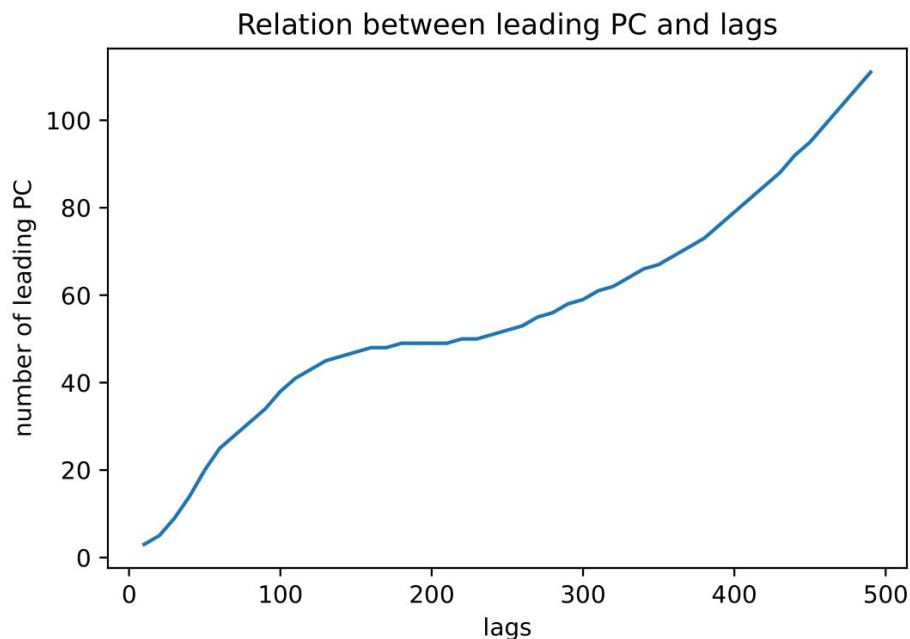






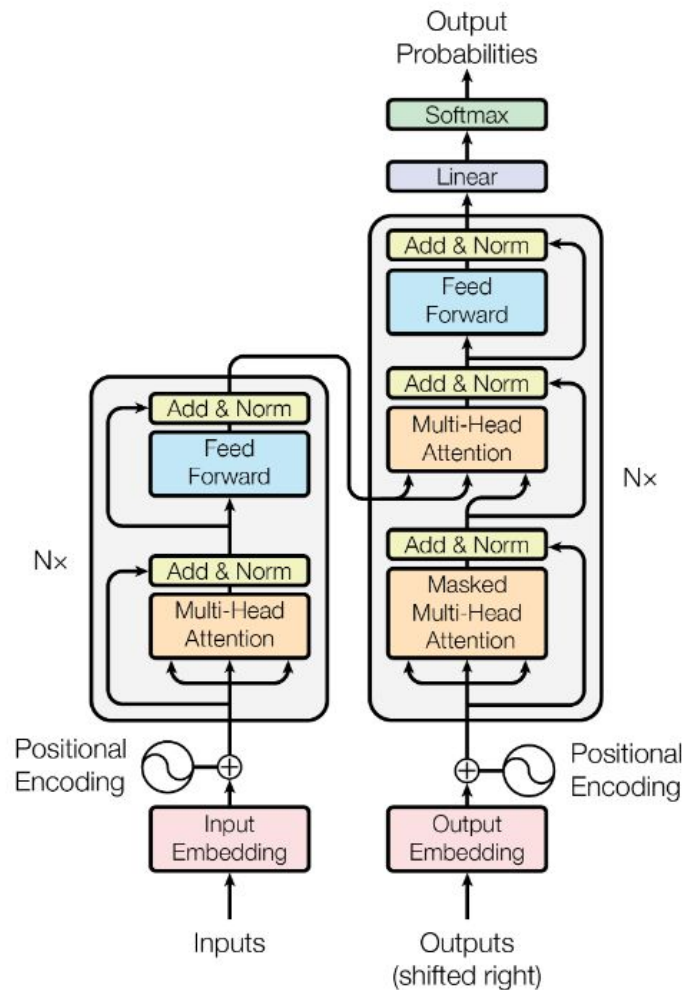
# Hľadanie veľkosti okna

- Veľkosť okna má vplyv na výkonnosť modelu, tréningový čas a pamäť.
- Analýza korelácie dát v čase pomocou techniky hlavných komponentov (PCA) za účelom maximalizovať informačný obsah v čo najmenšom časovom okne.
- Použité dáta sú v 60 minútovej samplovacej frekvencii.



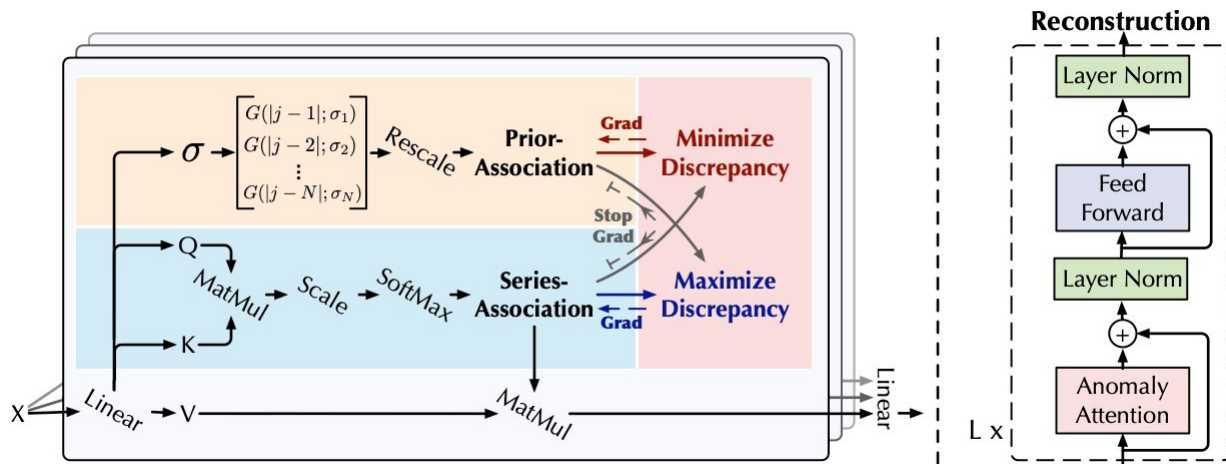
# Transformer architektúra

- Za účelom spracovania prirodzeného jazyka bola vyvinutá Transformer architektúra.
- RNN vs Transformer
- Najdôležitejší komponent je Mechanizmus pozornosti.
- Využitie Transformer architektúry pri časových radoch.



# Anomaly transformer

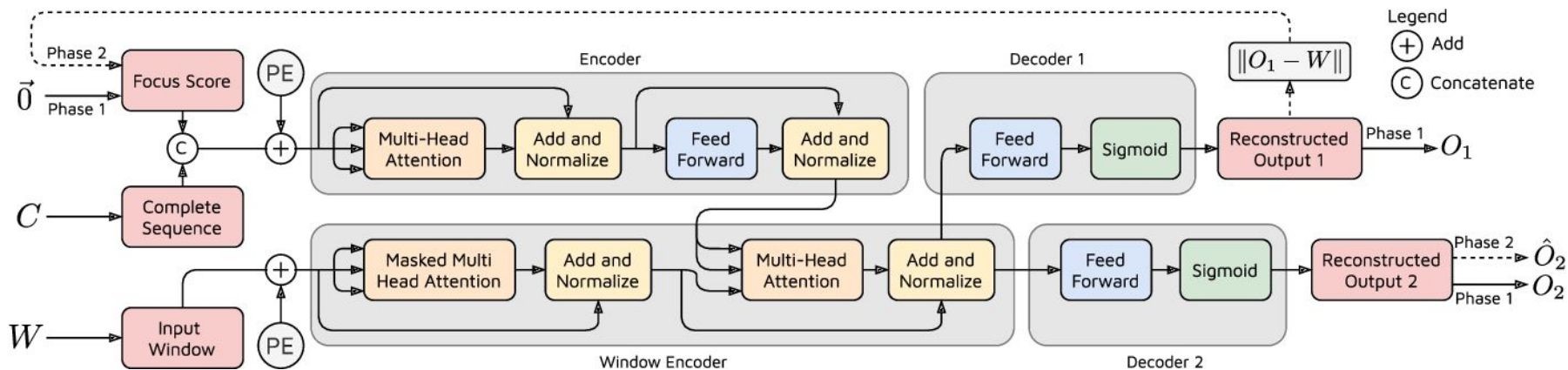
- Rekonštruje vstup na detekciu anomálií.
- Kvantifikuje mieru výchyľky podľa schopnosti tvoriť asociácie medzi bodmi.
- Identifikuje anomálie iba v čase a nie v jednotlivých dimenziách vstupných dát.
- Trénovanie prebiehalo na GPU.



<https://arxiv.org/abs/2110.02642>

# TranAD

- Schopnosť identifikovať anomálie v čase ale aj v priestore veličín časového radu.
- TranAD je predtrénovaný model, tréning prebiehalo na CPU.



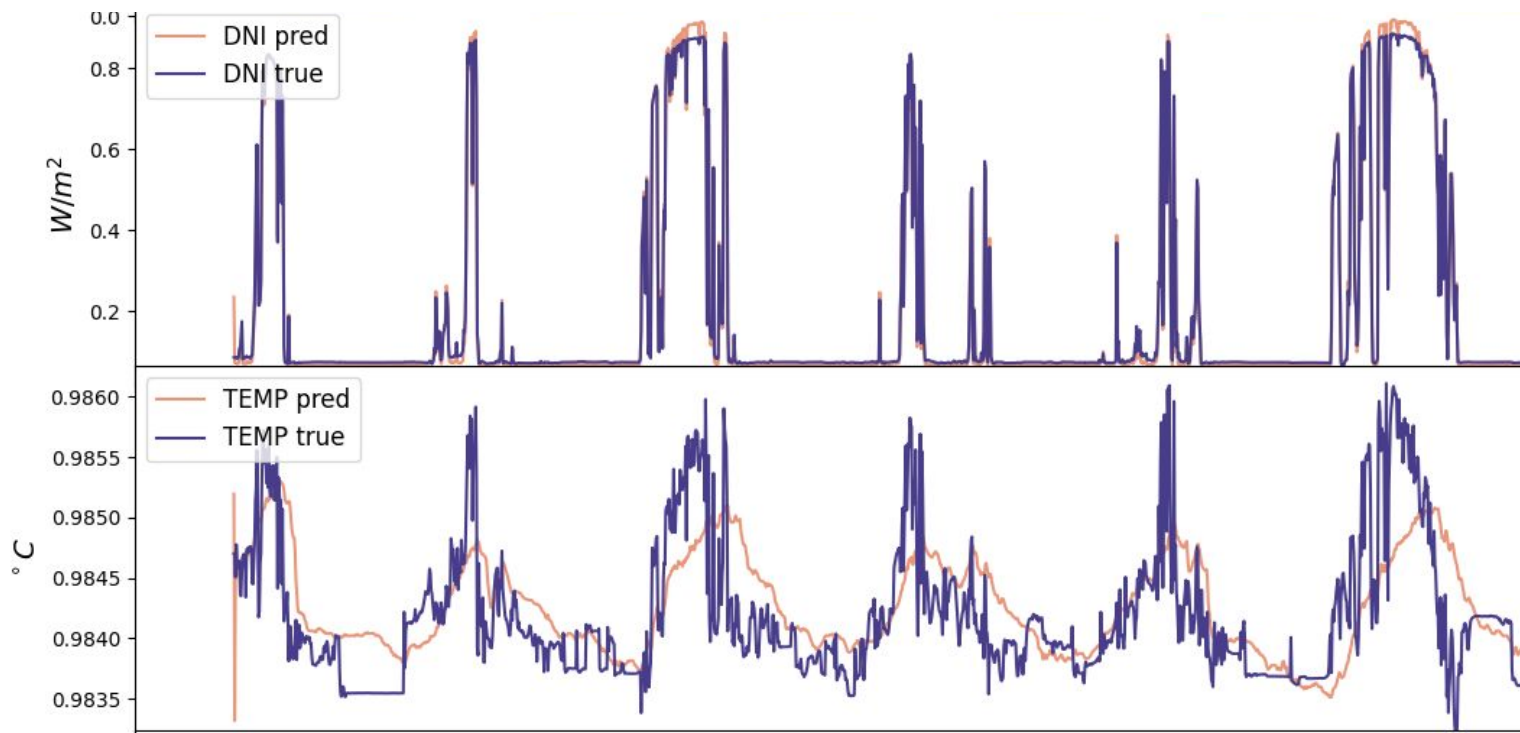
<https://arxiv.org/abs/2201.07284>

# Komplikácie a pozorovania

- Veľkosť okna identifikovaná pomocou PCA bola príliš veľká pre použitú pamäťovú kapacitu.
- Implementácia TranAD modelu neposkytovala parametrizáciu veľkosti okna.
- Trénovací čas bol v priemere pre TranAD 1 deň a pre Anomaly Transformer 3 dni.
- Nastaviť správny pomer medzi learning rate faktorom, trénovacou dobou a výkonnosťou modelu.

# Vysledky experimentu

Model	Precision	Recall	F1-score
Anomaly Transformer	0.7416	0.3567	0.4817
TranAD	0.1392	0.9615	0.2432



# Zhrnutie

- Podľa doposiaľ dosiahnutých výsledkov nedokážeme potvrdiť ani vyvrátiť hypotézu práce.
- Výsledok z analýzy veľkosti okna pomocou PCA metódy sa kvôli obmedzeným zdrojom nedal použiť.
- Možnosť budúcich prác:
  - Trénovať model len na iradiačných parametroch.
  - Vyskúšať jednotlivé modely trénovať pre jednu meteorologickú stanicu.
  - Vyskúšať väčší počet kombinácií hyperparametrov.

**Ďakujem za pozornosť**