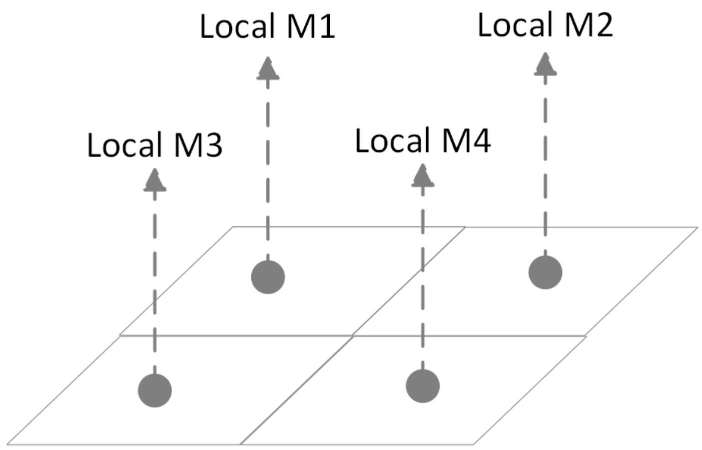
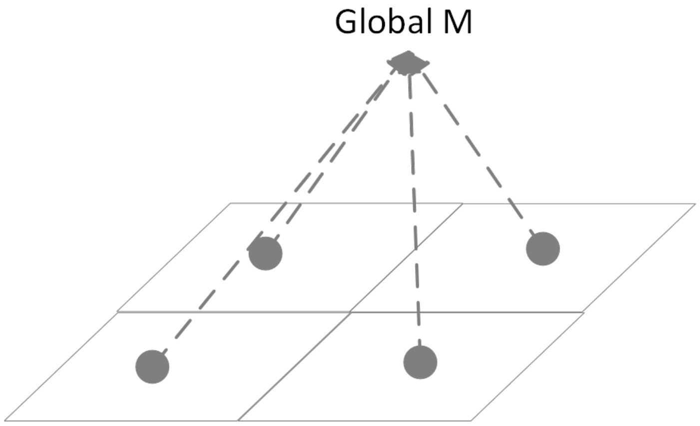
Decomposition learning based on spatial heterogeneity: A case study of COVID-19 infection forecasting in Germany

Ximeng Cheng, Jost Arndt, Emilia Marquez, and Jackie Ma
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Berlin, Germany
ximeng.cheng@hhi.fraunhofer.de

For each spatial unit, train a local model



Train a global model for the entire study area

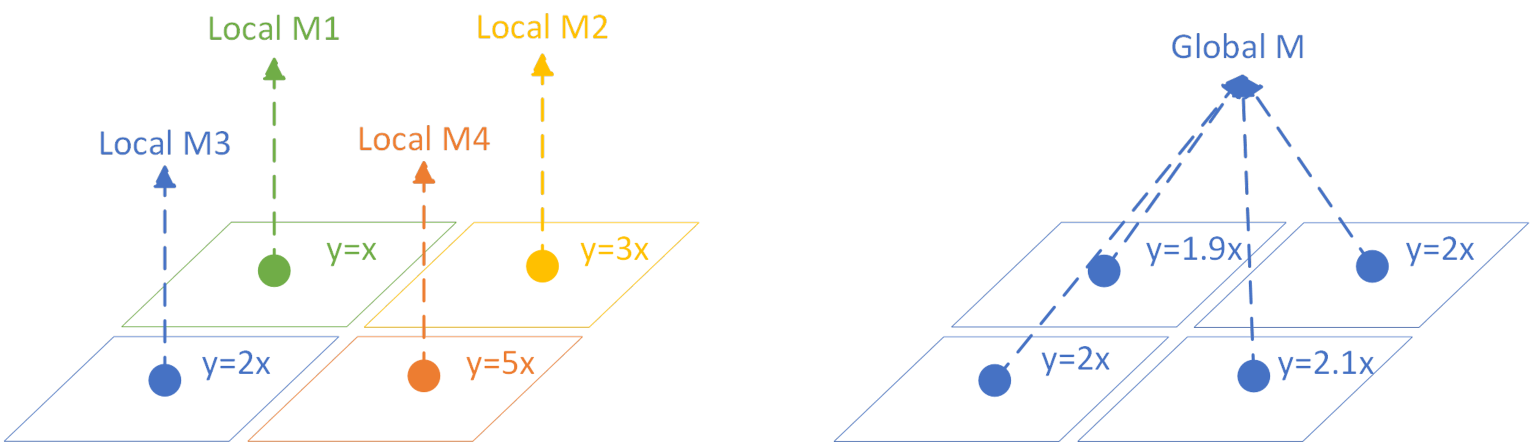


Multiple local models VS Single global model, which one is better?

Decomposition learning based on spatial heterogeneity: A case study of COVID-19 infection forecasting in Germany

Ximeng Cheng, Jost Arndt, Emilia Marquez, and Jackie Ma
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Berlin, Germany
ximeng.cheng@hhi.fraunhofer.de

Given x as model inputs and y as model outputs, y=ax+b as the linear model



Multiple local models VS Single global model, which one is better?
Depends on the degree of spatial heterogeneity and other training factors (e.g., data volume and parameter tuning)

Decomposition learning based on spatial heterogeneity: A case study of COVID-19 infection forecasting in Germany

Ximeng Cheng, Jost Arndt, Emilia Marquez, and Jackie Ma
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Berlin, Germany
ximeng.cheng@hhi.fraunhofer.de

	Advantage	Disadvantage
Local model	a. Ideally, the model performance is better; b. Take less time to train models, especially using parallel/distributed computing	a. Take much time to tune model parameters, especially for deep learning; b. Fewer data used for training; c. Difficult to achieve the best performance
Global model	a. More data used for training; b. Take less time to tune model parameters	a. Take more time/memory to train models

In our case, for ML methods (i.e., XGBoost), local- is better than global- models;
For DL methods (i.e., LSTM), local- is worse than global- models;
Whether to perform decomposition learning (from global to local) depends on the actual case

Background

- AI methods (machine learning/deep learning) have been applied in various application areas of geoscience based on spatio-temporal data
- In geoscience, spatial heterogeneity is a unique intrinsic feature
- Existing studies of decomposition learning
  - Spatial dimension: deep network transformation and moderation framework for data with spatial heterogeneity (Xie et al., 2021)
  - Temporal dimension: deep spatio-temporal residual networks for flow prediction (Zhang et al., 2017)
- This study takes the COVID-19 infection case forecasting in Germany as an example, from the spatial dimension, to compare a single global model and multiple local models

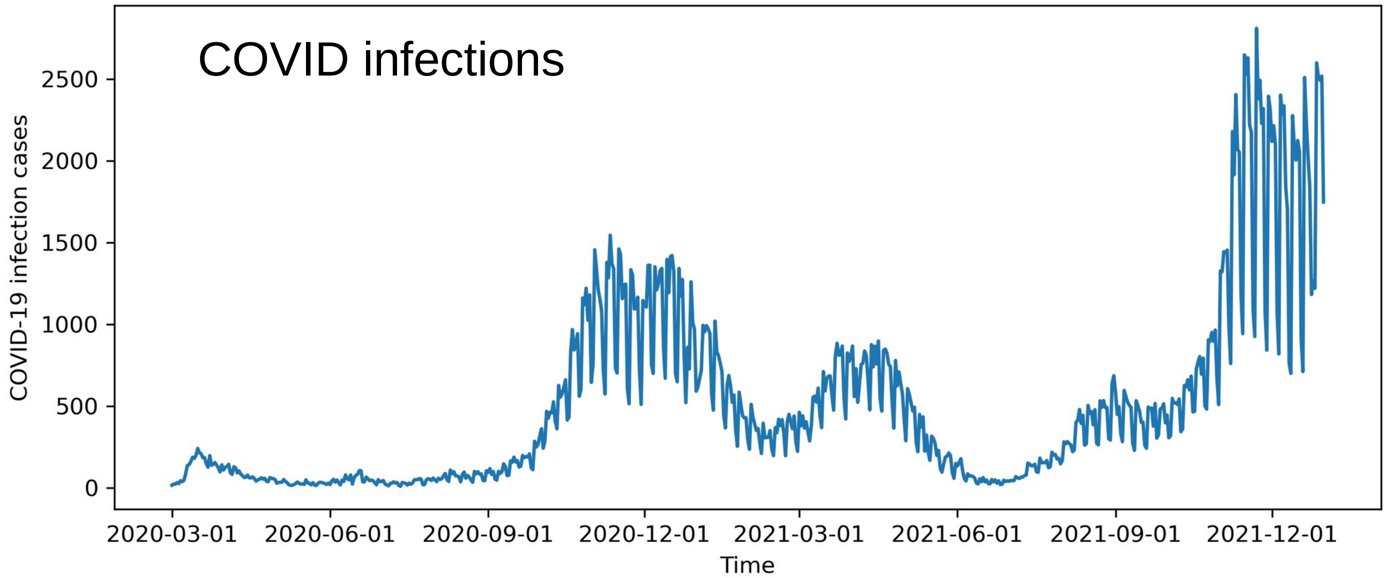
Xie Y, He E, Jia X, et al. A statistically-guided deep network transformation and moderation framework for data with spatial heterogeneity[C]//2021 IEEE International Conference on Data Mining (ICDM). IEEE, 2021: 767-776.
Zhang J, Zheng Y, Qi D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

Experiments

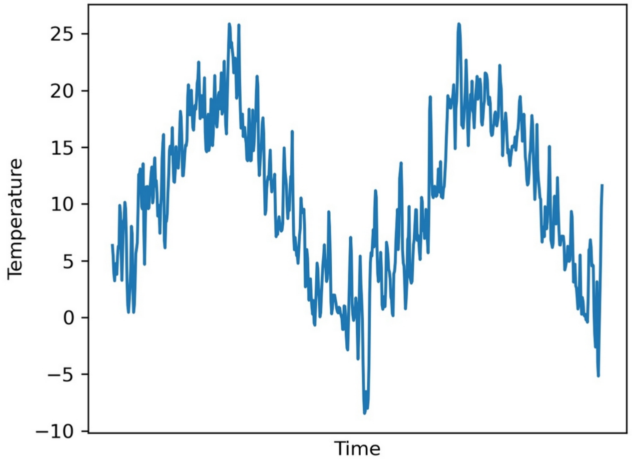
- Task: time-series forecasting, to forecast daily COVID-19 infection cases in Germany (NUTS 1 level, 16 states)
- Output (real value): German COVID-19 infection data from Robert Koch-Institute (RKI)
- Inputs: historical infection cases, humidity and temperature (from Deutscher Wetterdienst, DWD), traffic volumes (from HERE), contact index/mean (from NetCheck), and policy index (from Infas 360)
- Thanks to the data- and AI- supported early warning system (DAKI-FWS) project and to the partners for their support of research data

Experiments

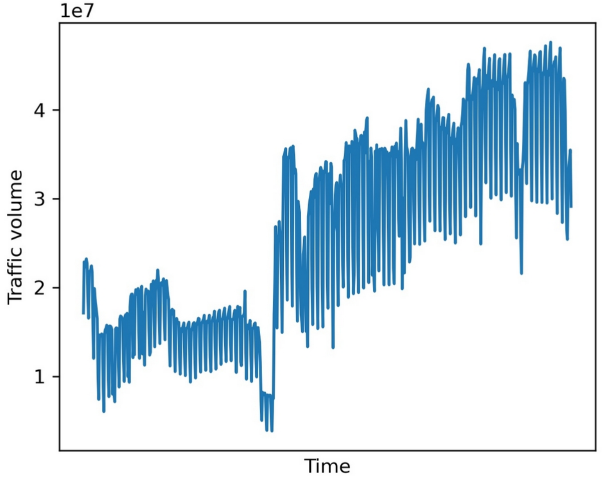
Data in Berlin (DE3)



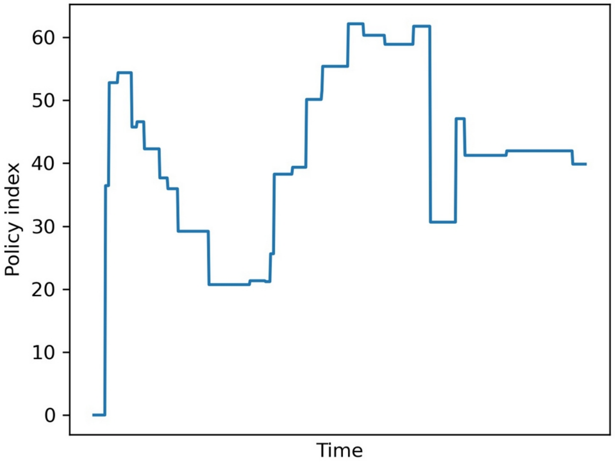
Temperature



Traffic volume



Policy index



Experiments

- Train and validation data: from 2020-03-01 to 2021-12-24
- Test data: from 2021-12-25 to 2021-12-31 (7 days)
- Spatial units: 16 states in Germany (NUTS 1 level)
- Time-series forecasting: previous 28 days as input, next 7 days as output
- Model: traditional time-series analysis method (ARIMAX/SARIMAX, only local model), machine learning method (XGBoost, local and global), deep learning method (LSTM, local and global)
- How to train local and global models?
  - Local model: for each state in Germany, train one specific model only based on the corresponding data in such state, 16 local models in total
  - Global model: for 16 states in Germany, train a single global model based on the entire data, ID of states is also an input feature

Results

RMSE

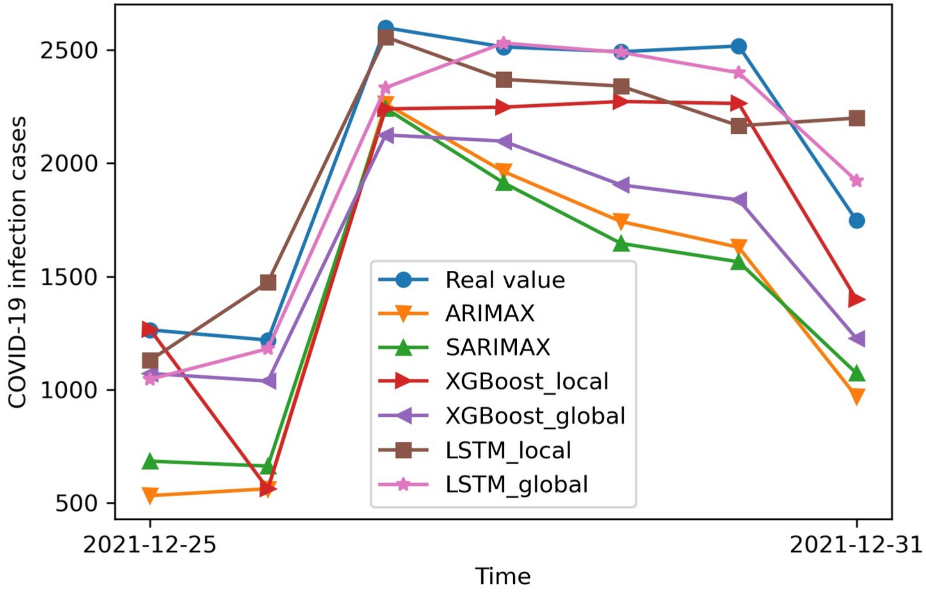
Region	ARIMAX	SARIMAX	XGB_L	XGB_G	LSTM_L	LSTM_G
DE1	1735.72	1755.20	**937.65**	975.41	1191.75	1012.51
DE2	1490.11	882.75	**348.66**	603.43	758.84	464.81
DE3	690.42	677.13	351.53	469.83	255.35	**153.10**
DE4	542.94	385.98	**191.04**	278.10	376.88	266.71
DE5	236.18	234.07	163.39	222.15	103.01	**92.01**
DE6	142.57	128.50	177.40	192.89	**120.83**	126.37
DE7	768.11	560.02	**345.22**	448.02	635.61	348.10
DE8	112.58	95.63	132.72	78.99	155.78	**73.38**
DE9	1040.18	907.81	402.14	679.58	567.82	**277.09**
DEA	2981.32	1908.87	830.05	958.24	1376.35	**691.57**
DEB	427.73	344.55	**130.22**	253.46	512.51	136.61
DEC	108.70	97.35	**23.45**	81.62	70.56	61.18
DED	1136.55	741.76	309.08	723.41	427.05	**274.60**
DEE	385.88	293.51	415.28	384.53	421.41	**247.26**
DEF	783.45	757.06	445.14	768.66	211.29	**160.79**
DEG	404.18	276.70	217.26	265.53	247.48	**170.88**
Average	811.66	627.93	338.77	461.49	464.53	**284.81**

MAPE

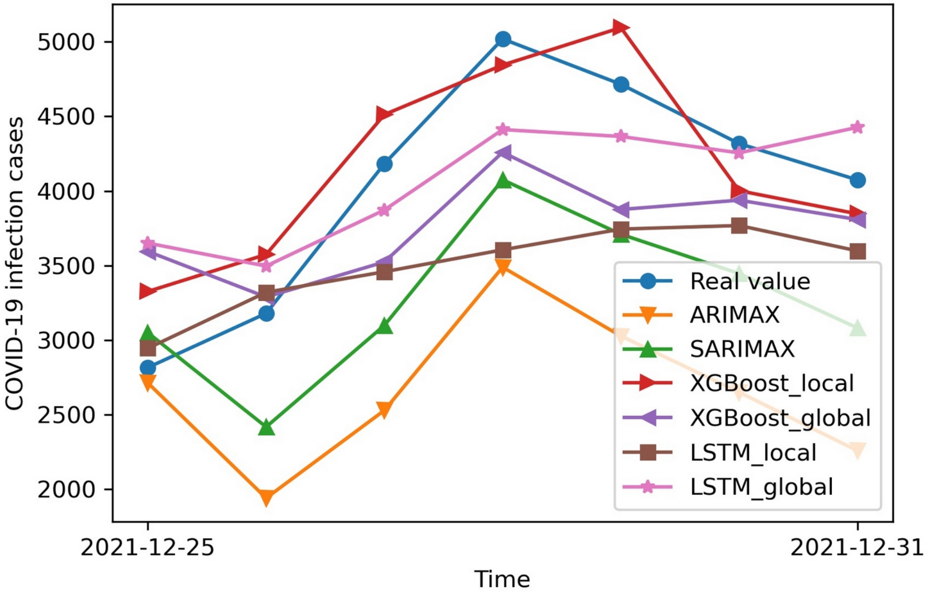
Region	ARIMAX	SARIMAX	XGB_L	XGB_G	LSTM_L	LSTM_G
DE1	0.4421	0.4497	**0.3229**	0.3529	0.3786	0.3460
DE2	0.3309	0.2040	**0.0898**	0.1359	0.1421	0.1094
DE3	0.3663	0.3420	0.1675	0.2075	0.1211	**0.0658**
DE4	0.3458	0.2349	**0.1419**	0.1868	0.2989	0.2090
DE5	0.3582	0.3650	0.4108	0.2708	0.2175	**0.1831**
DE6	0.0939	0.0938	0.1395	0.1305	0.0915	**0.0895**
DE7	0.3672	0.2752	**0.1646**	0.2126	0.2497	0.1863
DE8	0.1327	0.1142	0.1835	0.1068	0.2187	**0.0984**
DE9	0.3914	0.3438	0.1722	0.2773	0.2107	**0.1378**
DEA	0.4318	0.2731	0.1309	0.1537	0.1801	**0.1217**
DEB	0.3435	0.2473	**0.1260**	0.2562	0.4180	0.1436
DEC	0.3615	0.3232	**0.0731**	0.3211	0.2698	0.2362
DED	0.5347	0.3832	0.1680	0.4403	0.2550	**0.1220**
DEE	0.4125	0.3035	0.4586	0.4784	0.4558	**0.2421**
DEF	0.4567	0.4363	0.3562	0.4385	0.1596	**0.0871**
DEG	0.3333	0.2157	0.1705	0.2058	0.1643	**0.1150**
Average	0.3564	0.2878	0.2047	0.2609	0.2395	**0.1558**

Results

Berlin (DE3)



Bavaria (DE2)



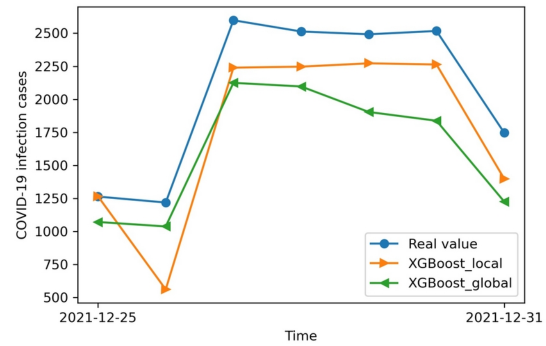
Methods have different forecasting patterns, e.g., ARIMAX underestimated the infection cases, and the forecasting time series of LSTM are relatively flat;
The model performance is different across regions;
Overall, the model performance LSTM > XGBoost > SARIMAX > ARIMAX

Comparison of local- and global- models

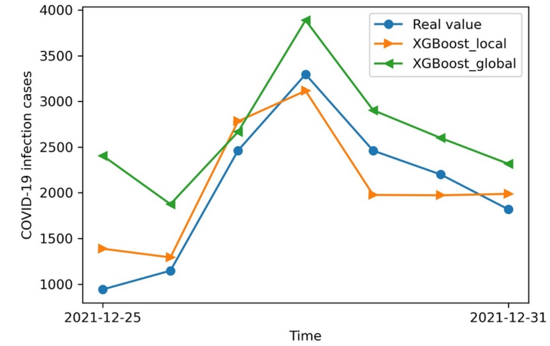
XGBoost

Region	Local RMSE	Global RMSE	Local MAPE	Global MAPE
DE1	**937.65**	975.41	**0.3229**	0.3529
DE2	**348.66**	603.43	**0.0898**	0.1359
DE3	**351.53**	469.83	**0.1675**	0.2075
DE4	**191.04**	278.10	**0.1419**	0.1868
DE5	**163.39**	222.15	0.4108	**0.2708**
DE6	**177.40**	192.89	0.1395	**0.1305**
DE7	**345.22**	448.02	**0.1646**	0.2126
DE8	132.72	**78.99**	0.1835	**0.1068**
DE9	**402.14**	679.58	**0.1722**	0.2773
DEA	**830.05**	958.24	**0.1309**	0.1537
DEB	**130.22**	253.46	**0.1260**	0.2562
DEC	**23.45**	81.62	**0.0731**	0.3211
DED	**309.08**	723.41	**0.1680**	0.4403
DEE	415.28	**384.53**	**0.4586**	0.4784
DEF	**445.14**	768.66	**0.3562**	0.4385
DEG	**217.26**	265.53	**0.1705**	0.2058
Average	**338.77**	461.49	**0.2047**	0.2609

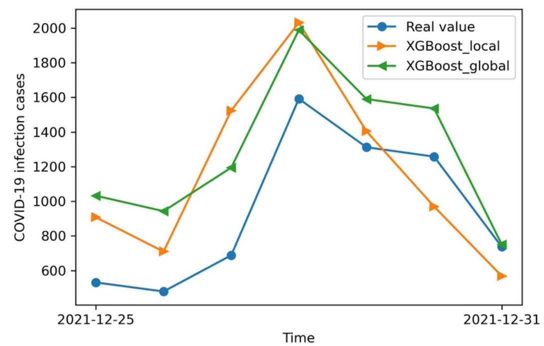
Berlin (DE3)



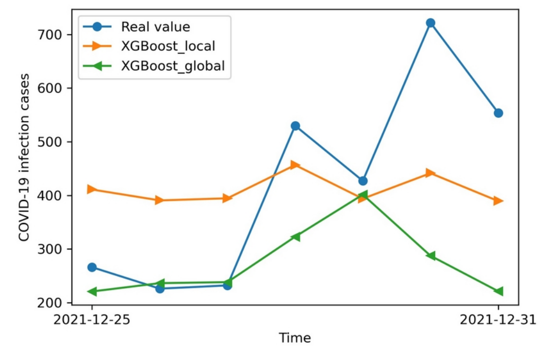
Saxony (DED)



Saxony-Anhalt (DEE)



Bremen (DE5)



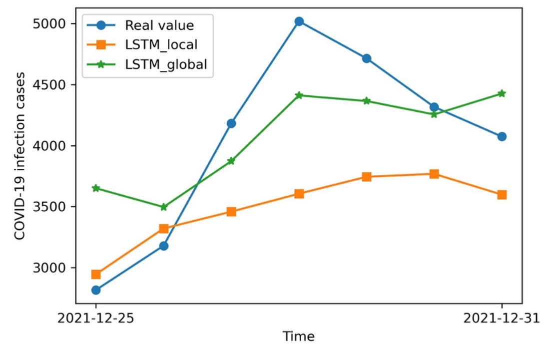
For most regions, the performance of the local model is better than the global model;
For XGBoost, it is easy to tune parameters to achieve the best performance of each local model

Comparison of local- and global- models

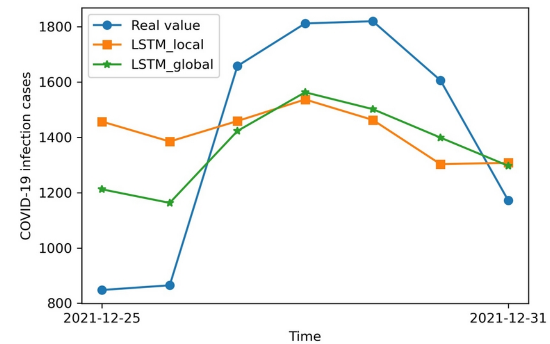
LSTM

Region	Local RMSE	Global RMSE	Local MAPE	Global MAPE
DE1	1191.75	**1012.51**	0.3786	**0.3460**
DE2	758.84	**464.81**	0.1421	**0.1094**
DE3	255.35	**153.10**	0.1211	**0.0658**
DE4	376.88	**266.71**	0.2989	**0.2090**
DE5	103.01	**92.01**	0.2175	**0.1831**
DE6	**120.83**	126.37	0.0915	**0.0895**
DE7	635.61	**348.10**	0.2497	**0.1863**
DE8	155.78	**73.38**	0.2187	**0.0984**
DE9	567.82	**277.09**	0.2107	**0.1378**
DEA	1376.35	**691.57**	0.1801	**0.1217**
DEB	512.51	**136.61**	0.4180	**0.1436**
DEC	70.56	**61.18**	0.2698	**0.2362**
DED	427.05	**274.60**	0.2550	**0.1220**
DEE	421.41	**247.26**	0.4558	**0.2421**
DEF	211.29	**160.79**	0.1596	**0.0871**
DEG	247.48	**170.88**	0.1643	**0.1150**
Average	464.53	**284.81**	0.2395	**0.1558**

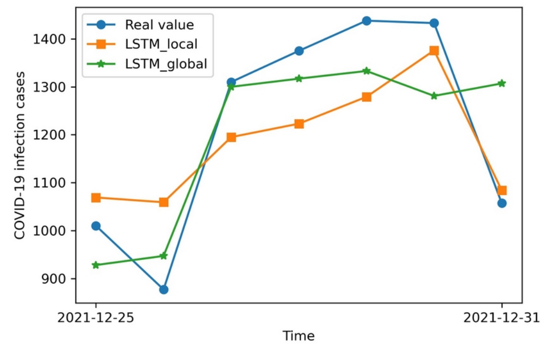
Bavaria (DE2)



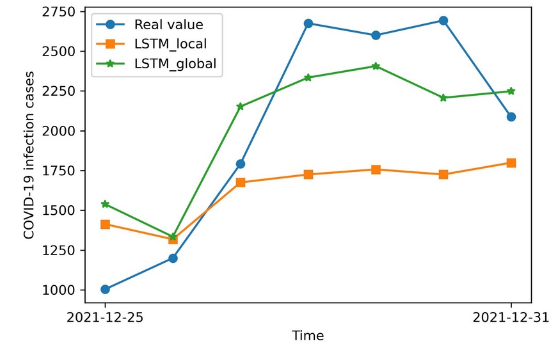
Brandenburg (DE4)



Hamburg (DE6)



Hesse (DE7)



For all of the regions, the performance of the local model is worse than the global model;
For LSTM, fewer train data makes it difficult to adequately train the local model;
For LSTM, a large number of parameters makes it difficult to achieve the best performance of the local model

Comparison of local- and global- models

Training time

Time	XGB_L	Sum_XGB_L	XGB_G	LSTM_L	Sum_LSTM_L	LSTM_G
Max	1.16s	4.15s	2.76s	2148.84s	34381.44s	33560.58s
Min	0.04s	3.18s	2.45s	1129.94s	18079.04s	23770.60s

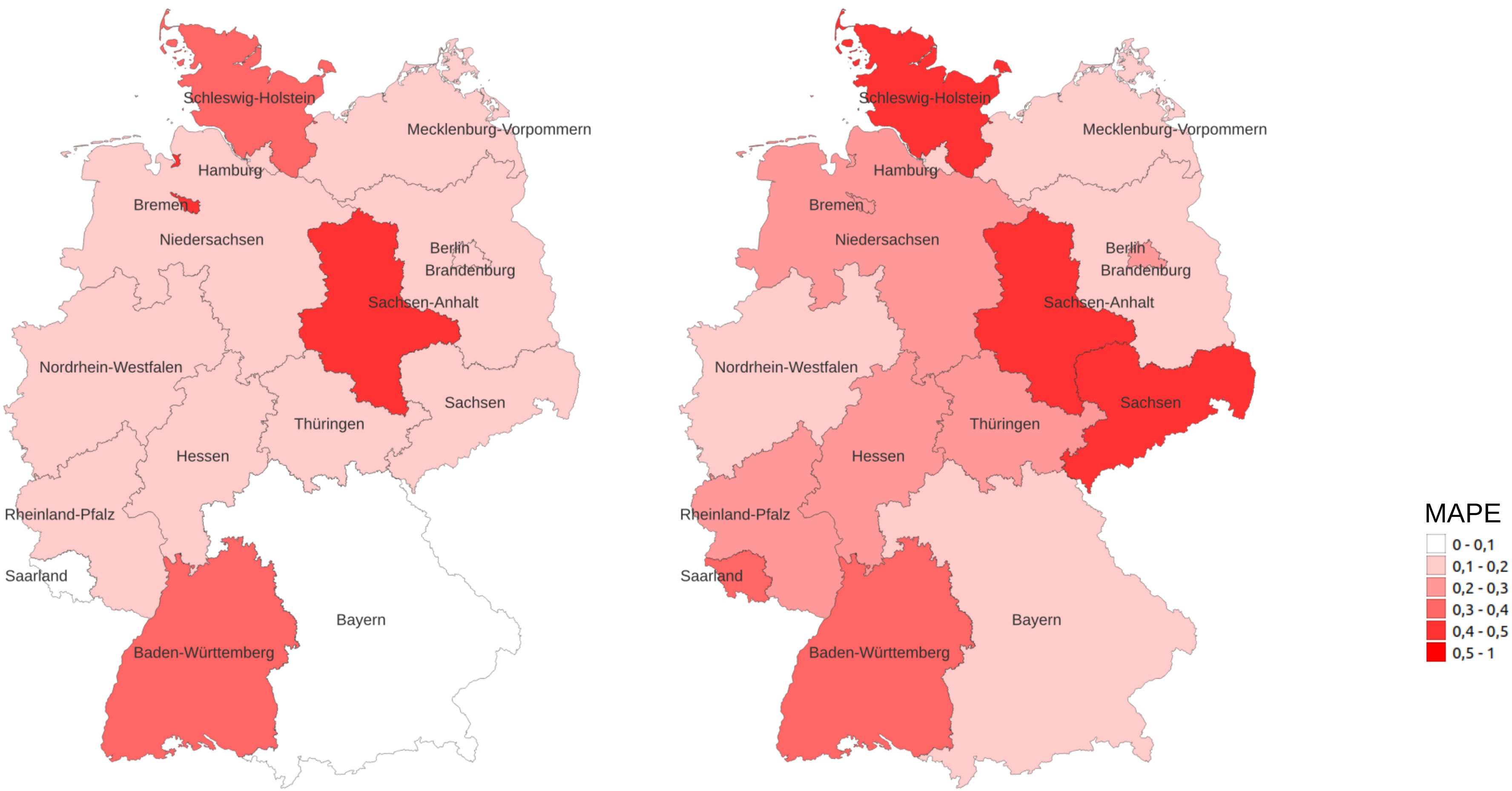
- Time of model training: single local model is shorter than the global model, and the time of sequentially training all local models is similar to the global model
- Time of tuning parameters: local model is much larger than the global model, especially for deep learning with many parameters
- Local model training is independent, techniques such as distributed or parallel computing can save training time

Comparison of local- and global- models

XGBoost

Local

Global



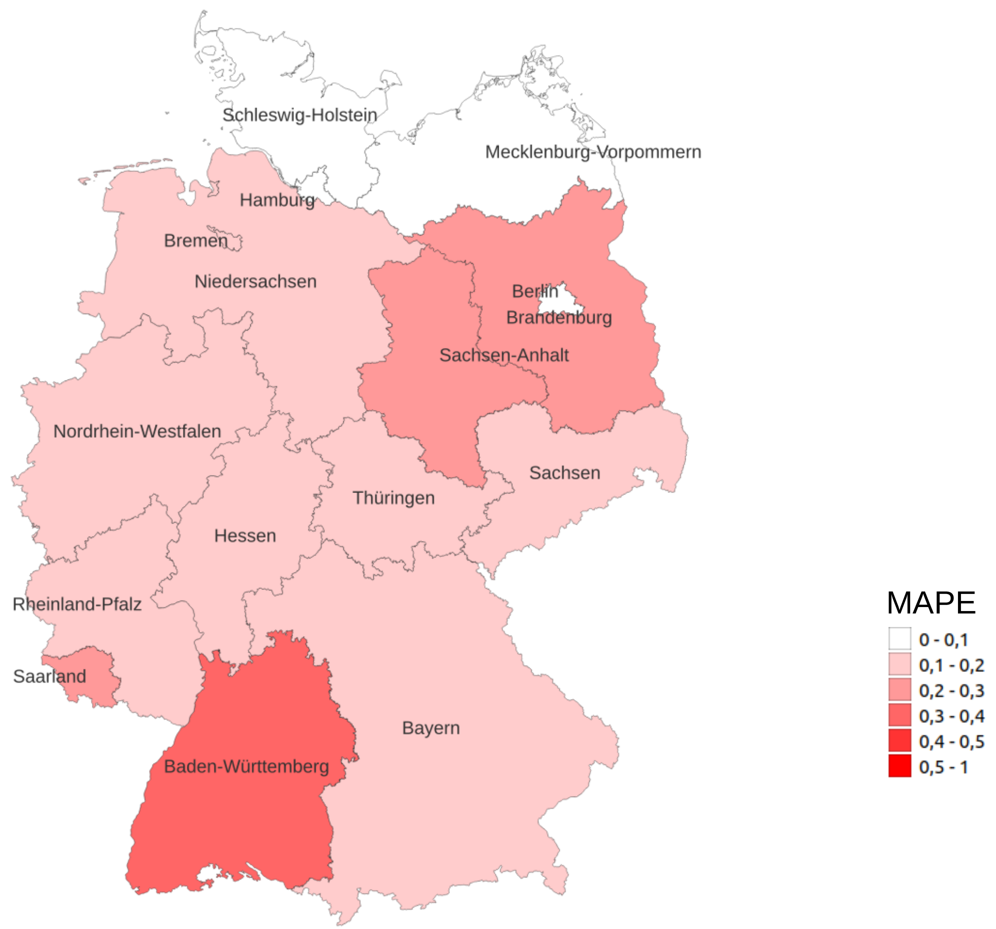
 Dr. Ximeng Cheng

Comparison of local- and global- models

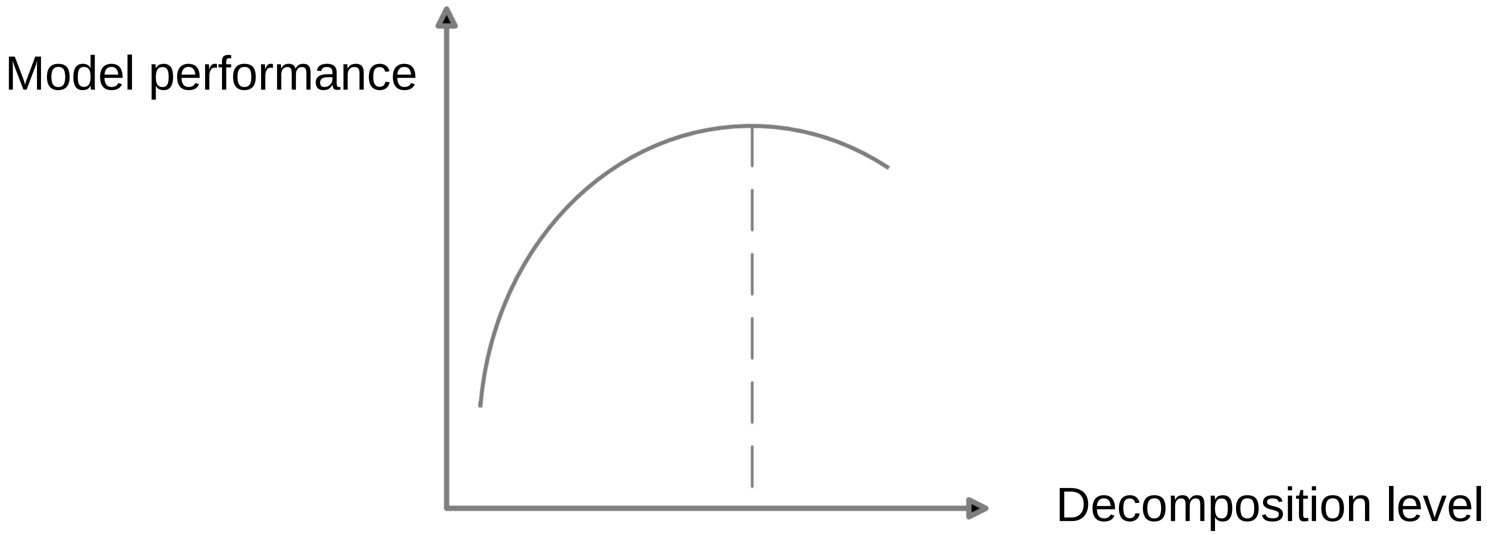
LSTM

Local

Global



Summary



- Ideally, the decomposed regions are more homogeneous, which leads the local model to achieve a better performance
- In reality, multiple local models use different parameters, making parameter tuning difficult. Besides, decomposed data is few which hinders model training
- No absolute winner between local- and global- models. In practical application, an appropriate decomposition level should be selected

Future

- Explicitly account for spatial heterogeneity, decompose study areas based on environmental attributes of each spatial unit
- Conduct more experiments
  - More ML/DL methods (e.g., LightGBM, GRU, and Transformer)
  - Finer spatial resolution (e.g., NUTS 3/county level in Germany)
  - New application scenarios beyond COVID-19 infection forecasting
- Optimizing the comparison of local- and global- models
  - Reduce or quantify the impact of different train data volumes
  - Use the same model parameters or optimize parameter tuning strategy

Dr. Ximeng Cheng
Applied Machine Learning Group
Artificial Intelligence Department
Fraunhofer HHI
ximeng.cheng@hhi.fraunhofer.de

HHI website:

Personal homepage: