Interpretability and Causal Inference for Global Time Series Forecasting Methods

Facebook Forecasting Summit 2021

November 16, 2021

Presented by

Christoph Bergmeir

Dept. of Data Science and Artificial Intelligence
Faculty of Information Technology
Monash University
Melbourne, Australia

Overview

Acknowledgements

The work I'm presenting today is joint work with my PhD students:

- Dilini Rajapaksha (Interpretability)
- Priscila Grecov (Causal Inference)

Local Vs Global Time Series Forecasting

Local Models	**Global Models**
Build one forecasting model per each series	Build one forecasting model across many series
Fit local parameters separately on each series	Fit global parameters that are the same across many series
Amount of parameters is small per series, but grows with amount of series	Amount of parameters stays the same
Models need to be simple, as not much data	Models can be more complex
Examples: SES, Theta, ETS, ARIMA, Prophet, TBATS	Examples: ES-RNN, DeepAR, NBEATS, ...

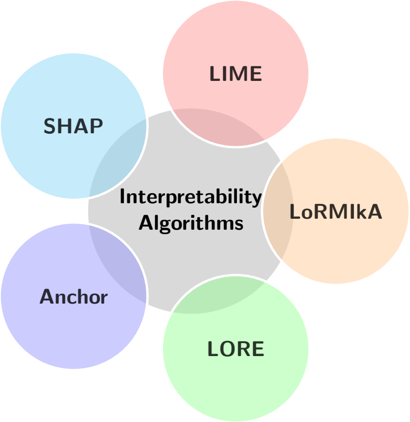
For a given time-series a global model is typically more complex and less interpretable than a local model

Accuracy vs Interpretability in Machine Learning

Accuracy ⇧ **Interpretability** ⇩

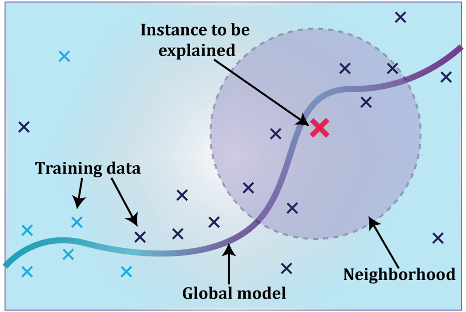
Trade-off between accuracy and interpretability

General Solution in Machine Learning: Post-hoc Local Model-Agnostic Interpretability



Local Interpretability For Classification

- We want an explanation for one particular instance
- Assumption: The behaviour of the instance to be explained is similar to the behaviour of the instances in its neighbourhood



- Select a neighbourhood
- Sample new instances from the neighbourhood (change the features)
- Run your global model on the sampled instances to get the labels
- Train an interpretable model on the neighbourhood plus generated labels, to mimic the global model in the neighbourhood of the instance

Our goal

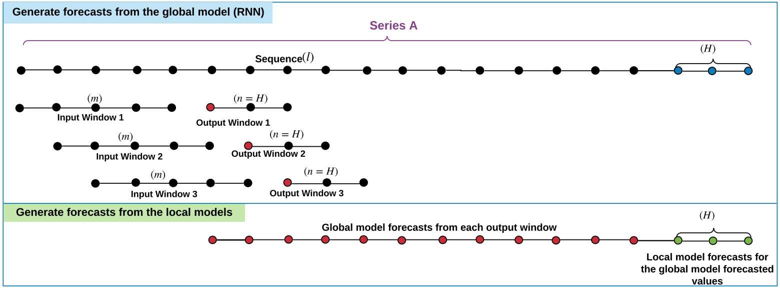
Transfer the concept of local interpretability as closely as possible to global time-series forecasting models:

- The global forecasting model is seen as a black-box model, the framework should be model-agnostic
- We want (local) explanations for one particular time series out of a big set of series
- Our neighbourhood is a particular series (could also use more series chosen with a similarity like DTW)
- Problems:
  - How to sample from the neighbourhood?
  - How to run the global model on the samples?
  - Which local explainer to use?
  - How to train the local explainers on the output of the global model in the neighbourhood?

Proposed Solution

How to sample from the neighbourhood? / How to run the global model on the samples?

- No sampling, only actual instances from the neighbourhood: Use the fit of the global model to the particular time series.



Proposed Solution Cont.

- Using the in-sample fit of the global model without sampling may seem like a bad idea, especially if the global model overfits.
- **But: Finding of Montero-Manso and Hyndman (2021):**

  The in-sample error of the global model is usually higher than the local model while it generalizes better

Bootstrapping:

- Generate new time series using bootstrapping, by bootstrapping the remainder of STL decomposition from the original series using a Moving Block Bootstrap (MBB) technique
- Bootstrap the remainder of the fit of the time series using MBB

Proposed Solution Cont.

- We assume that local statistical methods are interpretable, though this can be disputed.
- The goal of our work is to make global models as interpretable as local models.

Which local explainer to use?

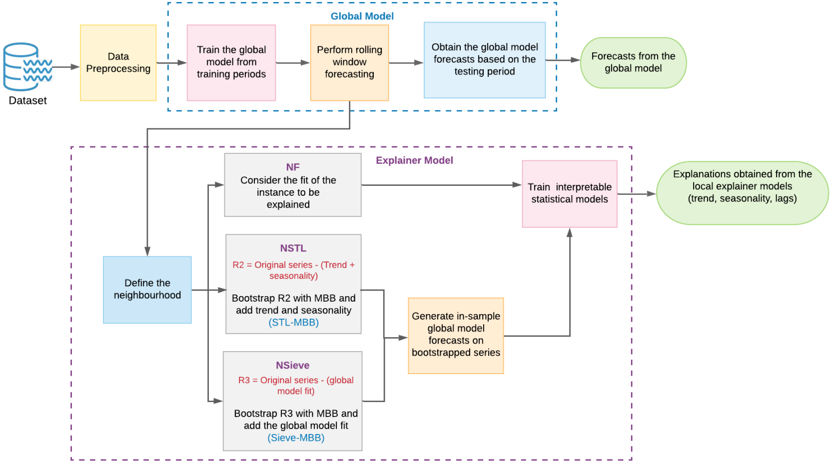
- ETS
- TBATS
- MSTL & STL
- Prophet

$= \text{Decompositions (trend, seasonality, remainder)}$

- DHR-Arima
- Theta

$= \text{Coefficents}$

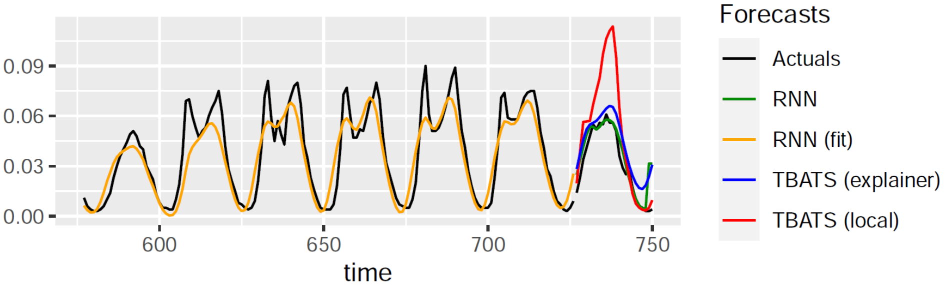
Proposed Methodology



Example 1

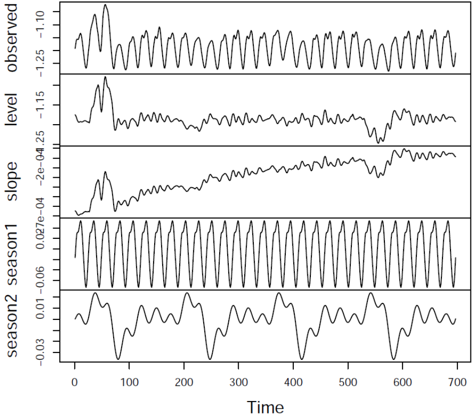
San Francisco Traffic Hourly Dataset

- **TBATS interpretation:** Forecasts plot



Example 1 Cont.

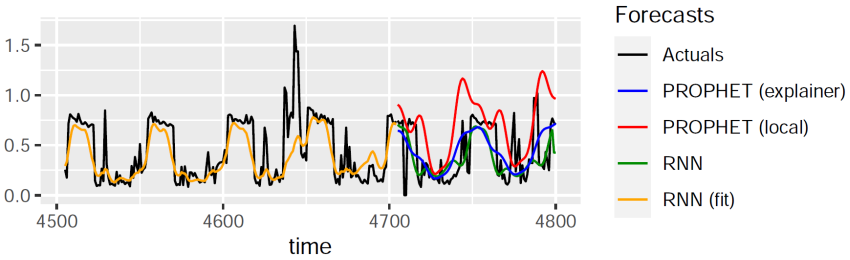
- **TBATS interpretation:** Decomposition plot



Example 2

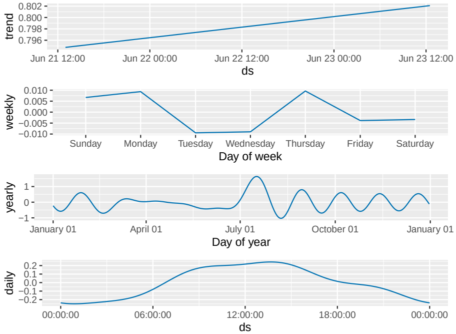
Ausgrid Half-hourly Dataset

- **Prophet interpretation:** Forecasts plot



Example 2 Cont.

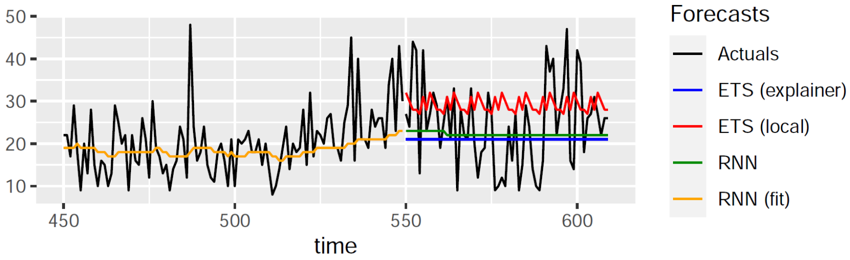
- **Prophet interpretation:** Decomposition plot



Example 3: Bootstrapped Models

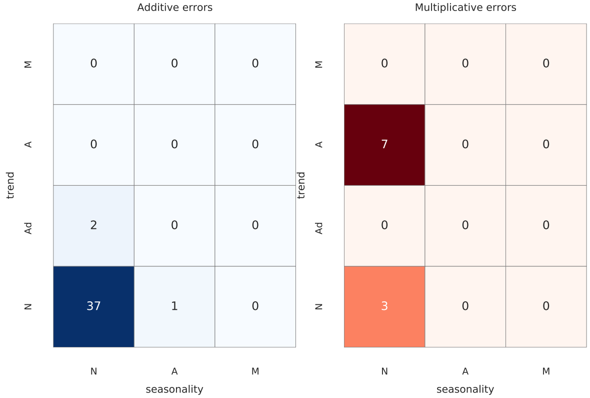
Kaggle Web Traffic Dataset

- **ETS interpretation:** Forecasts plot



Example 3 Cont.

- **ETS interpretation:** Chosen Model Summary Plot



How to Evaluate Local Explainers?

- **Comprehensibility:** How interpretable are the explanations?
- **Fidelity:** How well does the explanation approximate the forecast of the global model?
- **Consistency:** How consistent are the explanations of the explainer across multiple runs?
- **Accuracy:** How well does an explainer forecast unseen data?

Fidelity_Actual

Error can be sMAPE, MASE, RMSE, or MAE

Fidelity_Actual $=$ Error(global,explainer) - Error(global,actual)

- **Fidelity_Actual** < 0, the global model is closer to the explainer model than the actual.
- **Fidelity_Actual** > 0, the global model is closer to the actuals than the explainer model.

Fidelity_Local

Fidelity_Local= Error(global,explainer) - Error(global,local)

- Verifies whether the local explainer is closer to the global model forecasts than the local model to the global model forecasts.
- **Fidelity_Local** $<$ **0**, explainer model is closer to the global model.
- **Fidelity_Local** $>$ **0**, local model is closer to the global model.

Fidelity_with_Explainer

$$\textbf{Fidelity\_with\_Explainer} = \text{Error(global,explainer)} - \text{Error(local,explainer)}$$

- Verify whether the local explainer is closer to the global model than the local model.
- **Fidelity_with_Explainer** $< \mathbf{0}$, the explainer model is closer to the global model
- **Fidelity_with_Explainer** $> \mathbf{0}$, the explainer model is closer to the local model

Accuracy Explainer & Local model

Explainer & Local model = Error(actual,explainer) - Error(actual,local)

- Verify whether the local explainer performs better than the local model.
- **Explainer & Local model** $<$ **0**, the explainer model performs better than the local model
- **Explainer & Local model** $>$ **0**, local model performs better than the local explainer.

Accuracy Explainer & Global model

Explainer & Global model = Error(actual,explainer) - Error(actual,global)

- Verify whether the local explainer performs better than the global model.
- **Explainer & Global model** $< \mathbf{0}$, the explainer model performs better than the global model
- **Explainer & Global model** $> \mathbf{0}$, global model performs better than the local explainer

Accuracy Global and Local Model

Global and Local Model = Error(actual,global) - Error(actual,local)

- Verify whether the global model performs better than the local model.
- **Global and Local Model** $<$ **0**, the global model performs better than the local model
- **Global and Local Model** $>$ **0**, local model performs better than the global model

Experiments Results

Table 1: Mean SMAPE measures of the datasets for different local explainer models

Dataset	Local Explainer	Fidelity_Actual	Fidelity_Local	Explainer and Local Model	Explainer and Global Model	Fidelity_with Explainer	Global and Local Model
nn5_weekly	es	-0.082	-0.066	-0.008	0.006	-0.072	-0.013
	prophet	-0.050	-0.024	-0.000	0.003	-0.011	-0.004
	dhr_arima	-0.058	-0.021	-0.003	0.000	-0.011	-0.003
	tbats	-0.069	-0.036	-0.022	0.006	-0.022	-0.028
ausgrid_weekly	es	-0.074	-0.043	-0.024	0.019	-0.001	-0.043
	prophet	-0.063	-0.020	-0.015	0.018	0.067	-0.033
	dhr_arima	-0.082	-0.030	-0.015	0.009	0.038	-0.024
	tbats	-0.089	-0.022	-0.007	0.021	0.009	-0.028
	stl	-0.076	-0.039	-0.026	0.015	0.014	-0.041
kaggle_web_traffic_daily	ets	-0.401	-0.210	-0.074	0.003	-0.196	-0.077
	theta	-0.404	-0.164	-0.047	0.003	-0.151	-0.049
sf_traffic_hourly	tbats	-0.017	-0.020	-0.006	0.000	-0.018	-0.006
	prophet	-0.011	-0.016	-0.013	0.008	-0.009	-0.021
	dhr_arima	-0.013	-0.025	-0.014	0.000	-0.013	-0.014
	mstl	0.000	-0.015	-0.006	0.010	-0.008	-0.016
ausgrid_half_hourly	tbats	-0.149	-0.197	-0.119	0.000	-0.168	-0.119
	prophet	-0.070	-0.237	-0.167	0.054	-0.207	-0.221
	dhr_arima	-0.110	-0.121	-0.063	0.016	-0.019	-0.079
	mstl	0.034	-0.160	-0.108	0.131	-0.180	-0.239

Conclusions: LoMEF

- General trade-off between accuracy and interpretability
- Global Forecasting Models typically more complex than Local Models
- Translate Local Model Agnostic Post-Hoc interpretability to Global Forecasting Models
  - Neighbourhood - fit of the global model on original or bootstrapped series
  - Local interpretable models - Statistical models
- **Goal:** Focus on interpretability, not predictive accuracy
- Proposed local explainer performed models performed well in terms of Accuracy, Fidelity, and Consistency
- With this model, global models can be made as interpretable as local models

Local Interpretable Model Agnostic Rule-based Explanations for Forecasting (LIMREF)

Introduction

- The method was developed for the "FUZZ-IEEE Competition on Explainable Energy Prediction", where it was the winning method
- **Competition: Forecast the monthly electricity consumption for 3248 households in a coming year**
  - Half-hourly energy consumption data during for 1 year period
  - Prediction horizon - 12 months

Goal

Provide accurate and explainable forecasts for one year period

Proposed Methodology

Accurate Predictions

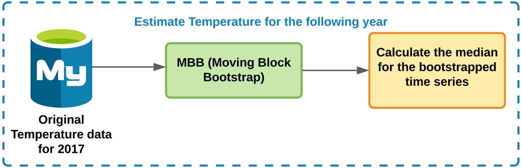
Introduced an energy prediction framework using global models to generate accurate forecasts

Explaining Predictions

Introduced a novel algorithm to produce local rule-based explanations for global time-series forecasting

Methodogy: How to Generate Accurate Predictions

- **Our model is based on the model that achieved 4th place in a past competition on the same data**
  - Uses daily series to trade-off granularity and complexity
  - Learns across series (global models), which is state of the art in forecasting - CIF2016, M4, M5 Competitions
- **Exogenous Variables**
  - Calendar effects
  - Temperature - Uses bootstrapping to generate future temperatures



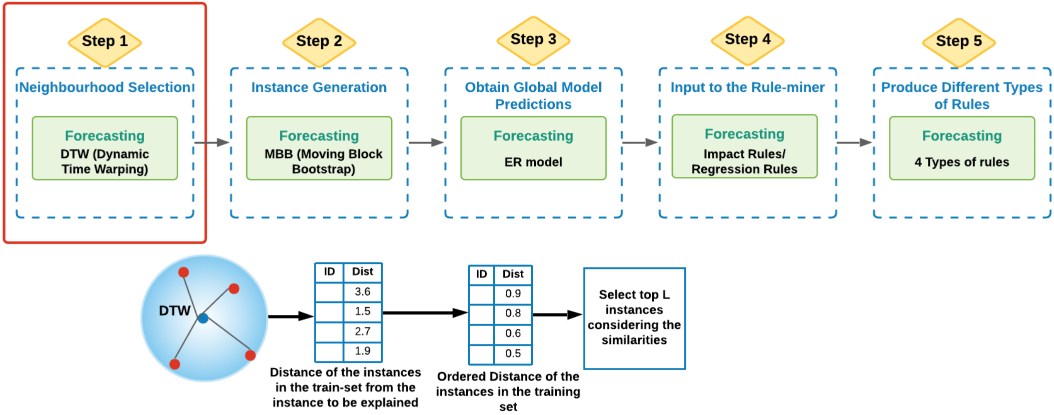
Methodology: Proposed Solution

LoRMIkA ? $\longrightarrow$ Time-series forecasting

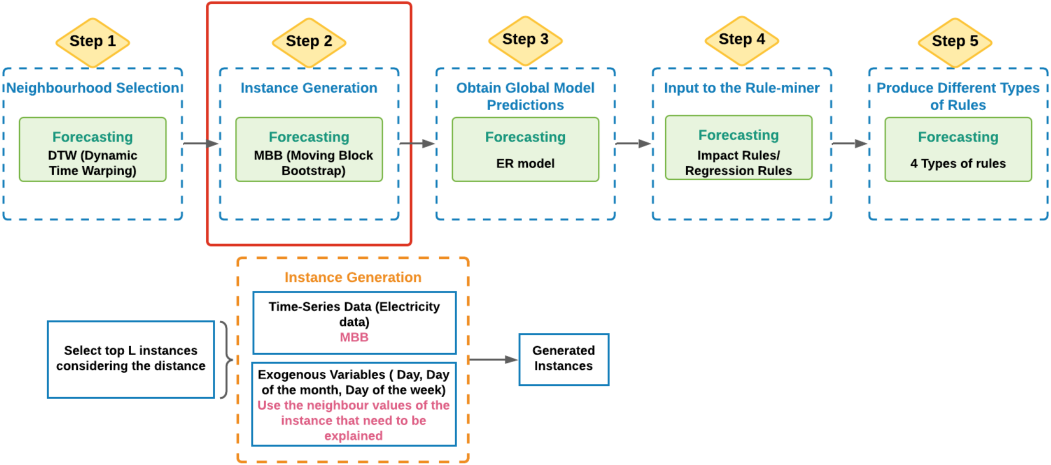
Proposed novel local model-agnostic interpretability approach to explain the forecast produced by global time-series models and produce Rule-based explanations

Proposed framework will keep the global forecasting model as a black-box model, in a model-agnostic way

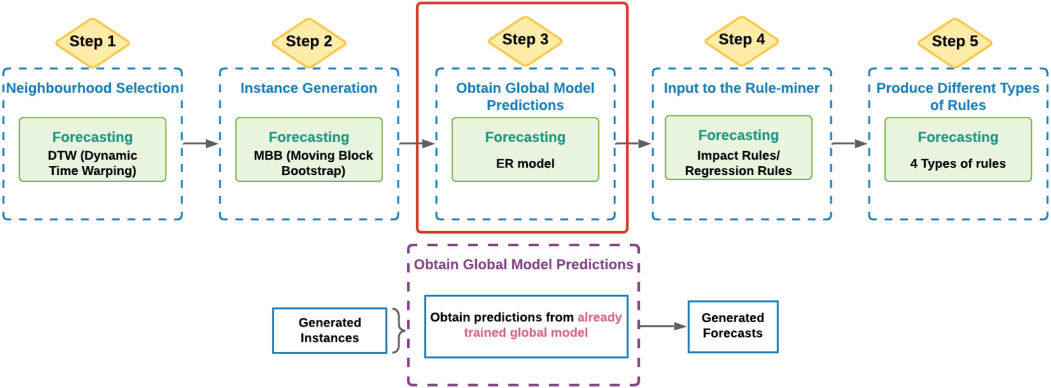
Methodology: LIMREF: Proposed Novel Interpretability Algorithm



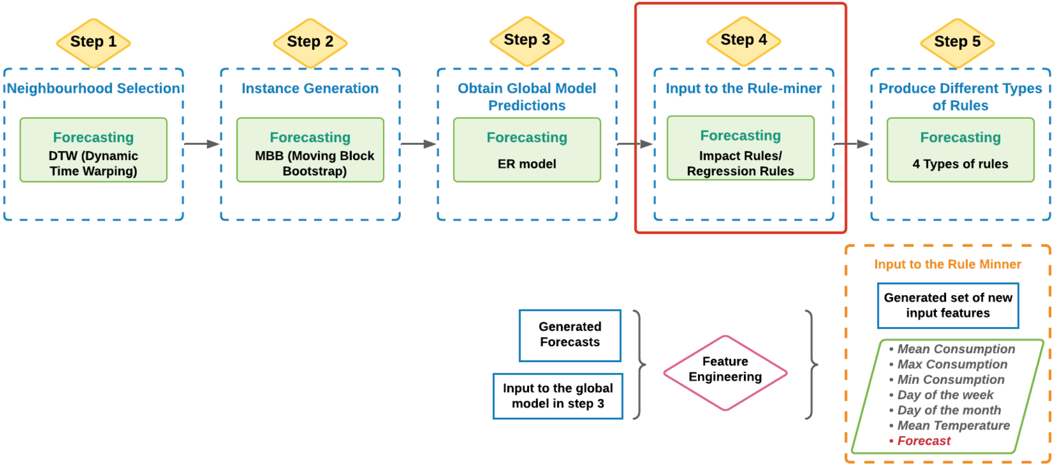
Methodology: LIMREF: Proposed Novel Interpretability Algorithm



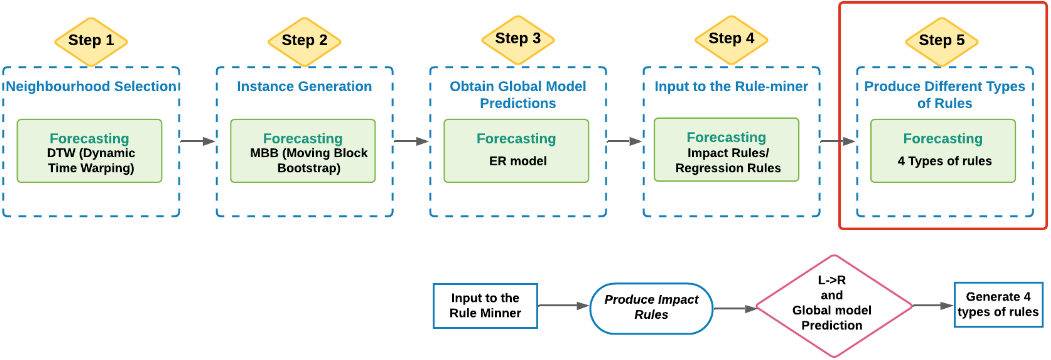
Methodology: LIMREF: Proposed Novel Interpretability Algorithm



Methodology: LIMREF: Proposed Novel Interpretability Algorithm



Methodology: LIMREF: Proposed Novel Interpretability Algorithm



Results: Examples for Each Rule Type

Table 2: Original Values of the Instance to be Explained

Mean Consumption	Max Consumption	Min Consumption	Temperature	Month	Monthly Forecast
3.5708kWh	6.131kWh	0kWh	5.98C	February	129.27kWh

- **Current Supporting Rules**
  - (What are the current conditions that support the global model prediction?)

Your predicted consumption is 129.27kWh. Because you are in month February.

Results: Examples for Each Rule Type

Table 3: Original Values of the Instance to be Explained

Mean Consumption	Max Consumption	Min Consumption	Temperature	Month	Monthly Forecast
3.5708kWh	6.131kWh	0kWh	5.98C	February	129.27kWh

- **Current Contradicting Rules**
  - (What are the current conditions that contradict the global model prediction?)

The conditions that currently exist that indicate a risk of an increased consumption by 118.85kWh for the particular month are 5.82 $<$ average temperature $\leq$ 6.20 & month $=$ February & minimum consumption $\leq$ 0.00.

Results: Examples for Each Rule Type

Table 4: Original Values of the Instance to be Explained

Mean Consumption	Max Consumption	Min Consumption	Temperature	Month	Monthly Forecast
3.5708kWh	6.131kWh	0kWh	5.98C	February	129.27kWh

- **Hypothetical Supporting Rules**
  - (What are the hypothetical conditions that support the global model prediction?)

The conditions that need to be satisfied to maintain the monthly predicted consumption would be $6.01 < $ average temperature ≤ 6.31.

Results: Examples for Each Rule Type

Table 5: Original Values of the Instance to be Explained

Mean Consumption	Max Consumption	Min Consumption	Temperature	Month	Monthly Forecast
3.5708kWh	6.131kWh	0kWh	5.98C	February	129.27kWh

- **Hypothetical Contradicting Rules**
  - (What are the hypothetical conditions that could potentially invert the global model prediction? **Counterfactual Rules**)

If you have mean consumption $>$ 9.16 it will increase your consumption by 275.53kWh.

A Complete Explanation of LIMREF

Table 6: Original Values of the Instance to be Explained

Mean Consumption	Max Consumption	Min Consumption	Temperature	Month	Monthly Forecast
3.463kWh	6.973kWh	0kWh	6.308C	February	119.01kWh

Your predicted consumption is 119.02kWh. Because you have month=February. If you have mean consumption $>$ 13.99 it will increase your consumption by 388.48kWh. The conditions that currently exist that indicate a risk of an increased consumption by 179.55kWh for the particular month are 6.05 $<$ average temperature $\leq$ 6.31 & month=February & min consumption $\leq$ 0.00. The conditions that need to be satisfied to maintain the monthly predicted consumption would be average temperature $>$ 6.31. Please note that here the mean, min and max consumption is calculated over the last 20 days. The temperature is the average temperature throughout the predicted month.

Conclusions: LIMREF

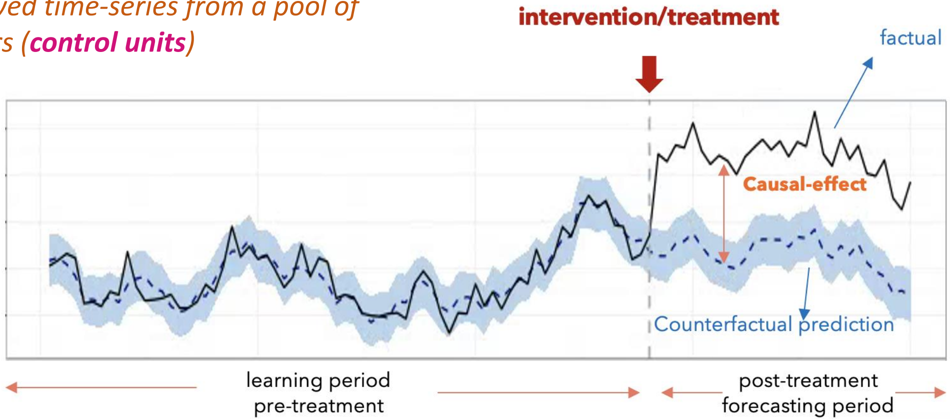
- Provide accurate and self-explaining predictions of the future monthly electricity consumption in a coming year
  - Accurate predictions - Train a global model (Expectile Regression)
  - Interpretable predictions - Proposed a novel explainability algorithm
- Explanations
  - Simple rules
  - Complete rules

Causal Inference Using Global Forecasting Models for Counterfactual Prediction

ARTIFICIAL

COUNTERFACTUAL

***IDEA:** to forecast an artificial **counterfactual trend** for the units affected by intervention (**treated units**) based on a large-dimensional panel of observed time-series from a pool of untreated peers (**control units**)*



HOW TO MEASURE

THE ACCURACY

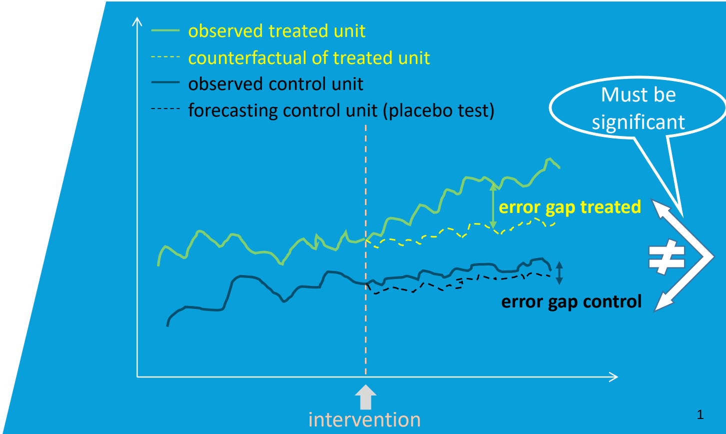
OF A

COUNTERFACTUAL?

PLACEBO TESTS:

1. **NULL EFFECT of the treatment for the control units:** errors of the **control group** (= effect of the intervention over the controls) **must be very low**

2. **SIGNIFICANT DIFFERENCE between the errors** from treated and control units (error gaps must be statistically significantly different from each other)



MONASH University

OUR EXPERIMENTS

I. Real-world datasets

AUSTRALIAN EMS CALLS

- 79 time series
- Training period: 41 points
- Forecasting period: 12 points
- **Intervention:** Effect of the increase of number liquor licenses over the EMS calls demand

911 EMERGENCY CALLS

- 62 time series
- Training period: 41 points
- Forecasting period: 7 points
- **Intervention:** Effect of the COVID-19 lockdown over 911 calls demand

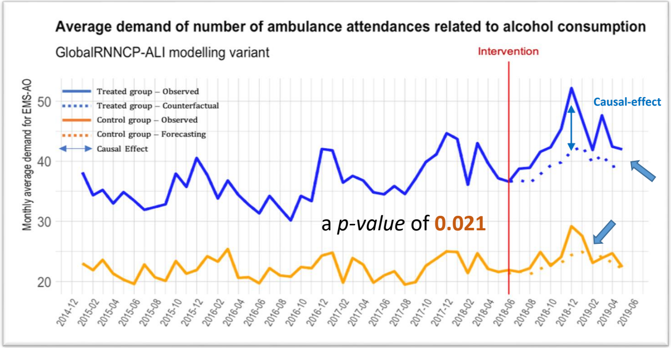


Table 1. Results for the 79 monthly series of NASS.

Method	Mean sMAPE	Median sMAPE	Mean MASE	Median MASE
DeepCPNet-ALI				
- All LGAs	0.1396	0.1341	0.9786	0.9467
- Treated Group	0.1405	0.1341	0.9812	**0.9413**
- Control Group	**0.1332**	0.1357	**0.9604**	0.9733
LSTM-univariate				
- All LGAs	0.1557	0.1537	1.0828	1.0205
- Treated Group	0.1570	0.1560	1.0900	1.0205
- Control Group	0.1464	**0.1323**	1.0331	0.9716
ARIMA				
- All LGAs	0.1560	0.1458	1.0888	0.9853
- Treated Group	0.1572	0.1427	1.0973	0.9643
- Control Group	0.1476	0.1490	1.0300	1.0069
ETS				
- All LGAs	0.1507	0.1441	1.0647	1.0078
- Treated Group	0.1515	0.1427	1.0705	0.9981
- Control Group	0.1450	0.1555	1.0244	1.0876

OUR EXPERIMENTS

I. Real-world datasets

AUSTRALIAN EMS CALLS

- 79 time series
- Training period: 41 points
- Forecasting period: 12 points
- **Intervention:** Effect of the increase of number liquor licenses over the EMS calls demand

911 EMERGENCY CALLS

- **62 time series**
- **Training period: 41 points**
- **Forecasting period: 7 points**
- **Intervention:** Effect of the COVID-19 lockdown over 911 calls demand

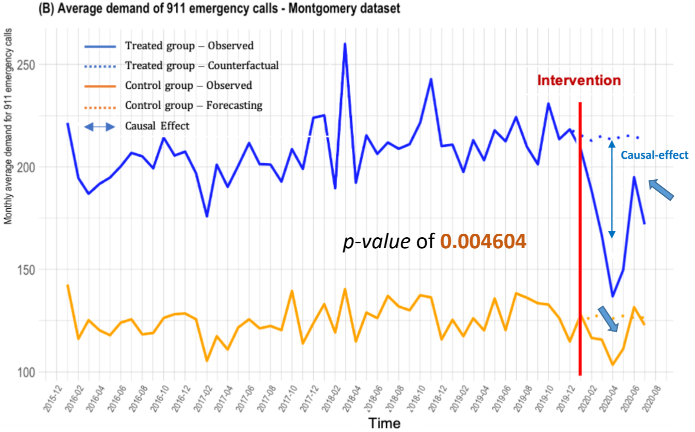


Table 2. Results for the 62 monthly series of 911 emergency calls.

	METHODS							
	Control Group				Treated Group			
Error Metric	DeepCPNet	LSTM-uni	ARIMA	ETS	DeepCPNet	LSTM-uni	ARIMA	ETS
Mean sMAPE	0.1899	0.1945	**0.1847**	0.1920	0.2525	0.2525	0.2508	0.2624
Median sMAPE	0.1740	0.1783	**0.1717**	0.1746	0.2290	0.2256	0.2250	0.2336
Mean MASE	0.8521	0.8698	**0.8288**	0.8616	1.3939	1.3926	1.3857	1.4498
Median MASE	**0.9176**	0.9370	0.9272	0.9430	1.3100	1.2842	1.2848	1.3334

OUR EXPERIMENTS – Simulated datasets

Monte Carlo experiment

S.1) Linear DGP: ARIMA + Var(3,4)

a. Exogenous Variable (regressor) = **RGCE** (with four lags) **using a stationary LINEAR ARIMA DGP** to build **100 time-series**

S.2) Non-linear DGP: SETAR + Var(3,4)

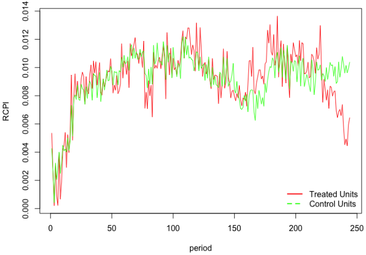
a. Exogenous Variable (regressor) = **RGCE** (with four lags) **using NONLINEAR SETAR DGP** to build **100 time-series**

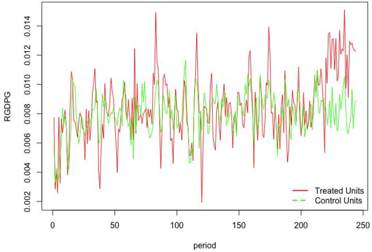
b. Endogenous Variables on VAR(3,4) **multivariate DGP process + inclusion of 1 regressor (RGCE)**:

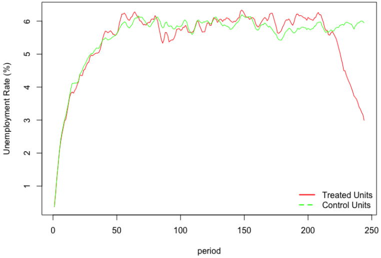
- **RCPI** --> within NN modelling will enter as **exogenous**
- **RGDPG** --> within NN modelling will enter as **exogenous**
- **UNRATE** --> within NN modelling will enter as **endogenous**

100 time-series

- **70 control**
- **30 treated**





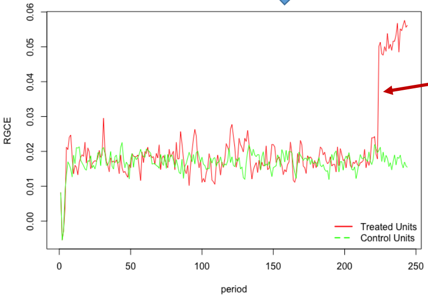


Using VAR(3,4) estimator:

causes

causes

Exogenous Variable (regressor) => RGCE



Intervention:

Adding an increase of 3.5% in the government expenditures from the period 224 towards

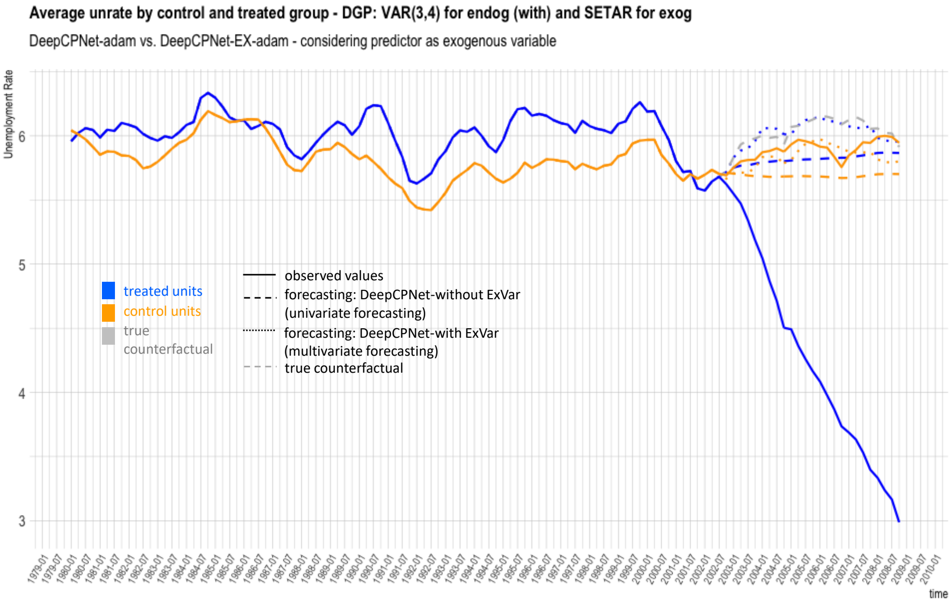
- quarterly data
- Training period: 199 time points
- Horizon forecasting: 21 time points

VAR(3,4) DGP for endogenous and SETAR DGP for exogenous

OUR

RESULTS

FOR SIMULATION 2: NON-LINEAR DGP



MONASH University

RESULTS for Simulated Data – simulation 2 using non-linear DGP:

DeepCPNet achieves the best results in the scenario with non-linear relationships across the time series

Control Units Prediction Errors

Model	meanSMAPE	medianSMAPE
DeepCPNet	0.03457	0.03833
DeepCPNet_EX	**0.01406**	**0.01227**
Arima	0.05706	0.06374
VAR4	0.01771	0.01582
CausalImpact	0.02478	0.02260
SCM	0.02832	0.03066

Counterfactual Prediction Errors

Model	meanSMAPE	medianSMAPE
DeepCPNet	0.03578	0.03351
DeepCPNet_EX	**0.00672**	0.00799
Arima	0.06597	0.07375
VAR4	0.05156	0.06027
CausalImpact	0.00684	**0.00663**
SCM	0.03515	0.02812

Idea: What about Probabilistic Forecasting?

- Study of target interventions affecting only the tails or the variance of the treated units' distribution rather than their mean or median (i.e., interventions affecting only parts of the distribution)

- Inspection of the impacts of interventions over skewed or heavy-tailed distributions

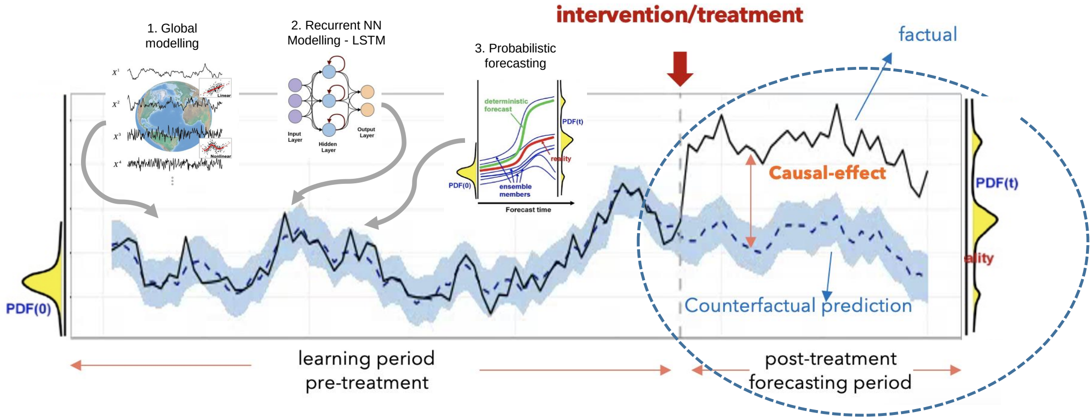
- It allows for decision-making under uncertainty by precisely predicting intervals to quantify forecast uncertainties

OUR NOVELTY PROBABILISTIC GFM

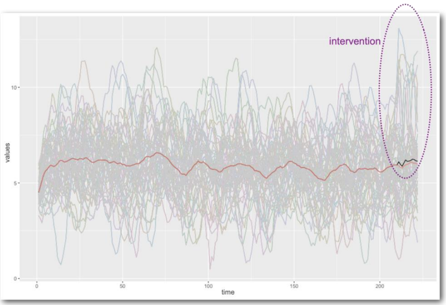
COUNTERFACTUAL and CAUSAL EFFECT distribution predictions

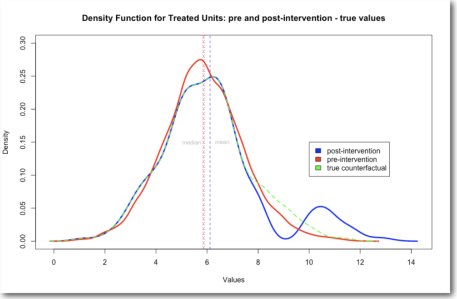
DeepProbCP:

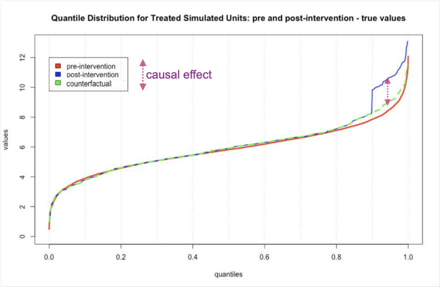
SOLUTION: global + Recurrent NN + probabilistic forecasting and inference framework to perform counterfactual distributional predictions



SIMULATION DESIGN OF THE
INTERVENTION OVER TREATED UNITS







- **Treated units**: 50 units (randomly chosen)
- **length** = 220 observations

Intervention:

- **how:** adding a constant of 1.574 (= 1 standard deviation from the treated units before the intervention) for the values > quantile 90%
- **The intervention causes:**
  - + 1.5% in the median
  - + 4.3% in the mean
  - > + 25% left tail

STEPS OF
DeepProbCP:

where:
- τ … some quantile
- y … observed data
- $\hat{y}$ … forecast data
- $(\cdot)_+ = \max(0,\cdot)$
- J … each time series
- T_0 … intervention time
- Θ … set of parameters
- h_t … hidden states

Step 1) DeepCPNet (GFM-RNN-LSTM based stacked model):

- *the MAE error function is replaced by the* ***"pinball loss"*** *(quantile loss) during the training and hyperparameter tuning processes*

$$L_\tau(y,\hat{y}) = \tau(y-\hat{y})_+ + (1-\tau)(\hat{y}-y)_+$$

- *estimate the future trajectory for each control and treated time series post-intervention, based on their (and covariates) pre-intervention data:*

$$\{y_{j,T_0+1:T}\}_{j=1}^{J} = m_G(y_{j,1:T_0}, \mathbf{x}_{j,1:T_0}, \Theta)$$

where $m_G(\cdot)$ is a multi-layer RNN with LSTM cells under the stacked design:

$$\mathbf{h}_{j,t} = r(\mathbf{h}_{j,t-1}, y_{j,t-1}, \mathbf{x}_{j,1:T_0}, \Theta)$$

Step 2) The step 1 is performed for each chosen quantile:

$$\begin{cases} (\hat{y}^{(\tau_1)}_{t+1,:j}, \dots, \hat{y}^{(\tau_1)}_{t+k,:j}) & = m_G(h_t, x^{(h)}_{:t,j}, \tau_1) \\ (\hat{y}^{(\tau_2)}_{t+1,:j}, \dots, \hat{y}^{(\tau_2)}_{t+k,:j}) & = m_G(h_t, x^{(h)}_{:t,j}, \tau_2) \\ \dots & \\ (\hat{y}^{(\tau_Q)}_{t+1,:j}, \dots, \hat{y}^{(\tau_Q)}_{t+k,:j}) & = m_G(h_t, x^{(h)}_{:t,j}, \tau_Q) \end{cases}$$

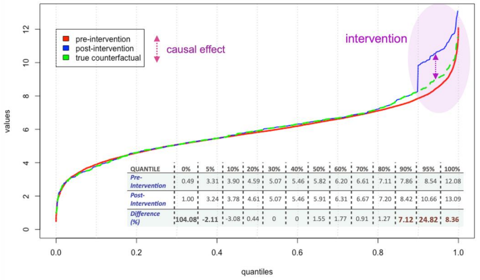
Step 3) Each forecasted quantile from step 2 is used to build the quantile distribution

- *using cubic splines functions, we interpolate each one the forecasted quantiles to build the "Quantile Distribution Function".*
- *we also estimate the quantile distribution function by splines for the observed data post-intervention*

Results of the EXPERIMENTS (1)

1) Simulated Data:

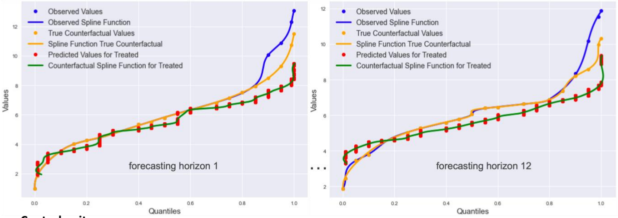
a. Simulated intervention



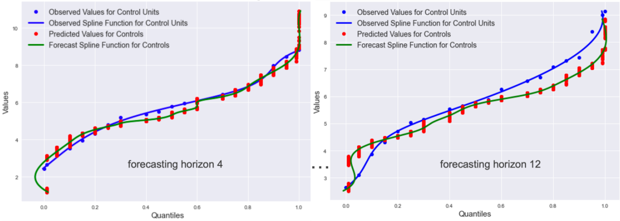
QUANTILE	0%	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%	100%
Pre-Intervention	0.49	3.31	3.90	4.59	5.07	5.46	5.82	6.20	6.61	7.11	7.86	8.54	12.08
Post-Intervention	1.00	3.24	3.78	4.61	5.07	5.46	5.91	6.31	6.67	7.20	8.42	10.66	13.09
Difference (%)	**104.08**	**-2.11**	-3.08	0.44	0	0	1.55	1.77	0.91	1.27	**7.12**	**24.82**	**8.36**

b. Estimated quantile distributions using splines

Treated units:



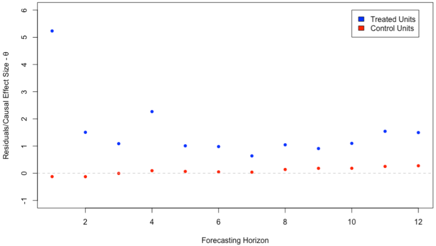
Control units:



Results of the EXPERIMENTS (1)

1) Simulated Data*:

c. Placebo Tests



Null effect of the intervention over the control units:

** 51 control time series and 50 treated time series using the nonlinear DGP employing SETAR models*

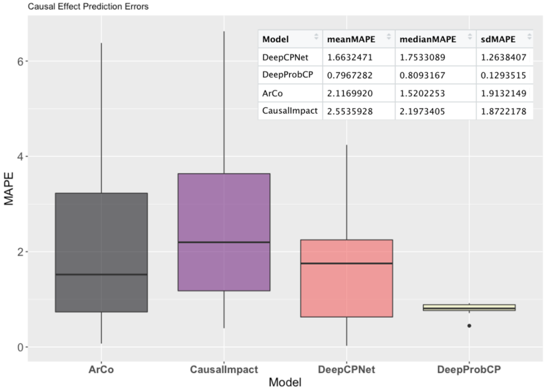
d. Performance comparing with benchmark models

Error gaps estimation for future trajectory:

POINT METRIC ERROR RESULTS FOR THE 101 SIMULATED SERIES GENERATED BY SETAR.

Method	Mean sMAPE	Median sMAPE	Mean MASE	Median MASE
DeepProbCP-0.50				
- Treated Group	**0.2088**	**0.1770**	**1.2202**	**0.9420**
- Control Group	**0.1892**	**0.1407**	**1.0322**	**0.8665**
CausalImpact				
- Treated Group	0.3237	0.2915	1.8509	1.6781
- Control Group	0.3561	0.3016	1.9302	1.8983
ArCo				
- Treated Group	0.2834	0.2488	1.6575	1.3993
- Control Group	0.2494	0.2206	1.3851	1.0676

Causal Effect errors comparing with the recovered true causal effect:

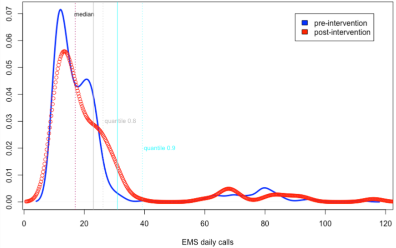


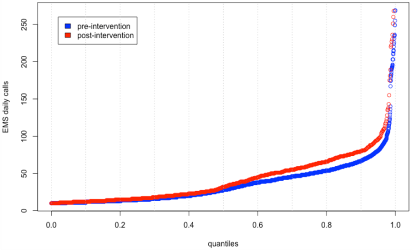
Model	meanMAPE	medianMAPE	sdMAPE
DeepCPNet	1.6632471	1.7533089	1.2638407
DeepProbCP	0.7967282	0.8093167	0.1293515
ArCo	2.1169920	1.5202253	1.9132149
CausalImpact	2.5535928	2.1973405	1.8722178

Results of the EXPERIMENTS (2)

2) Real-world data – Australian EMS calls:

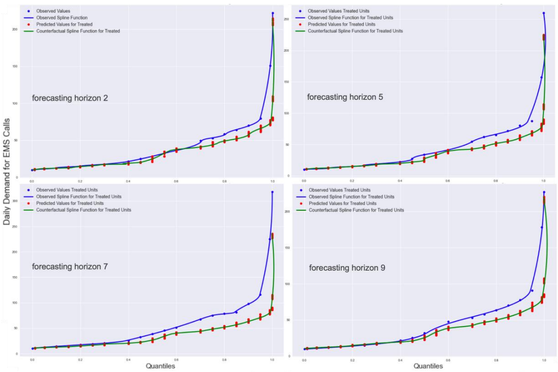
a. Intervention:





b. Estimated quantile distributions using splines

Treated units:



Conclusions

- **Global RNN-based approaches** (combination of global modelling, stacked architecture, LSTM cells, and COCOB optimizer) should emerge as a **new generation of machine learning models** **in the causal inference task** of **predicting counterfactual outcomes**, not being restricted to the pure forecasting tasks.

- The powerful causal inference mechanism developed in this research can **contribute to leverage the estimation of causal effect's measurement** which is crucial **to guide and support the policymakers**: (i) in their **decision making processes**, and (ii) **evaluation of their policies**

Conclusions Cont.

With this **causality-inspired DNN model** combined with **non-parametric probabilistic techniques** we can:

- improve the generalisation and adaptivity of the framework by leveraging the causality analysis to achieve reliable causal effect identifications in real-world scenarios.

- enrich the decision-making process with powerful tools capable of not only delivering point predictions of counterfactuals but also predictions of their distributions.

MONASH University

- Dilini Rajapaksha, Christoph Bergmeir, Rob J Hyndman (2021) LoMEF: A Framework to Produce Local Explanations for Global Model Time Series Forecasts. https://arxiv.org/abs/2111.07001
- Dilini Rajapaksha, Christoph Bergmeir, Rob J Hyndman (2020) Local Model-Agnostic Interpretability in Global Time Series Forecasting. In: 40th International Symposium on Forecasting (ISF) 2020, October 26 - 28, Virtual (Rio de Janeiro, Brazil).
- Priscila Grecov, Kasun Bandara, Christoph Bergmeir, Klaus Ackermann, Sam Campbell, Deborah Scott, Dan Lubman (2021) Causal Inference Using Global Forecasting Models for Counterfactual Prediction. In: PAKDD 2021: Advances in Knowledge Discovery and Data Mining, Springer International Publishing, pp. 282-294.
- P. Grecov, C. Bergmeir, K. Ackermann (2021) Causal Inference Using Global Forecasting Models for Counterfactual Prediction. In: 41st International Symposium on Forecasting (ISF) 2021, June 27 - July 08, virtual.

Thank you!

christoph.bergmeir@monash.edu