Continually learning out-of-distribution spatiotemporal data for robust energy forecasting

Dr. Arian Prabowo
Dr. Kaixuan Chen
Dr. Hao Xue
Dr. Subbu Sethuvenkatraman
Prof. Flora D. Salim

Centre for New Energy Technologies

Energy use forecasting – Why?

Energy use forecasting – Why?

- Australia has an **<u>energy market</u>** i.e. the price of electricity fluctuates according to supply and demand.

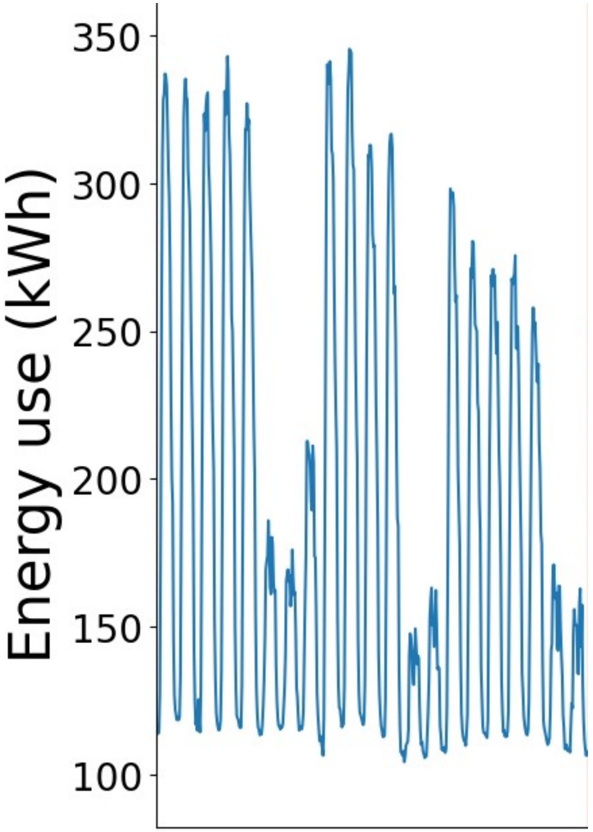
Energy use forecasting – Why?

- Australia has an **energy market** i.e. the price of electricity fluctuates according to supply and demand.
- Energy use forecasting is important to many stakeholders in the sectors. For examples:
  - Energy **retailers** need an accurate forecast to manage their infrastructure to effectively distribute electricity.
  - **Building managers** need an accurate forecast to efficiently schedule high energy operations e.g. pumping and heating water.
  - **Power plants operator** want to schedule maintenance only when the demand is low.

Energy use forecasting – Why?

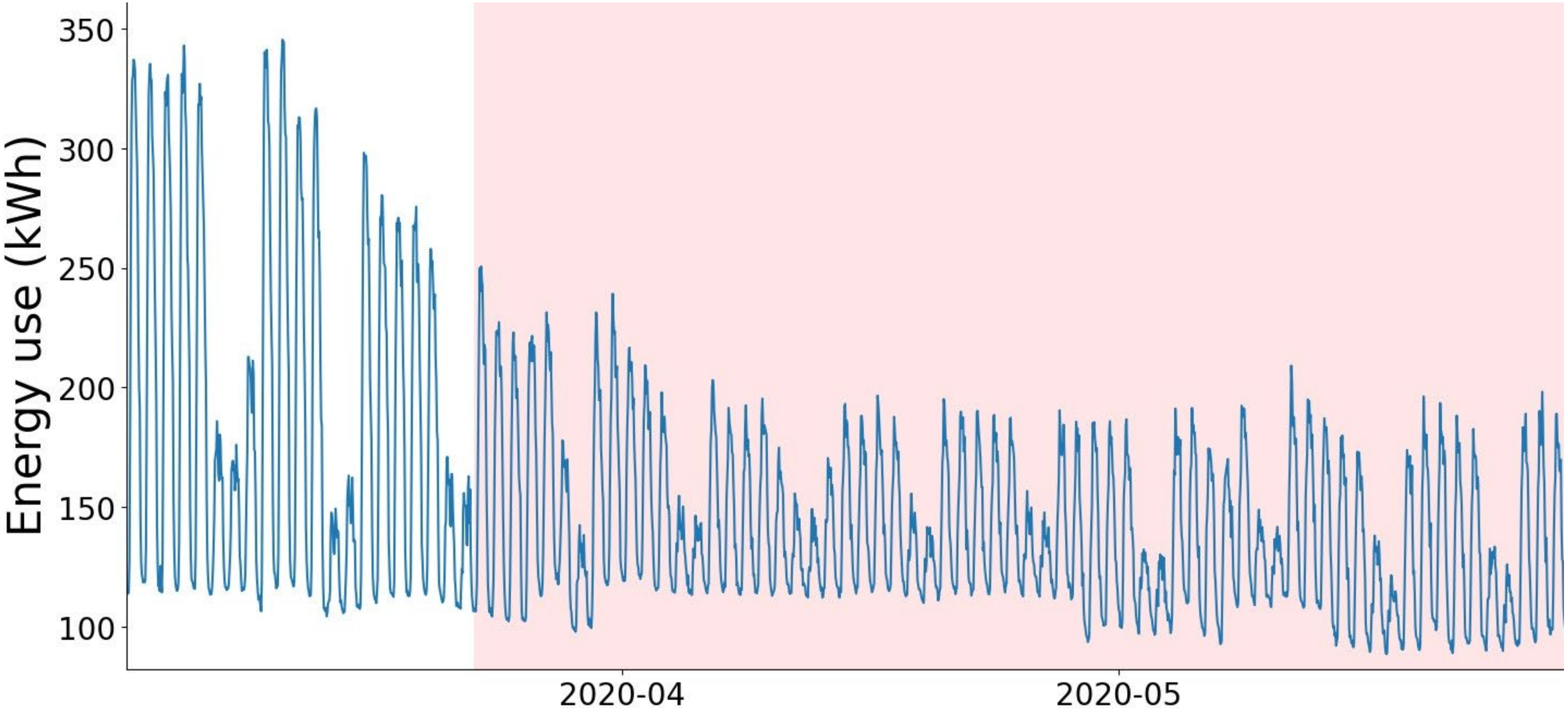
- Australia has an **<u>energy market</u>** i.e. the price of electricity fluctuates according to supply and demand.
- Energy use forecasting is important to many stakeholders in the sectors.
- Currently, the industry are only using traditional machine learning models.

Extreme events (e.g. COVID-19) change the energy use distribution patterns



Before lockdown

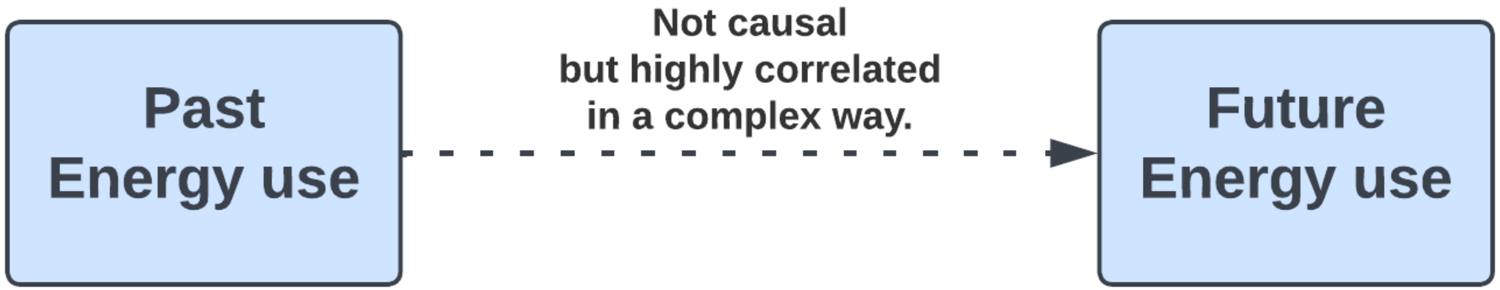
Extreme events (e.g. COVID-19) change the energy use distribution patterns



Before lockdown

After lockdown

Extreme events (e.g. COVID-19) change the energy use distribution patterns

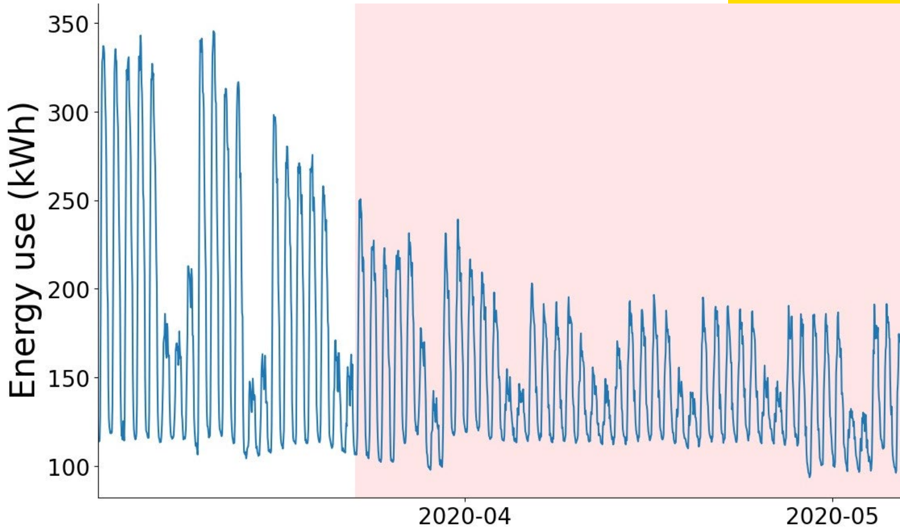


Extreme events (e.g. COVID-19) change the energy use distribution patterns

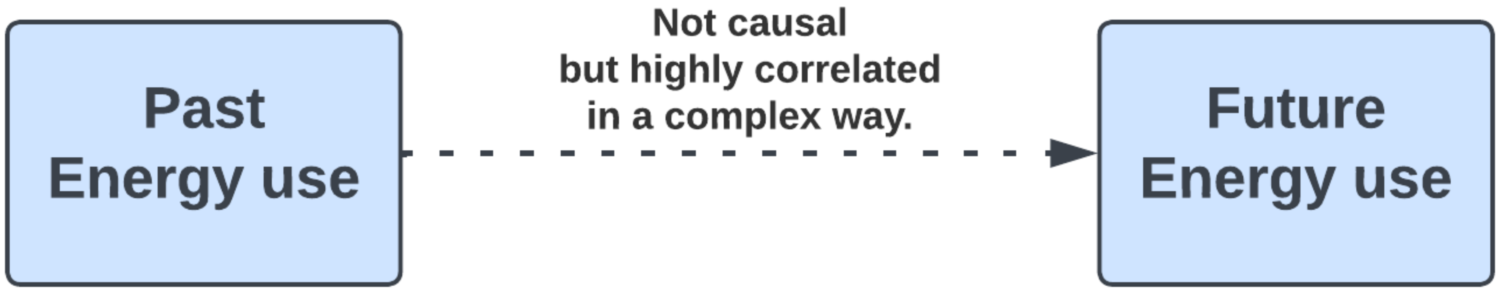
Past Energy use → **Future Energy use**

Not causal but highly correlated in a complex way.

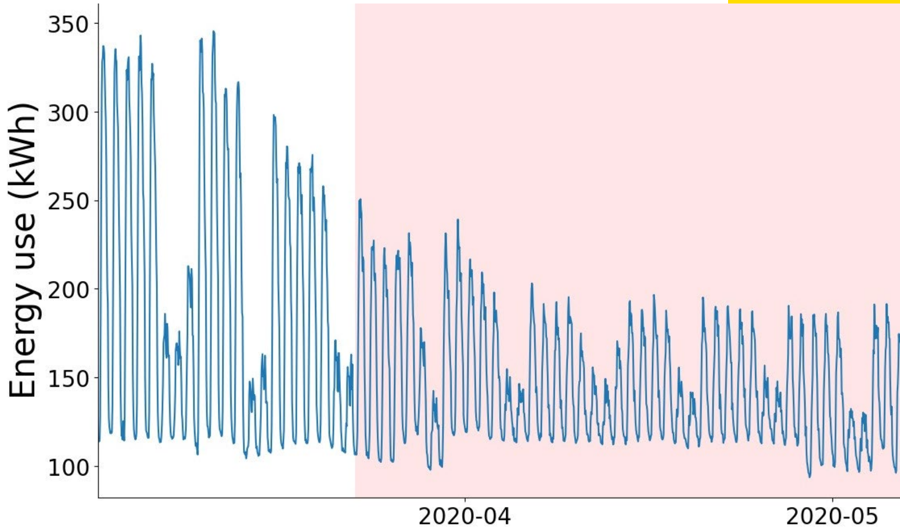
However, the complex correlation changes because of the changes in distribution due to lockdown.



Extreme events (e.g. COVID-19) change the energy use distribution patterns



We use **<u>continual learning</u>** to adapt to the changes in correlation due to changes in distribution.



Continual Learning

Traditional / Offline Learning	Continual Learning
Models are only trained once. Suffers when there are changes in distribution.	

Continual Learning

Traditional / Offline Learning	Continual Learning
Models are only trained once. Suffers when there are changes in distribution.	Models are updated as new data comes in. Adaptive to changes in distribution.

We use **continual learning** to perform accurate forecast even during **OoD** periods

Datasets: Energy

- 4 Building Complexes (BC)
  - To protect privacy, we aggregate nearby buildings, even with different building use (e.g. retail/office/residential), into one BC.

Datasets: Energy

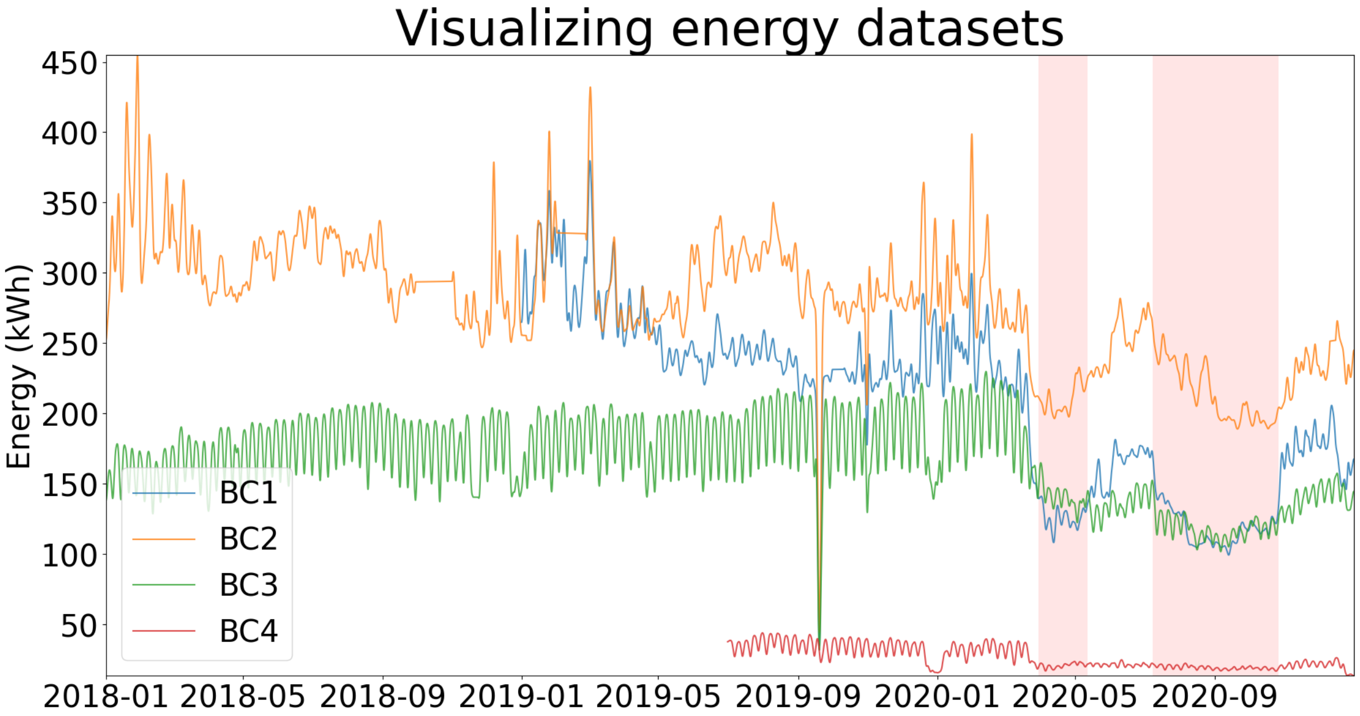
- 4 Building Complexes (BC)
  - To protect privacy, we aggregate nearby buildings, even with different building use (e.g. retail/office/residential), into one BC.
- All from Melbourne, Australia.
  - The city which experienced one of the longest and strictest COVID lockdown.

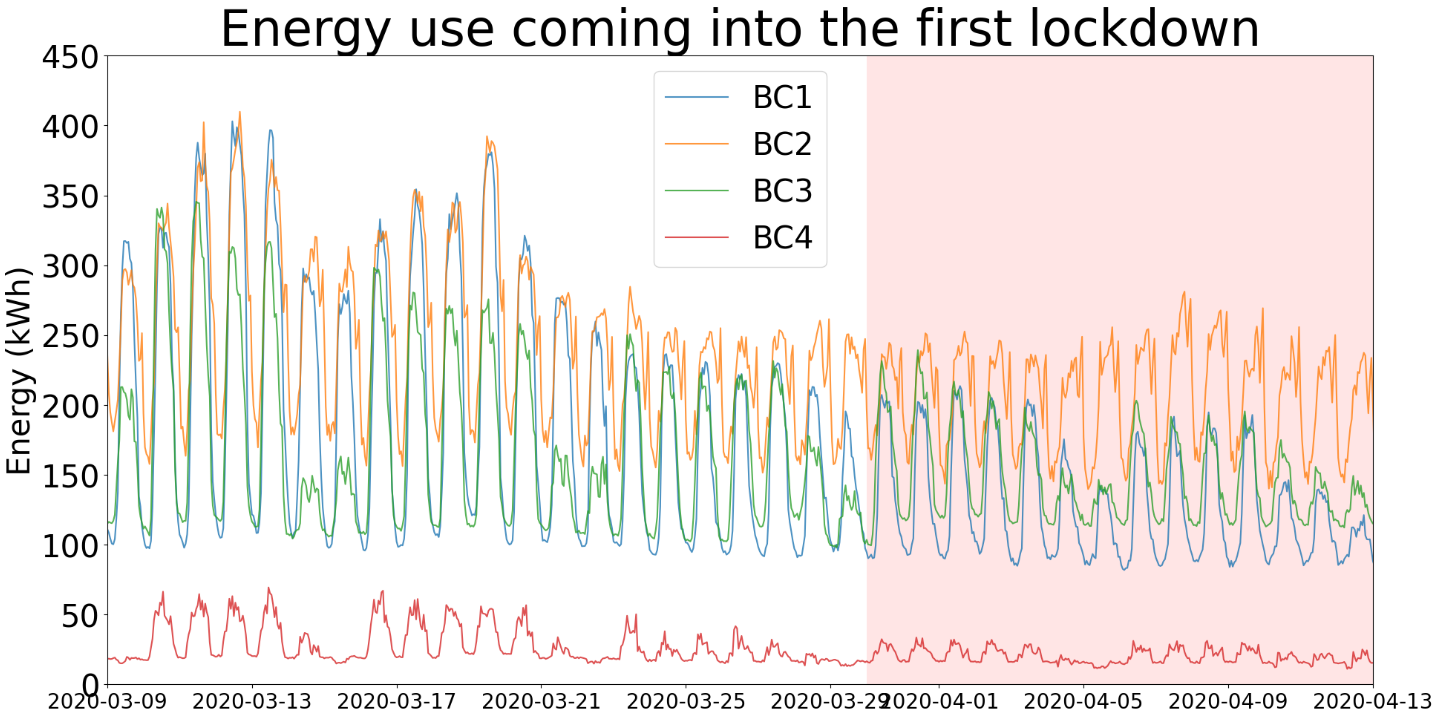
Datasets: Energy

The top plot
is smoothed.

The dataset
selection
is diverse.

They all have
different
distributions.





UNSW
SYDNEY

Datasets: Mobility

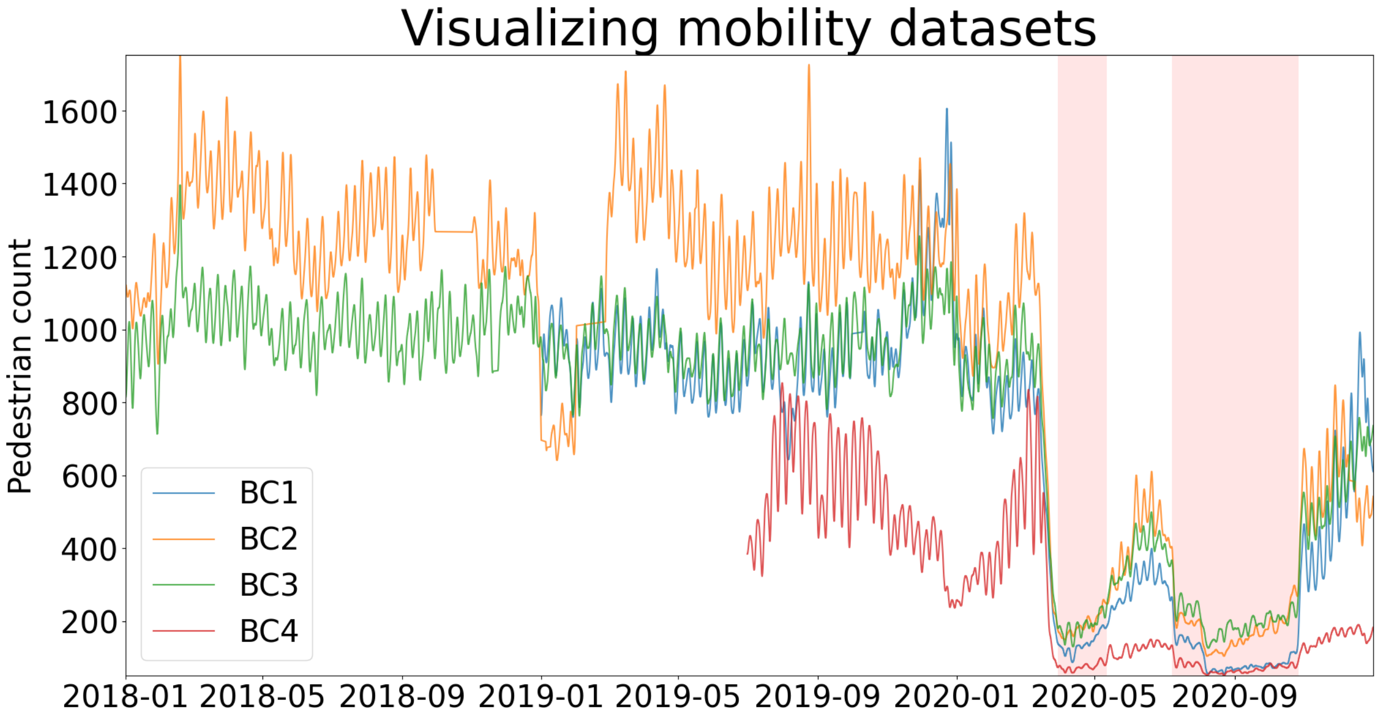
Installed in public space for by the City of Melbourne.

Does not capture images or videos to protect privacy.

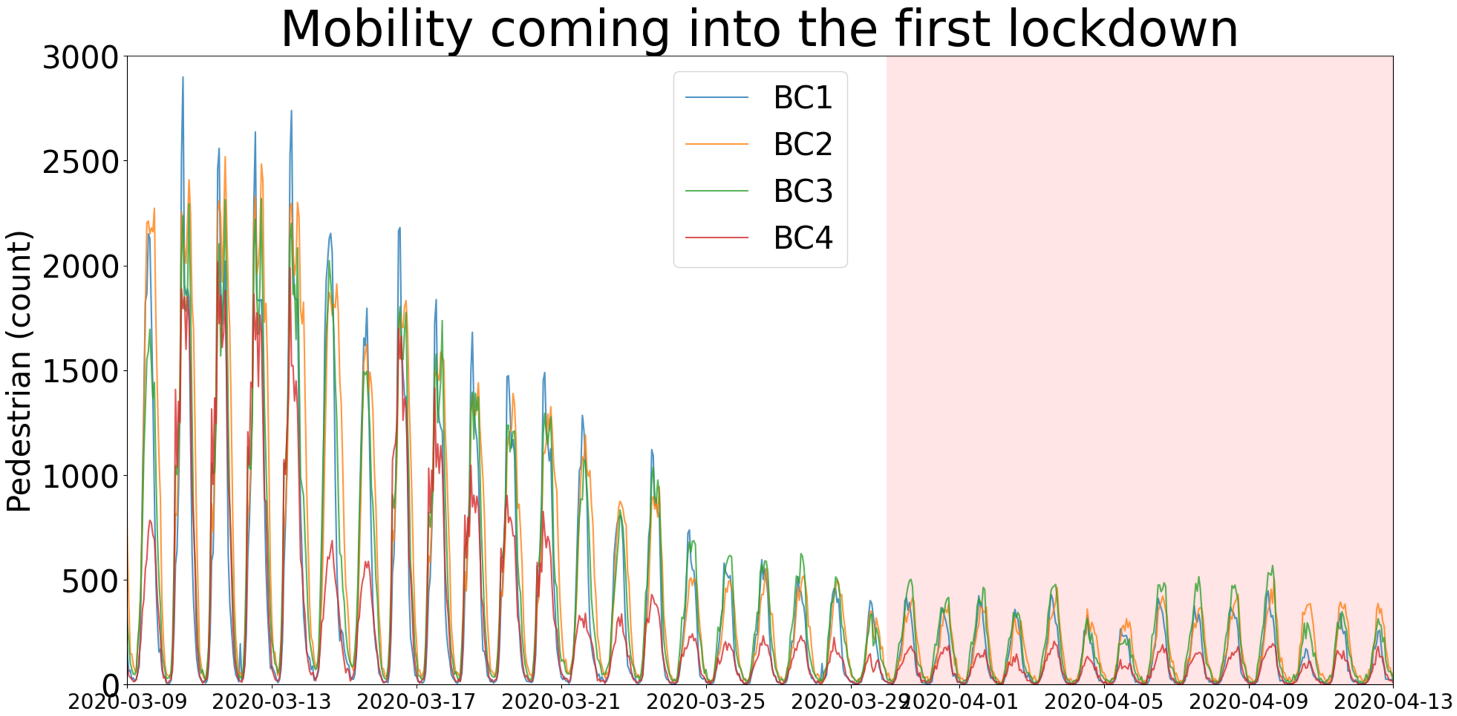
Publicly available, thus useful for stakeholders without access to data from inside the buildings e.g. energy retailer and power plants operator.

Datasets: Mobility

The top plot is smoothed.



The dataset selection is diverse.



They all have different distributions.

Table 1. The summary statistics of the four datasets, each of which represents an aggregated and anonymized building complex (BC).

		BC1	**BC2**	**BC3**	**BC4**
Temporal	start	2019-01-01	2018-01-01	2018-01-01	2019-07-01
	end	2020-12-31	2020-12-31	2020-12-31	2020-12-31
	num of record	17304	24614	26196	13200
	duration (years)	2.0	3.0	3.0	1.5
	granularity	hourly	hourly	hourly	hourly
Energy (kWh)	mean	207.27	278.17	166.42	26.64
	std	111.59	88.67	66.60	13.21
	min	3.25	1.32	5.38	0.00
	0.25	112.72	203.38	112.62	17.45
	median	169.29	272.68	144.34	21.75
	0.75	297.33	342.10	206.63	31.06
	max	611.67	709.41	371.64	83.01
Mobility (count)	mean	661.4	977.8	804.9	295.8
	std	876.8	936.2	761.4	387.3
	min	0	0	0	0
	0.25	37	149	127	33
	median	209	614	528	135
	0.75	1004	1818	1349	386
	max	6025	5053	3780	2984

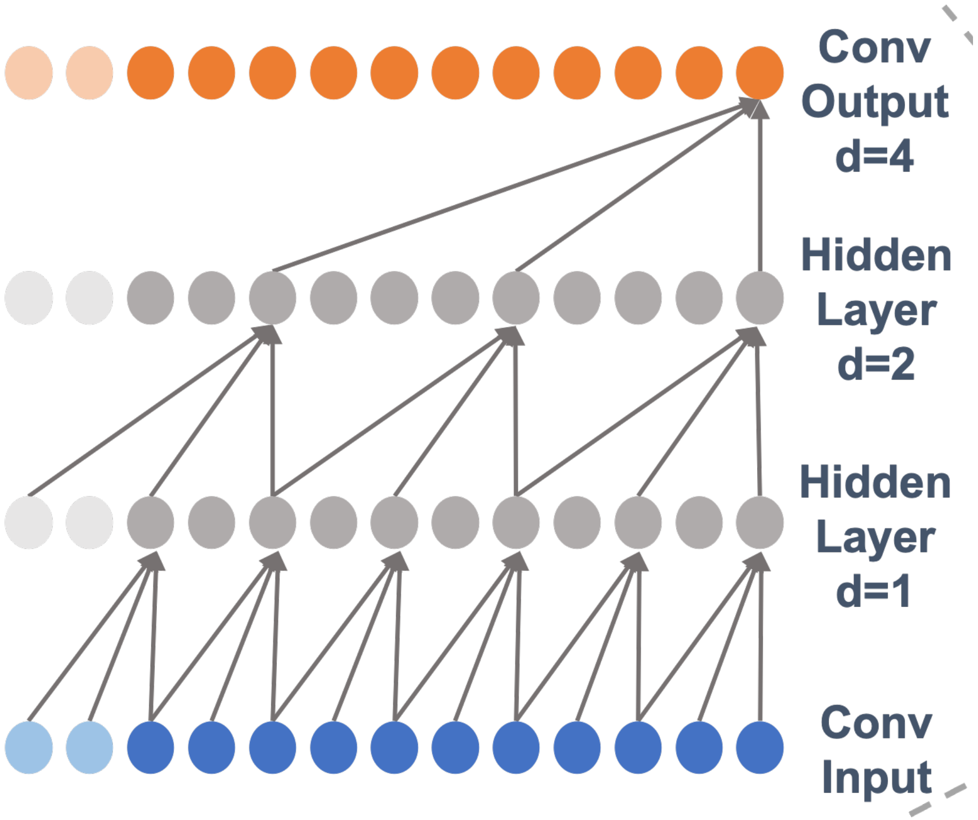
Experimental setup

- Last 24 hours as input to forecast the next 24 hours (i.e. autoregressive)
- We also have temperature data
- We divide our data into two periods: pre- and post- Lockdown

- We compare 4 different CL methods implemented on a TCN backbone.

UNSW SYDNEY

(Temporal Convolutional Network) TCN



(a) Dilated Causal Convolution

- TCN is a popular convolutional architecture in timeseries. (sometimes also known as WaveNet)
- The **dilation** operation exponentially increase receptive field of subsequent layers.
- The **causal** part means that the receptive field is only over preceding nodes, so they are not exposed to data from the future.

Without CL, DL perform worse after lockdown.

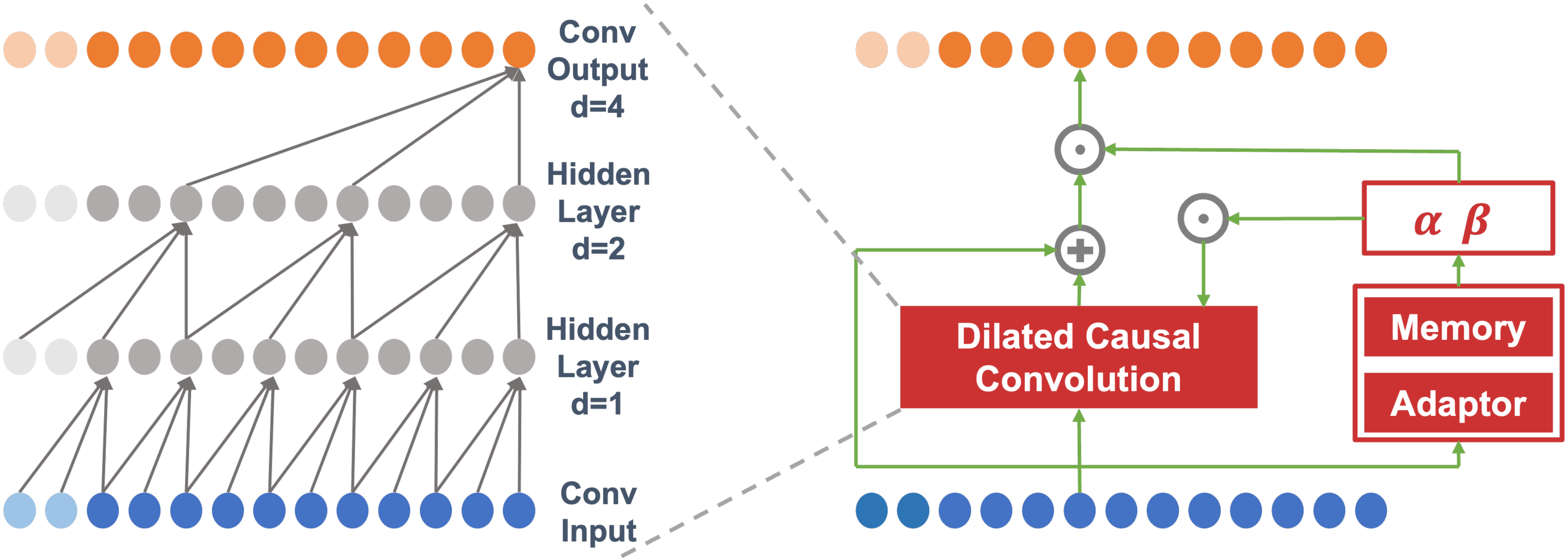
FSNet is consistently the best

Table 3. Comparing the performance of different algorithm with or without continual learning (CL). The metric used is MAE. Results are average over 10 runs with different random seed. The standard deviation is shown.

	dataset	FSNet (no CL)	FSNet	TCN (no CL)	OGD	ER	DR++
Pre-Lockdown	BC1	0.3703 ±0.0607	**0.1583** ±0.0280	0.3668 ±0.0379	0.2056 ±0.0413	0.1820 ±0.0217	0.1696 ±0.0130
	BC2	0.6272 ±0.0914	**0.1712** ±0.0063	0.5176 ±0.0607	0.2465 ±0.0105	0.2322 ±0.0056	0.2272 ±0.0062
	BC3	0.6750 ±0.0638	**0.2462** ±0.0151	0.6500 ±0.0698	0.3308 ±0.0812	0.2862 ±0.0432	0.2726 ±0.0334
	BC4	1.0018 ±0.1053	**0.2802** ±0.0312	1.1236 ±0.1040	0.3910 ±0.0520	0.3511 ±0.0323	0.3408 ±0.0210
Post-Lockdown	BC1	0.4537 ±0.0517	**0.1429** ±0.0275	0.4179 ±0.0443	0.1797 ±0.0342	0.1589 ±0.0168	0.1482 ±0.0094
	BC2	0.6506 ±0.0994	**0.1628** ±0.0057	0.5209 ±0.0535	0.2313 ±0.0085	0.2188 ±0.0060	0.2148 ±0.0068
	BC3	0.7168 ±0.0632	**0.2255** ±0.0145	0.7083 ±0.0793	0.3014 ±0.0709	0.2636 ±0.0373	0.2518 ±0.0286
	BC4	1.8415 ±0.2765	**0.3314** ±0.0520	1.8307 ±0.2319	0.4496 ±0.0643	0.4162 ±0.0475	0.4043 ±0.0338

UNSW SYDNEY

FSNet architecture with TCN backbone

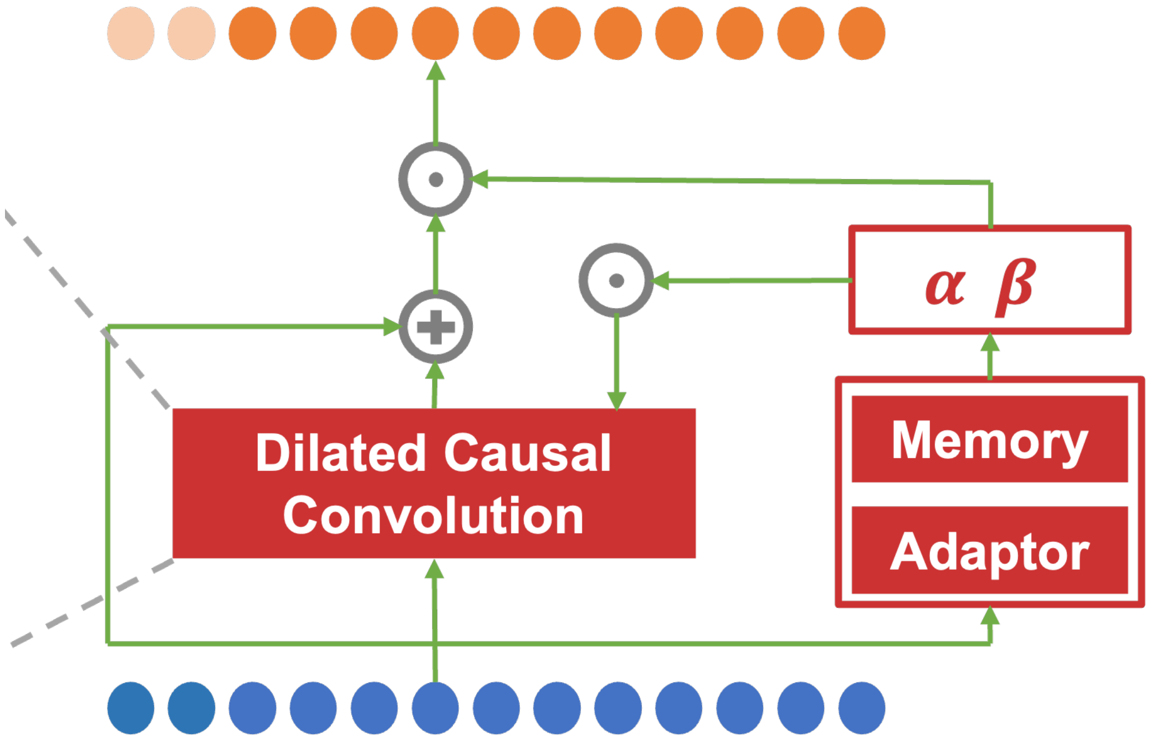


(a) Dilated Causal Convolution

(b) Convolution Layer with Adaptor and Memory

FSNet architecture with TCN backbone

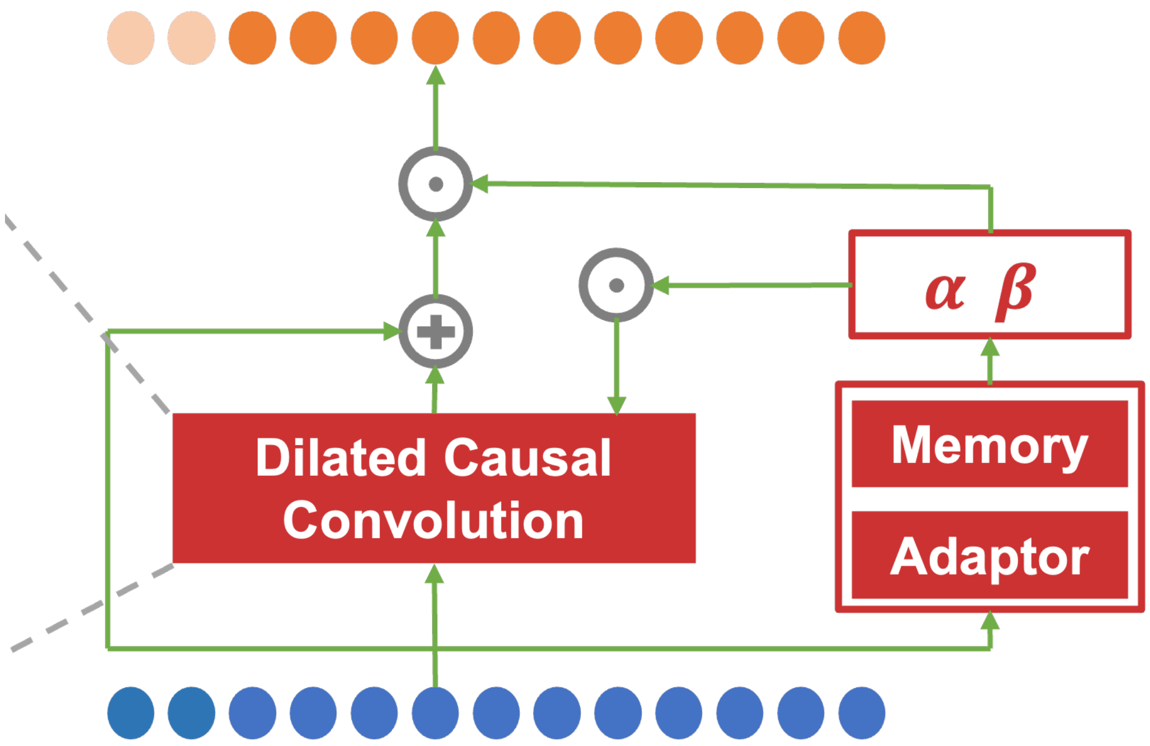
- Updating the entire model (OGD) whenever there's a new data is **inefficient**, and risk **catastrophic forgetting** where the model forget how to perform well on older task / distribution.



(b) Convolution Layer with Adaptor and Memory

FSNet architecture with TCN backbone

- Updating the entire model whenever there's a new data is inefficient, and risk catastrophic forgetting.
- So we only update the **memory module**.
- And the **adapter module** decides when to update the memory module.



(b) Convolution Layer with Adaptor and Memory

Mobility is important, especially post-Lockdown, especially with temperature.

Although the size of the improvement varied between BCs.

Table 2. Performance comparison between different contextual features. Results are average over 10 runs with different random seed. The standard deviation is shown. The algorithm used was FSNet with continual learning. +M is the improvement of adding mobility over no context, +T is the improvement of adding temperature over no context, T+M is the improvement of adding mobility over temperature only.

(MAE)	dataset	no context	mobility only	temp. only	both	+M	+T	T+M
Pre-Lockdown	BC1	0.1591 ±0.0252	0.1587 ±0.0334	0.1595 ±0.0269	0.1516 ±0.0332	0.0004	-0.0004	0.0079
	BC2	0.1711 ±0.0085	0.1993 ±0.0385	0.1947 ±0.0391	0.1708 ±0.0068	-0.0282	-0.0236	0.0239
	BC3	0.2629 ±0.0373	0.2866 ±0.0534	0.2509 ±0.0262	0.2403 ±0.0095	-0.0237	0.0120	0.0105
	BC4	0.2706 ±0.0370	0.2516 ±0.0206	0.3142 ±0.1581	0.2776 ±0.0312	0.0190	-0.0436	0.0366
Post-Lockdown	BC1	0.1484 ±0.0318	0.1475 ±0.0464	0.1434 ±0.0283	0.1369 ±0.0355	0.0033	0.0041	0.0041
	BC2	0.1636 ±0.0085	0.1902 ±0.0371	0.1849 ±0.0381	0.1624 ±0.0063	0.0072	-0.0194	0.0053
	BC3	0.2418 ±0.0374	0.2654 ±0.0537	0.2299 ±0.0252	0.2198 ±0.0089	-0.0014	-0.0251	0.0355
	BC4	0.3236 ±0.0602	0.2943 ±0.0294	0.4134 ±0.3215	0.3282 ±0.0502	0.0293	-0.1191	0.0852

UNSW SYDNEY

Current work
ECML PKDD ADS 2023

- 4 Building Complexes
- 2 periods:
  - Pre-lockdown
  - Post-lockdown
- 4 Continual Learning methods

Follow up work
BuildSys 2023

- 13 Building Complexes
- 5 periods:
  - Pre-lockdown
  - 1st lockdown
  - Inter-lockdown 1
  - 2nd lockdown
  - Inter-lockdown 2
- 13 methods
  - Naïve, statistical, classical ML, deep learning, online learning, and continual learning.

Thank you. Questions?

Conclusion:

For energy forecasting during anomalous periods:

1. Mobility data is at least as important as temperature, and it complements temperature.
2. Continual learning is superior even before lockdown, and it is even more pronounced after lockdown. Among all CL methods, FSNet has superior performance.

Link to this paper

Link to our follow up work, with 13 datasets!

UNSW SYDNEY

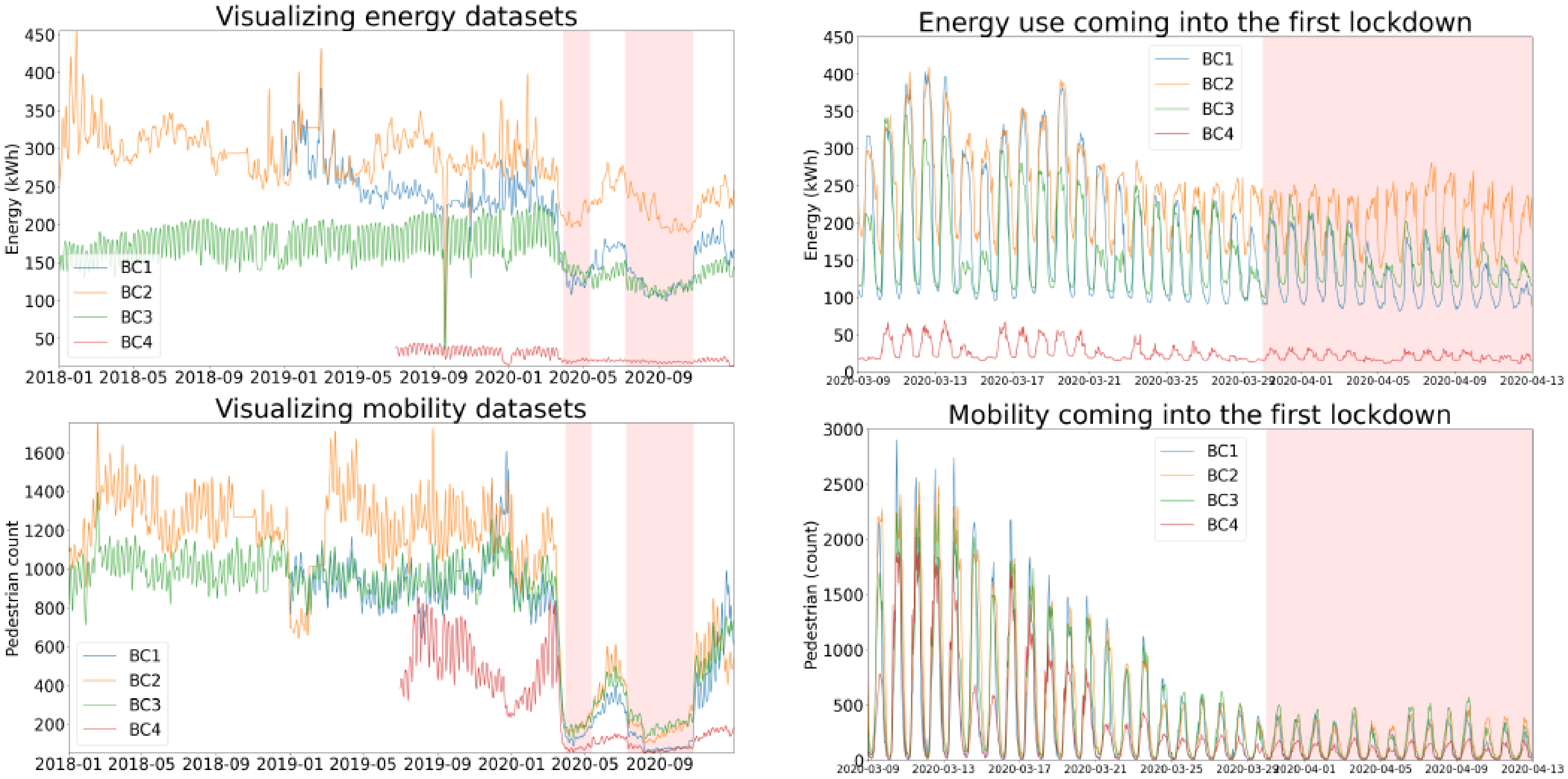


Fig. 2. Visualizing the four datasets and their features, showing the significant changes in distributions due to lockdowns. Plots on the left column are smoothed with a Gaussian filter with sigma = 24 hours. Red areas are lockdowns.