Digital Digital Twins
Some Experiences

Dr Marc Roper
Dept. Computer and Information Sciences
University of Strathclyde
Glasgow, UK

SANS II – 2nd-3rd May 2024

University of Strathclyde Glasgow

Digital Digital Twins?

Digital twins traditionally create a mapping between physical and virtual systems

- “A … high-fidelity model of the system which can be used to emulate the actual system.”

Digital digital twins create a mapping between virtual systems and virtual systems

- Few instances of these

Why do we need them?

Actual system may be too large/slow/complex/expensive… to provide answers in the time/budget available

- e.g. a CDN

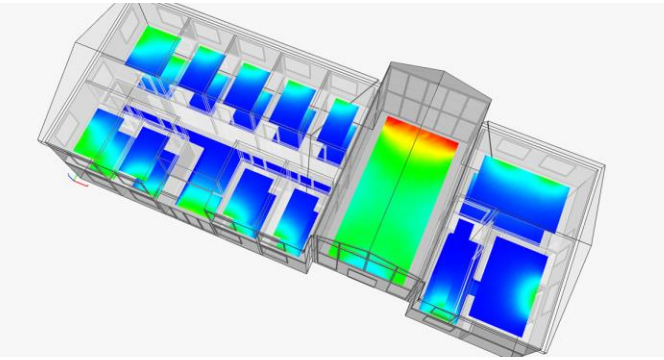
Need an approximation to the system to allow us to perform the necessary computations

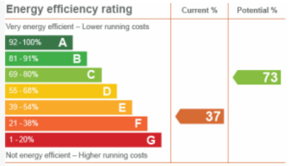
- e.g. Evaluate different configurations in the case of the CDN

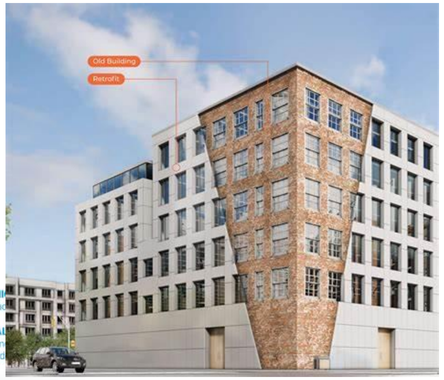
Predicting Building Energy Performance

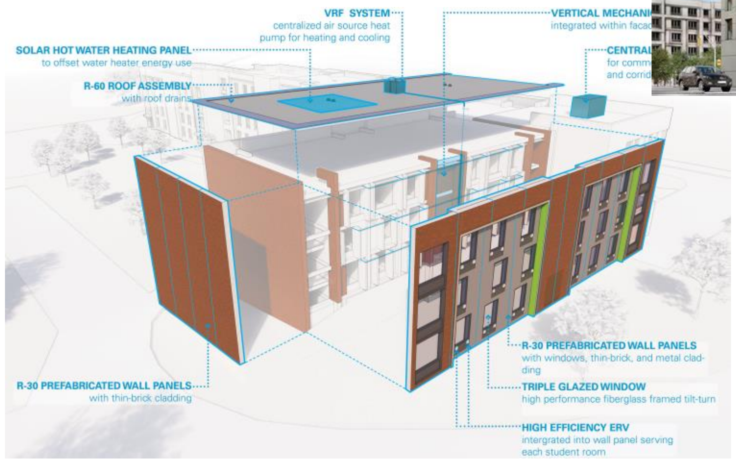
(Data Lab funded in conjunction with Architecture)

- Commercial buildings need to meet strict energy performance standards
- Retrofitting – process of upgrading the building
    - Huge number of options
    - Clients need to know the cost/benefit analysis
        - Multi-objective optimisation problem
    - Software available to simulate possible changes took days to run









Predicting Building Energy Performance

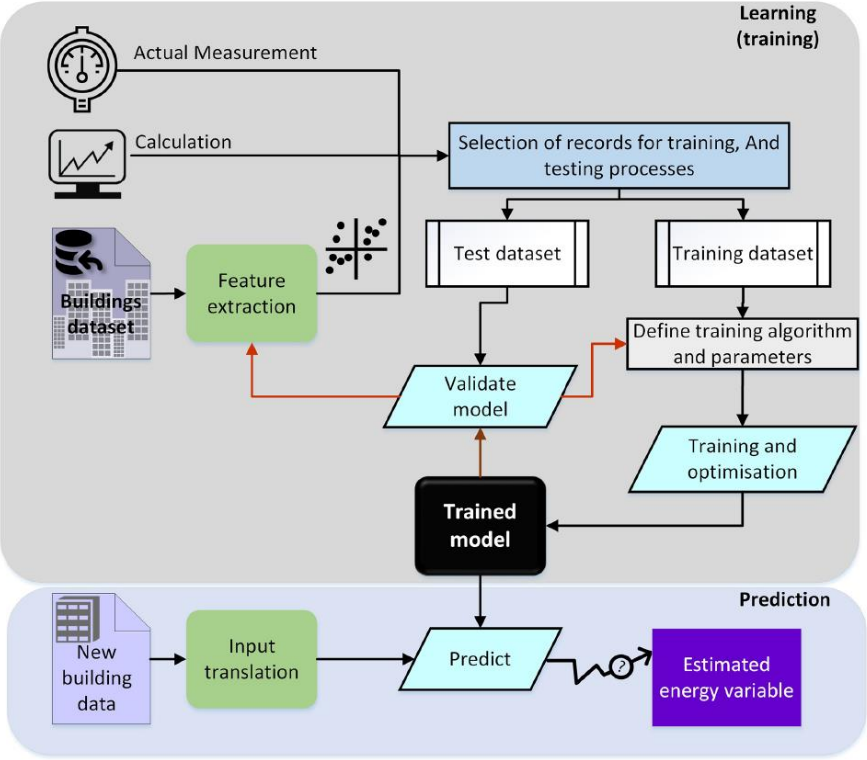
- Explored the performance of a range of machine learning models against two Building Performance Simulation tools: EnergyPlus and Ecotect
- Datasets already existed for both tools covering a variety of climate conditions, building types and parameters

Feature	Unit	Range	Variation	Code
Inputs				
Relative compactness	-	0.62 – 0.98	12	rc
Surface area	m^2	514 – 808	12	sa
Wall area	m^2	245 – 416	7	wa
Roof area	m^2	110 – 220	4	ra
Overall height	m	3.5, 7	2	oh
Orientation	-	2 – 5	4	ori
Glazing area	m^2	0 – 0.4	4	glza
Glazing area distribution		0 – 5	6	glzd
Targets				
Heating load	KWh/m^2	6 – 43	-	heat
Cooling load	KWh/m^2	10 – 48	-	cool

Also carried out feature engineering to reduce size of input space

Frequency and size of building types in EnergyPlus data

Building Usage	Type	Area (m^2)	Volume (m^3)	No. of E+ zones	No. of samples
Health	Hospital	22,422	88,864	55	3827
	Outpatient	3,804	11,932	118	5504
Home	Mid-rise	3,135	9,553	36	37173
	Apartment				
	Single Family	78532			
Hotel	Large	11,345	35,185	43	5504
	Small	4,014	11,622	67	5468
Office	Large	46,320	178,146	73	275345
	Medium	5503	4,982	18	19,741
	Small	511	1,559	5	5483
Restaurant	Full Service	5,502	55,035	2	3824
	Quick Service	232	708	2	5505
Retail	Stand Alone	2,294	13,993	5	5503
	Strip_Mall	2,090	10,831	10	5498
	Supermarket	45,002	900,272	6	5554
School	Primary	6,871	27,484	25	5505
	Secondary	19,592	95,216	46	5507
Warehouse	–	4,835	39,241	3	5492



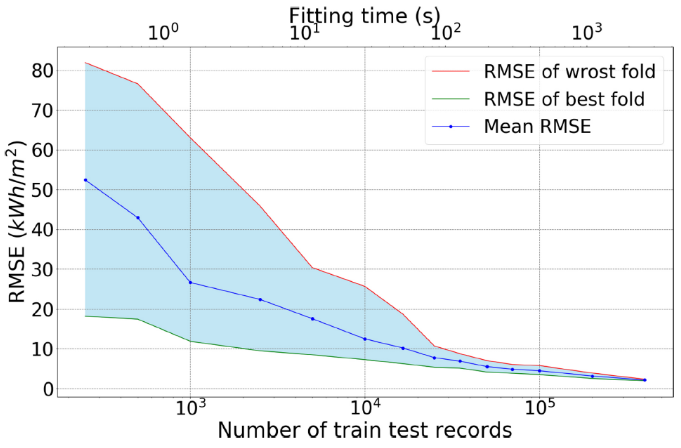
Predicting Building Energy Performance

- Compared a range of models :
  - ANN, SVM, GP, RF and GBRT
- Evaluated performance on heating and cooling over 25,000 data points (simulations)

	SVM		RF		NN		GBRT		XGBoost	
	Heat	Cool	Heat	Cool	Heat	Cool	Heat	Cool	Heat	Cool
R^2	0.965	0.973	0.973	0.968	0.966	0969	0.980	0.986	0.982	0.986
RMSE	14.318	8.763	12.720	9.400	14.068	9.376	10.721	6.296	10.386	6.270
MAE	5.622	3.465	5.057	4.841	7.472	4.932	4.400	3.365	4.130	3.143
Fit time (s)	177.66	406.31	6.35	34.873	126.29	10.88	6.363	1.789	4.897	4.871
Mean fit time (s)	1641.91	1197.16	17.6	19.54	21.32	17.19	4.85	4.92	4.61	4.55
Test time	0.483	0.507	0.333	0.595	0.008	0.010	0.244	0.078	0.228	0.219

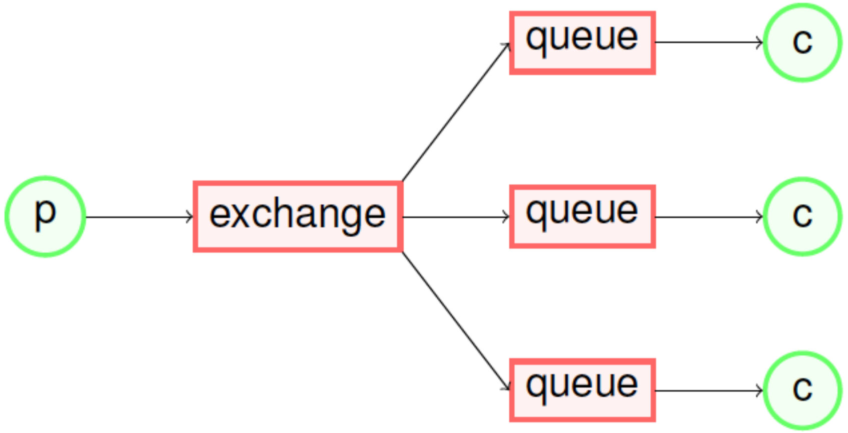
Some variation between models:

- GBRT and XGBoost most accurate
- SVM longest training time
- NN and XGBoost require a large amount of training data
- NN fastest prediction time



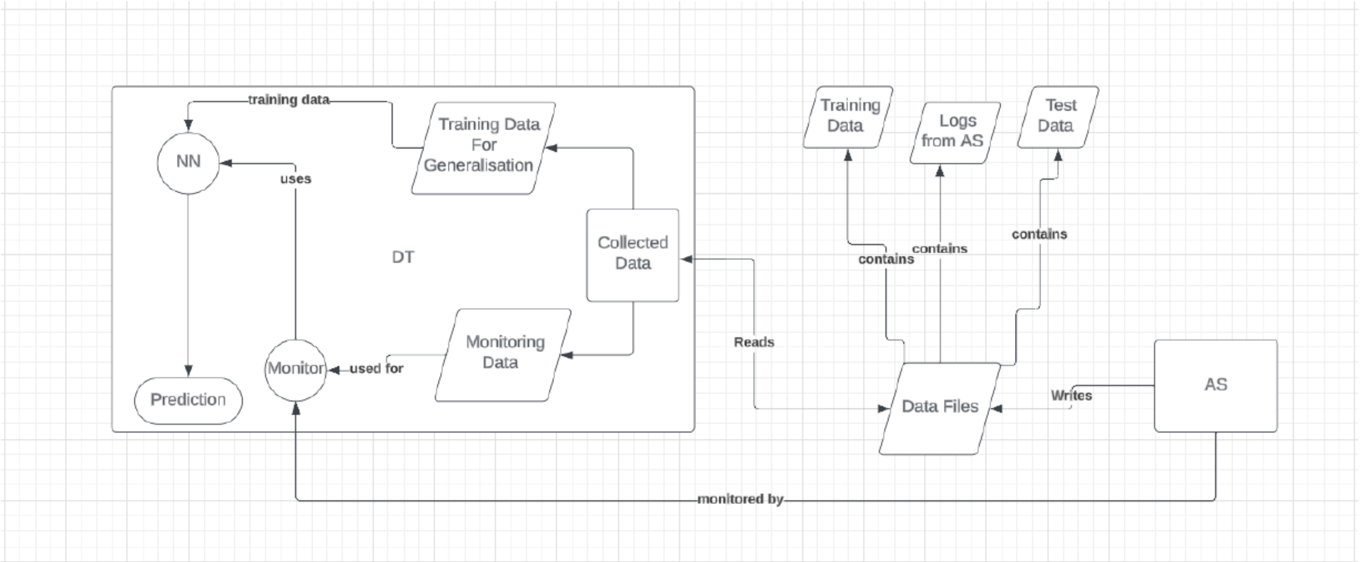
End result is a model that can rapidly produce a set of approximate results that can be shortlisted and explored by the full model for higher accuracy

RabbitMQ message broker



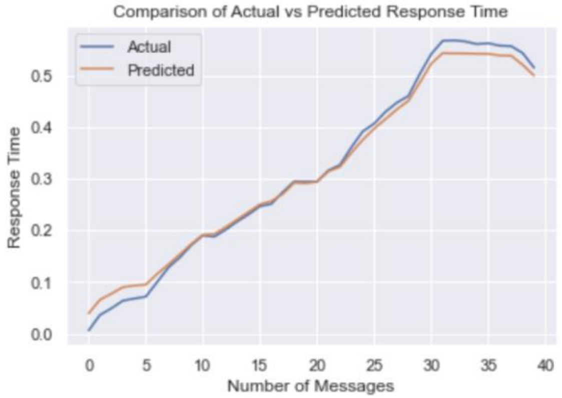
- Measured:
    - System Latency
    - Processing Time
    - Response Time
    - Message Delivery rate
    - Instances of Node Failure
- DT based on a NN model

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 10)                30

 dense_1 (Dense)             (None, 128)               1408

 dense_2 (Dense)             (None, 64)                8256

 dense_3 (Dense)             (None, 64)                4160

 dense_4 (Dense)             (None, 10)                650

 dense_5 (Dense)             (None, 1)                 11

=================================================================
Total params: 14,515
Trainable params: 14,515
Non-trainable params: 0
```

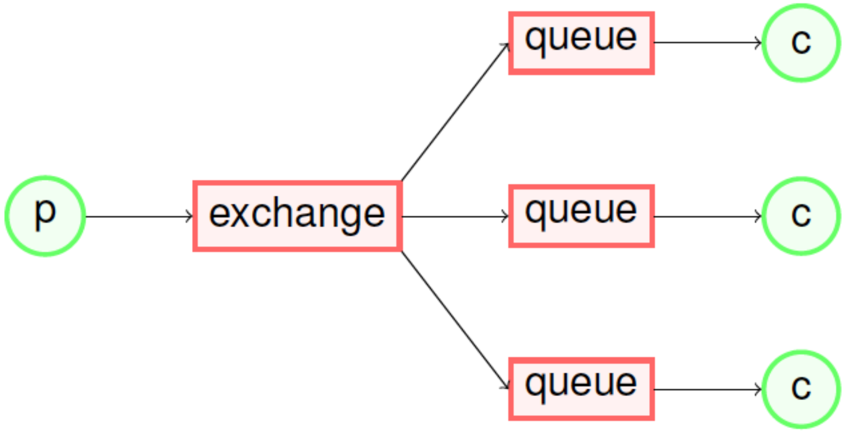


Trained on approx. 1200 instances

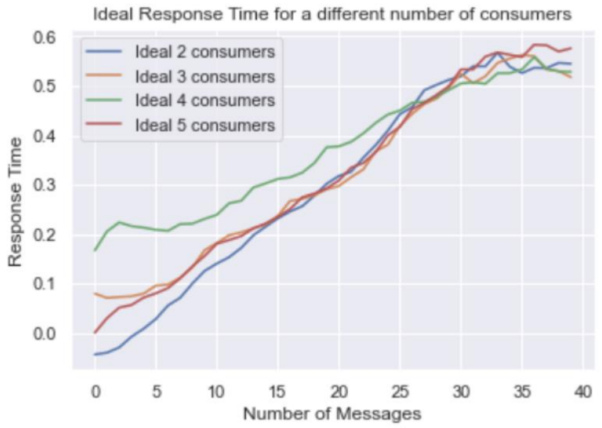
RabbitMQ message broker



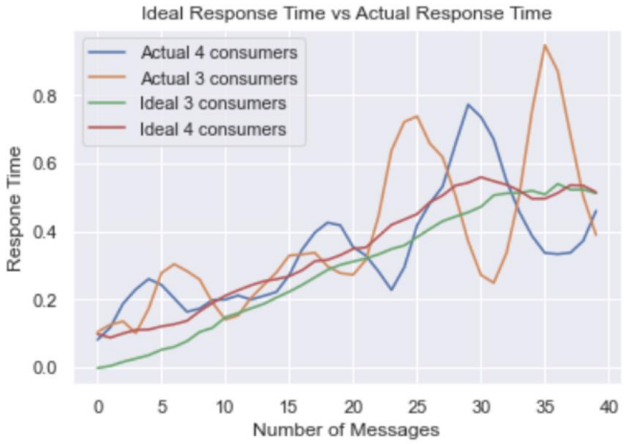
Good degree of accuracy when predicting response time



“Descriptive, predictive, and prescriptive”



Able to predict responses for unseen configurations

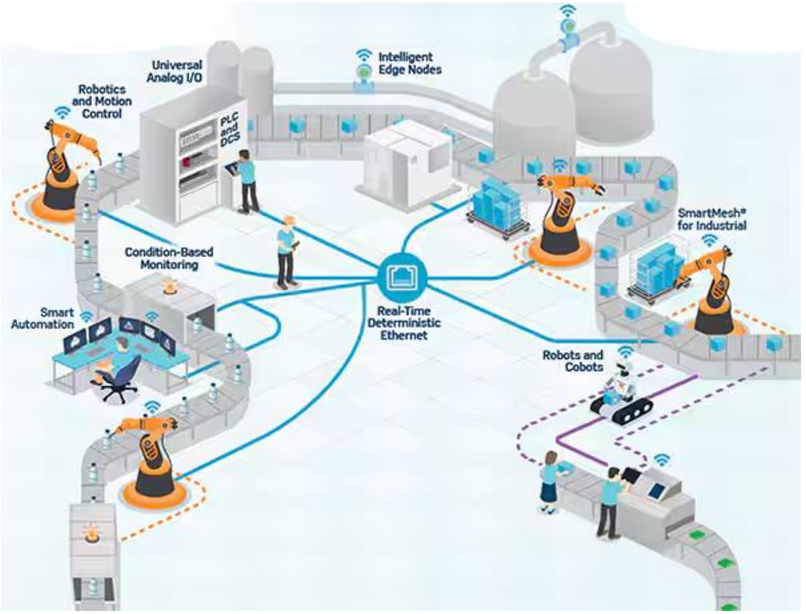


Also able to use variations from predictions to identify problems

```
Current value second consumer: 0.0024042129516601562
Current value first consumer: 0.002485990524291992
Current value (publisher): 4.0
#########################
the publisher has an issue!
#########################
main.py (PID: 2985) terminated successfully.
producer.py (PID: 2989) terminated successfully.
```

Challenges in building digital digital twins

- No obvious instrumentation points
  - What do we measure to best characterize the system?
- Generating data can be expensive
- How do we know…
  - What test scenarios to use
  - When we have enough data
- …to accurately characterise the system (response surface)?



https://feuernursingreview.teachable.com/p/feuer-nclex-rn-qbanl