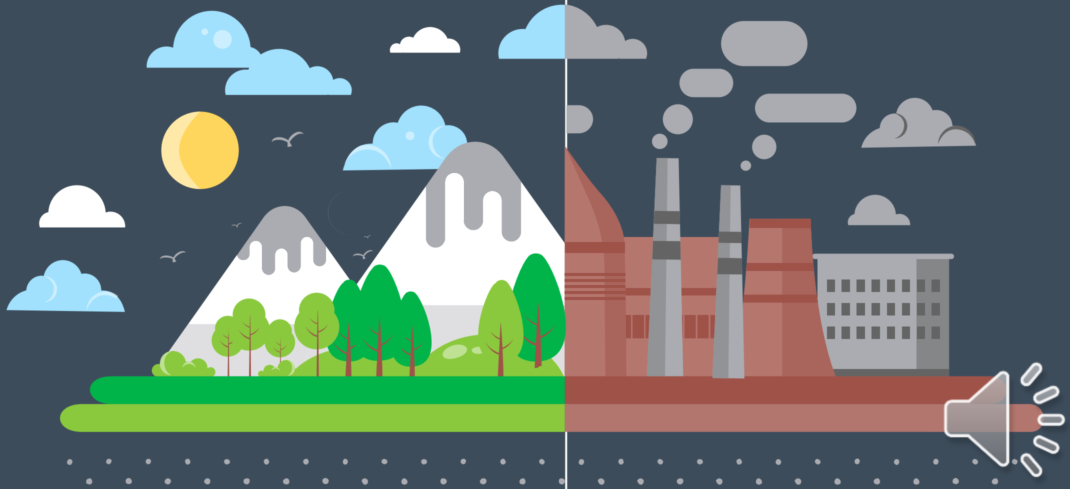
Using AI to Validate Carbon Containment in Illinois Basin

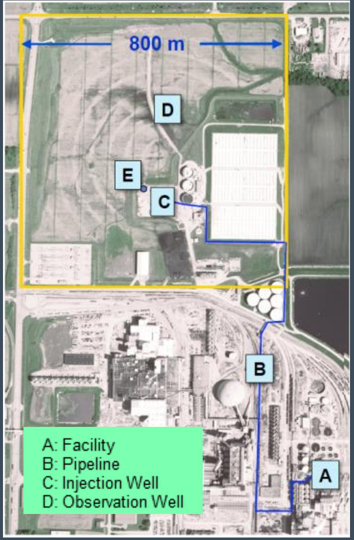
Machine Learning Challenge

Insights Inc

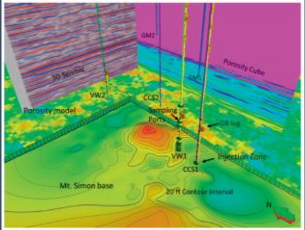


Background & Objective

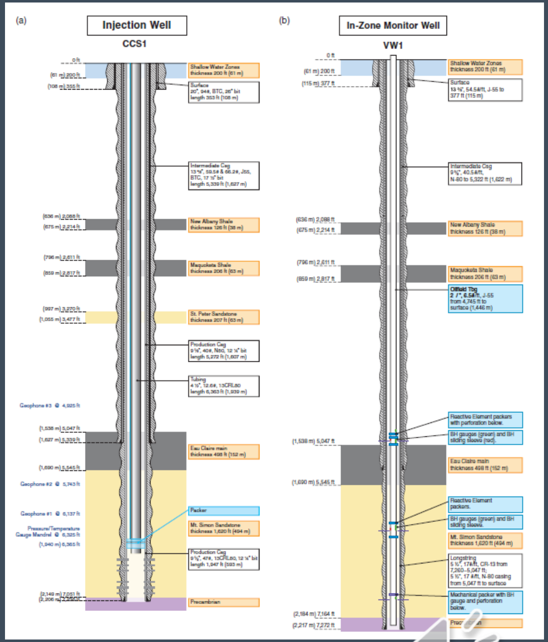
- Carbon Dioxide (CO_2) released from ethanol production at the ADM (Archer Daniels Midland) plant was compressed, dehydrated and injected into the Mt. Simon Sandstone
- Mt Simon Sandstone is a saline aquifer located in the Illinois Basin and is approximately ~7,000 ft deep
- Over three years (Nov-2011 to Nov-2014), ~1000 tonnes/d of CO_2 was injected into the reservoir
- Cumulatively, 999,215 tonnes of supercritical CO_2 was injected and has been geologically stored in Mt Simon Sandstone
- The injection and verification wells that are ~700ft apart were drilled downdip into the Mt Simon Sandstone
- The wells are equipped with downhole sensors to monitor pressure and temperature at various depths of interest
- The aim of this challenge is to use time series injection information and monitoring data from these wells to predict carbon capture well injection rate deltas



Source: modified from Locke, 2012



Source: Bauer et al, 2019

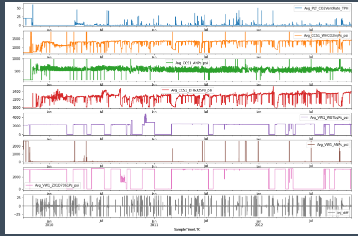


Source: Bauer et al, 2019

Methodology - Machine Learning

FieldID	WellID	Parameter	Units	Description	Depth ft
CCS1 ANPs psig	CCS1	pressure	psi	CCS1 annulus pressure in psi	Surface
CCS1 DHPs psig	CCS1	pressure	psi	CCS1 downhole pressure in psi	6325
CCS1 DHTp F	CCS1	temperatur e	F	CCS1 downhole temperature 0 in F	6325
CCS1_WHCO2InjPs_psi	CCS1	pressure	psi	CCS1 wellhead pressure in psi	Surface
CCS1_WHCO2InjTp_F	CCS1	temperatur e	F	CCS1 wellhead temperature in F	Surface
PLT CO2InjRate TPH	CCS1	rate	tonnes/hour	CO2 injection rate in tonnes per hour	Surface
PLT CO2VentRate TPH	CCS1	rate	tonnes/hou r	CO2 vent rate in tonnes per hour	Surface
VW1_ANPs_psig	WW1	pressure	psi	VW1 annulus pressure in psi	Surface
VW1 PTbgPs psig	VW1	pressure	psi	VW1 tubing pressure in psi	Surface
VW1 PTbgTp F	VW1	temperatur e	F	VW1 tubing temperature in F	Surface
Z 01D7061Ps psi	VW1	pressure	psi	VW1 Zone01 pressure (7061 ft) in psi	7061
VW1 Z01D7061Tp F	VW1	temperatur e	F	VW1 Zone01 temperature (7061 ft) in F	7061
VW1 Z02D6982Ps psi	VW1	pressure	psi	VW1 Zone02 pressure (6982 ft) in psi	6982
VW1 Z02D6982Tp F	VW1	temperatur e	F	VW1 Zone02 temperature (6982 ft) in F	6982
VW1 Z03D6945Ps psi	VW1	pressure	psi	VW1 Zone03 pressure (6945 ft) in psi	6945
VW1 Z03D6945Tp F	VW1	temperatur e	F	VW1 Zone03 temperature (6945 ft) in F	6945
VW1_Z04D6837Ps_psi	VW1	pressure	psi	VW1 Zone04 pressure (6837 ft) in psi	6837
VW1_204D6837Tp_F	VW1	temperatur e	F	VW1 Zone04 temperature (6837 ft) in F	6837
205D6720Ps psi	VW1	pressure	psi	VW1 Zone05 pressure (6720 ft) in psi	6720
VW1_205D6720Tp_F	VW1	temperatur e	F	VW1 Zone05 temperature (6720 ft) in F	6720
VW1 Z06D6632Ps p	VW1	pressure	psi	VW1 Zone06 pressure (6632 ft) in psi	6632
VW1 Z06D6632Tp F	VW1	temperatur e	F	VW1 Zone06 temperature (6632 ft) in F	6632
VW1 Z07D6416Ps psi	VW1	pressure	psi	VW1 Zone07 pressure (6416 ft) in psi	6416
VW1 207D6416Tp F	VW1	temperatur e	F	VW1 Zone07 temperature (6416 ft) in F	6416
VW1 Z08D5840Ps psi	VW1	pressure	psi	VW1 Zone08 pressure (5840 ft) in psi	5840
VW1 Z08D5840Tp F	WW1	temperatur e	F	VW1 Zone08 temperature (5840 ft) in F	5840
VW1_Z0910D5482Ps_ps	WW1	pressure	psi	VW1 Zone0910 pressure (5482 ft) in psi	5482
VW1 Z0910D5482Tp F	VW1	temperatur e	F	VW1 Zone0910 temperature (5482 ft) in F	5482
VW1_209D5653Ps_psi	VW1	pressure	psi	VW1 Zone09 pressure (5653 ft) in psi	5653
VW1 Z09D5653Tp F	VW1	temperatur e	F	VW1 Zone09 temperature (5653 ft) in F	5603
VW1_Z10D5001Ps_psi	VW1	pressure	psi	VW1 Zone10 pressure (5001 ft) in psi	5001
VW1_Z10D5001Tp_F	VW1	temperatur e	F	VW1 Zone10 temperature (5001 ft) in F	5001
VW1 211D4917Ps ; psi	VW1	pressure	psi	VW1 Zone11 pressure (4917 ft) in psi	4917
VW1_Z11D4917Tp_F	VW1	temperatur e	F	VW1 Zone11 temperature (4917 in F	4917

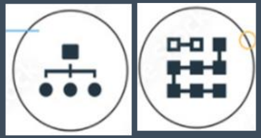
Data Collection



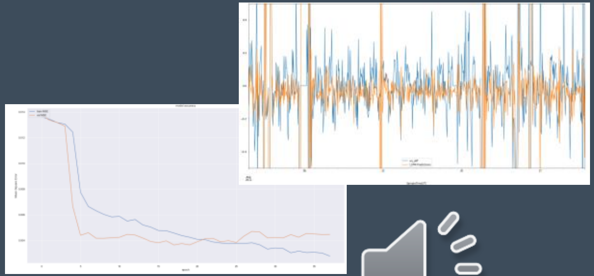
Data Preparation/Cleaning

```
# define model
model = Sequential()

# add LSTM layer with tanh activation. return_sequences=True is used so we can feed another
# LSTM layer to model. default is false.
model.add(LSTM(neurons, activation=activation, return_sequences=True, input_shape=(n_steps_in, n_features)))
model.add(Dropout(dropout))
# second layer of LSTM with tanh activation. Now return sequence is False, as there will be no other layer
# except output layer.
model.add(LSTM(neurons, activation=activation))
model.add(Dropout(dropout))

# final layer with output
model.add(Dense(n_steps_out))

# activation function of output layer
model.add(Activation('tanh'))

# compile model with loss funtion, optimizer and metrics
model.compile(loss='mse' , optimizer=opt , metrics=['mse'])

return model
```

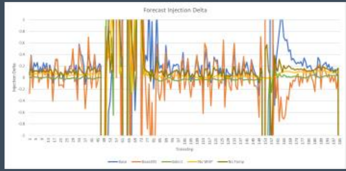
Model Selection - LSTM



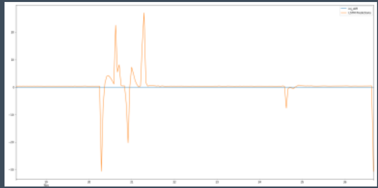
Model Training



Model Evaluation/Validation



Sensitivity runs on inputs and testing randomness

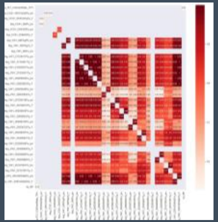


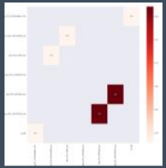
Retrain model on full dataset, deploy on blind data set using unseeded model

As new data is collected, it will appended to the entire dataset and the model will re-trained stochastically (unseeded) for future prediction

Model Deployment

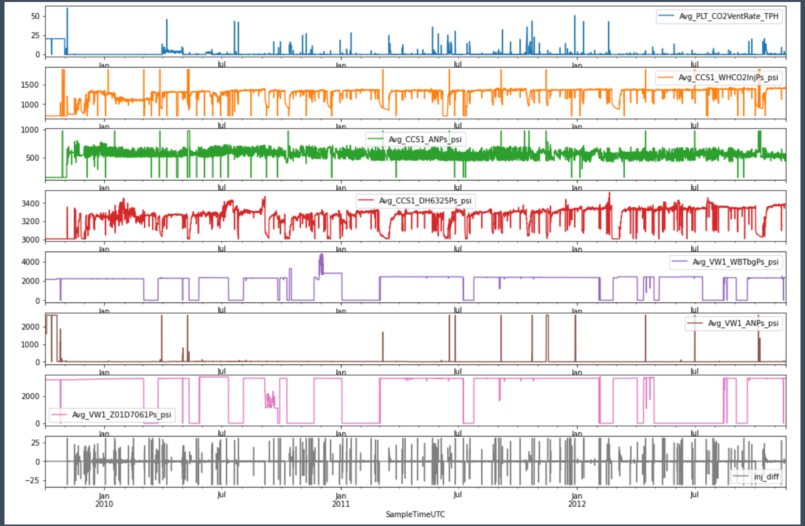
Data Preparation





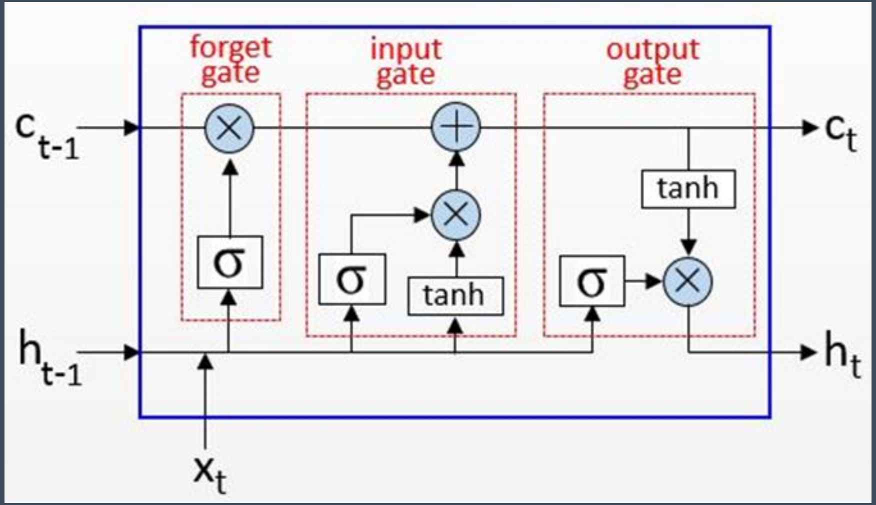
```
['Avg_PLT_CO2VentRate_TPH',
 'Avg_CCS1_WHCO2InjPs_psi',
 'Avg_CCS1_WHCO2InjTp_F',
 'Avg_CCS1_ANPs_psi',
 'Avg_CCS1_DH6325Ps_psi',
 'Avg_CCS1_DH6325Tp_F',
 'Avg_VW1_WBTbgPs_psi',
 'Avg_VW1_Z05D6720Ps_psi',
 'inj_diff'
]
```

- Most variables are highly correlatable, with values of > 0.5
- Retained only one variable from groups of highly correleatable variables.
- Exceptions are made if variables are found from different sources e.g. tubing and reservoir pressure at the observation well
- Excluding the target injection delta variable, six variables retained for the machine learning application.



- A modified Z-score is used to remove the outliers.
- A more lenient score is used for injection delta and CO2 vent rate due to the difference in the magnitude's range
- The zero values at the observation well sensors have been kept – this will be tested for removal during the sensitization stage

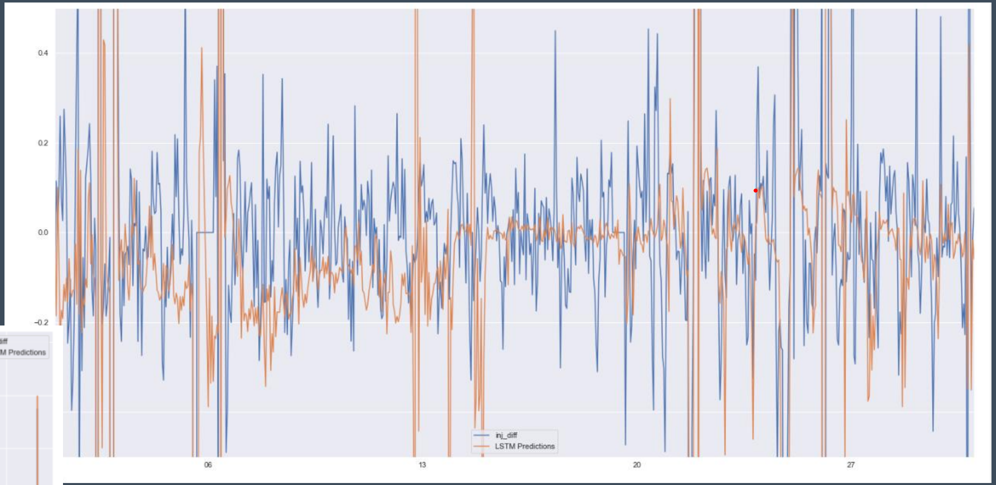
Long Short-Term Memory (LSTM) & Randomness



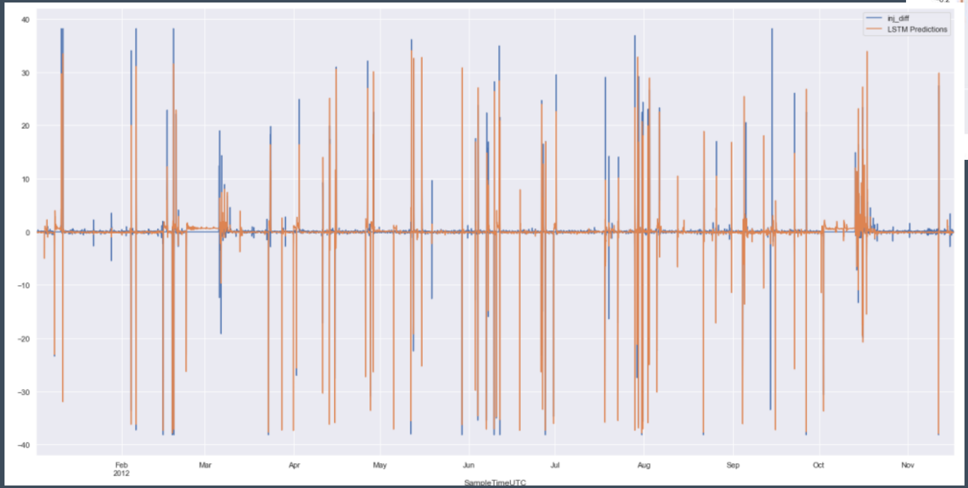
Note. By Rainardi, V., 2021, RNN and LSTM. Accessed at https://dwbi1.wordpress.com/2021/08/07/recurrent-neural-network-rnn-and-lstm/

- LSTM is a type of recurrent neural network (RNN)
- Learns from previous outputs to provide better results the following time
- Typical LSTM has 3 layers
  - Input gate which assigns weights based on the significance of different variables,
  - Forget gate to retain only useful information, and
  - Output gate which manages the information flow
- Randomness
  - Machine learning models are stochastic in nature
  - Randomness is present in the LSTM network employed
- Sensitivity runs on input parameters
  - Randomness controlled using a seed value – so that we reproduce the result if necessary
  - To see effect of input parameter
- Model Deployment and Prediction
  - Deployed without a seed to make use of the randomness –
  - if the model is run again, it will not be able to re-produce the exact same result
  - Multiple runs may be different, but the results will be in range
  - the anomalies will be predicted
  - objective of the model met.

Model Validation



Able to match anomalies in injection delta



Able to mimic the small variations in the injection delta

Sensitivity Runs

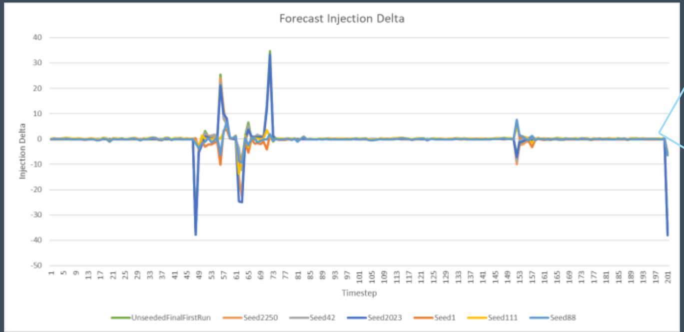
- Sensitivity runs on input parameters were performed using a single seed value of 2250. Several sensitivities were run but only selected are displayed here
- Best results (based on RMSE)
  - Scaling between -1 and 1
  - Use of DH sensor Z05 data with corrected values of 0
  - Use of temperature and whp sensors data
  - Z-score outlier replacement of 25, resulting inj_diff range between 30 and -30.

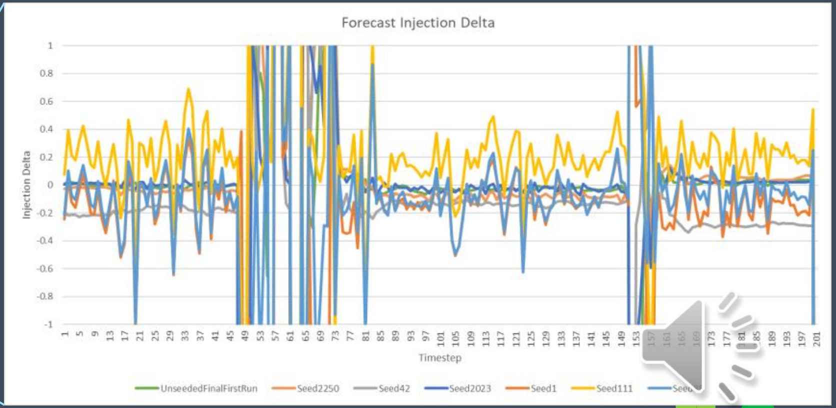
Vintage	**Select**	Base (0)	4	12	9
Scaling	**Y(-1,1)**	Y(0,1)	Y(0,1)	Y(-1,1)	Y(-1,1)
VW DH Sensor	**Z05**	Z09	Z01	Z05	Z05
Injection WHP Sensor	***Y***	Y	Y	Y	N
Temp Sensor	***Y***	N	N	N	Y
Z-Score Inj_Diff	**25**	20	25	25	25
VW Zero Values	**No**	Yes	No	No	No
Val RMSE	**1.55**	1.75	2.20	1.90	2.01
Val R2	**0.82**	0.73	0.63	0.72	0.69

Randomness

- Running with different seed values result in different results due to the stochastic nature of the model.
- 10 runs with various seed values were run and a selected six are displayed here to show the range in comparison to the first unseeded run
- All runs can predict the anomalies however the signature of the minor differences varies
- The RMSE and R2 values on the validation set is within a small range
- Thus, to include randomness and to be unbiased, an unseeded model is selected for deployment

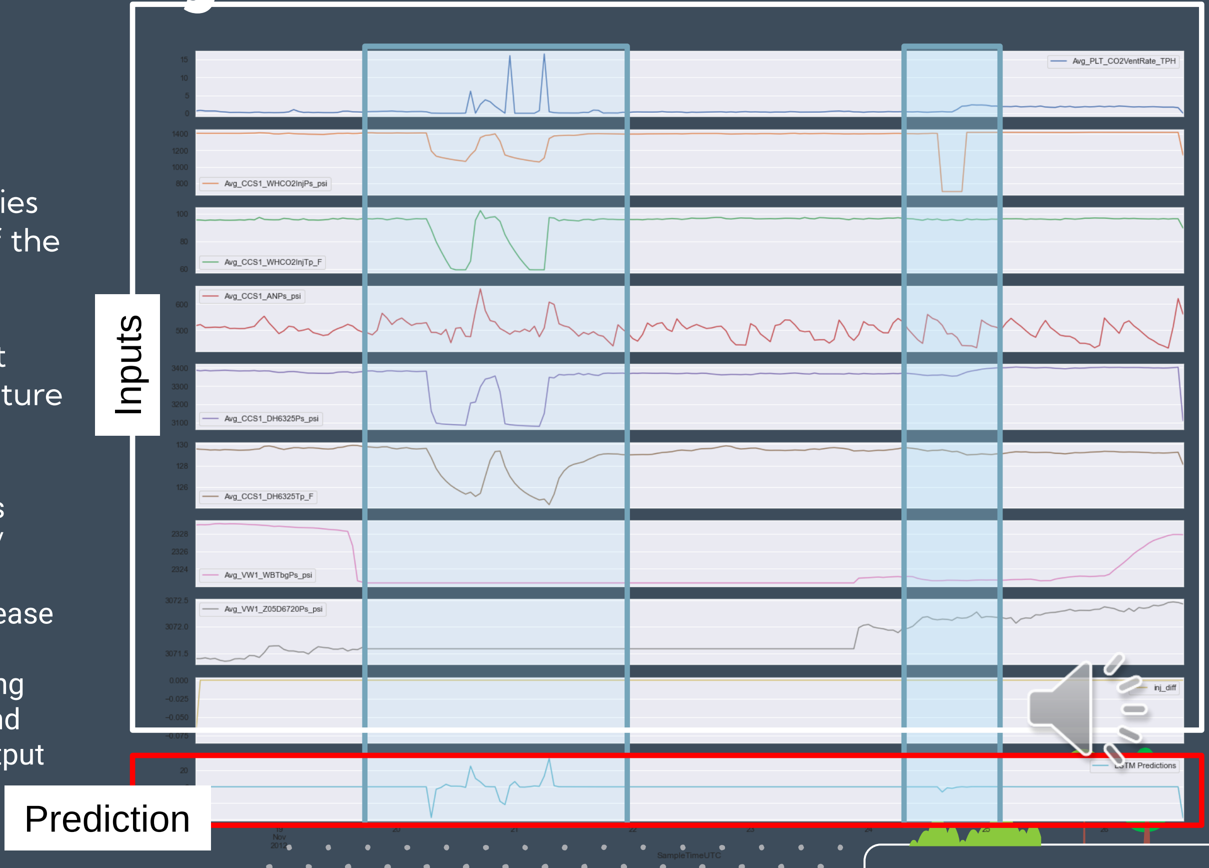
Seed Value	Unseeded	2250	42	2023	1	111	88
Val RMSE	1.73	1.55	1.69	1.69	1.64	1.87	1.70
Val R2	0.77	0.82	0.78	0.78	0.79	0.73	0.78





Predicted Injection Difference

- Predicted models does well based on good RMSE and R2 scores on test and train data
- Able to pick out large anomalies thus meeting the objective of the challenge
- Also able to predict small variation in rate though might have a slightly different signature
- Future recommendations for improvement:
    - Explore alternative models
    - Improved outlier removal / nan/zero replacements
    - Improve workflow to decrease runtime which will allow multiple set of runs creating an ensemble of models and taking the mean of the output



Thank you all!

Happy to share more about our algo and journey, feel free to reach out to us

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution