

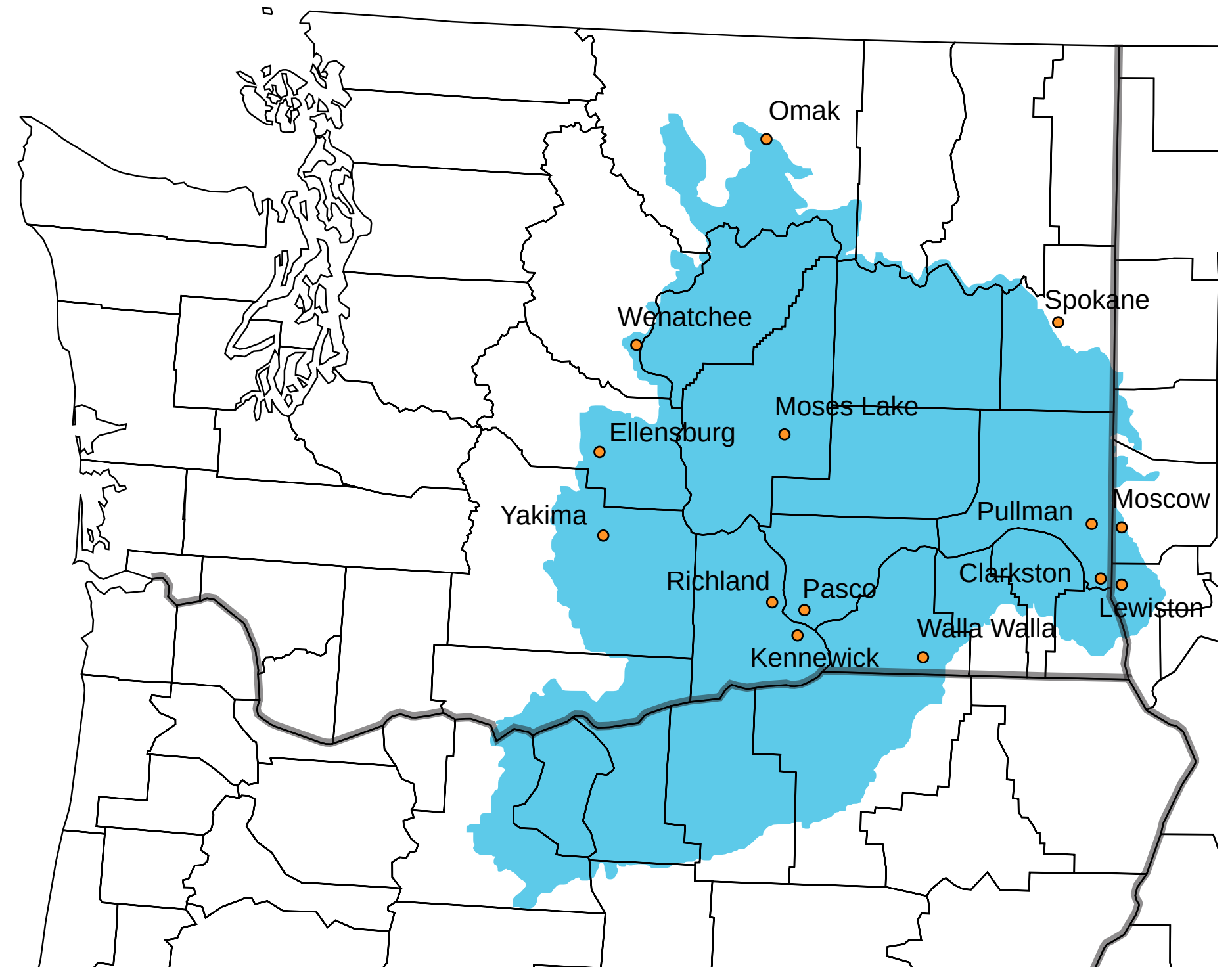


Predicting Groundwater Levels

Marcos Ortiz, Meredith Sargent, Riti Bahl, Chelsea Gary, and Anireju Emmanuel Dudun

Overview

- 1 Over the last 100 years, water abstraction volume has increased from 500 to $\sim 4000 \text{ km}^3/\text{yr}$ due to population growth, economic development, and rapid urbanization around the world.
- 2 More specifically, the Columbia Basin lies in an arid lowland area, making groundwater level prediction crucial to assist with water supply monitoring.



Why Spokane?

Spokane county boasts the second-largest number of **farms** in Washington state, with a total of 2,425 in 2017. Additionally, the availability of **consistent data** over several years from a cluster of wells makes Spokane an ideal location for modeling purposes. The area's **excellent weather and river data** further contribute to its appeal.



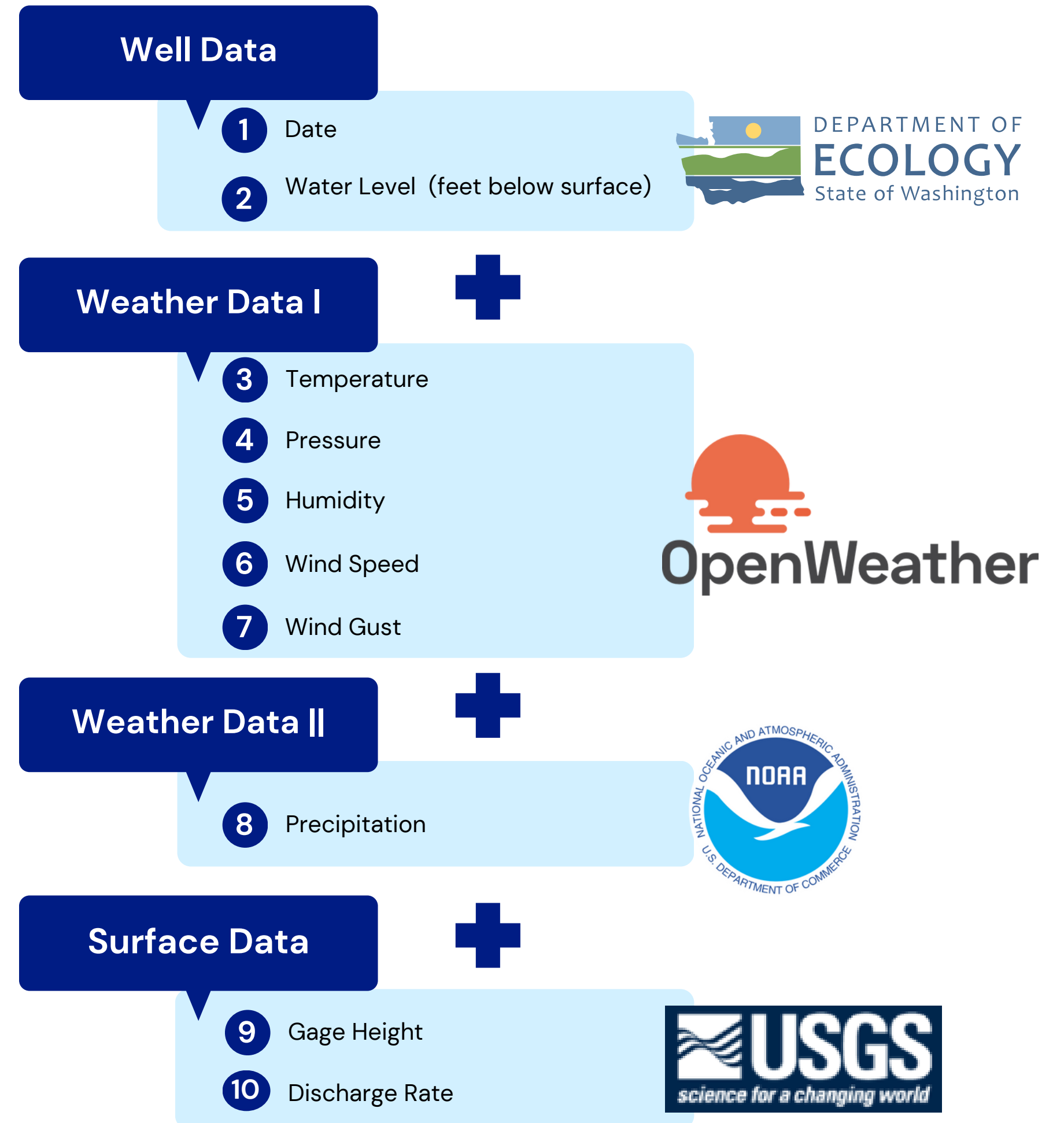
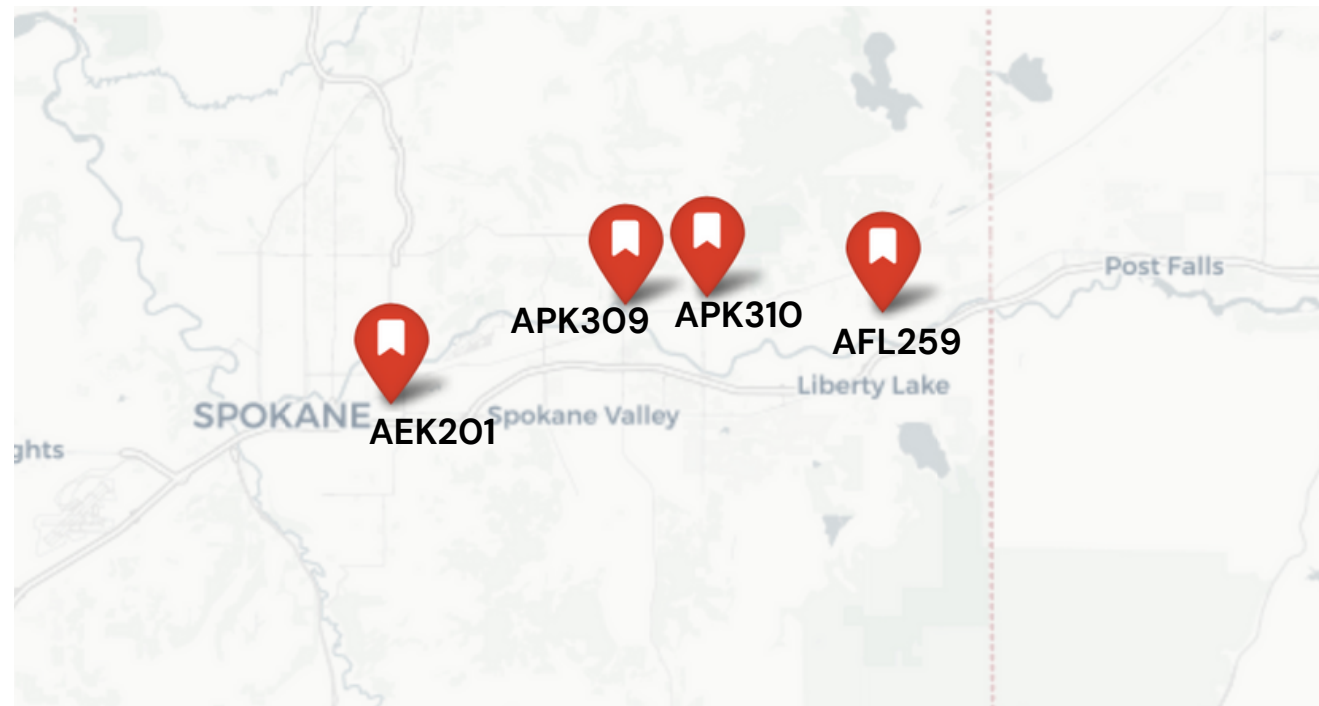
Stakeholders

- 1 Spokane residents.
- 2 Spokane businesses.
- 3 Local government agencies.

Datasets

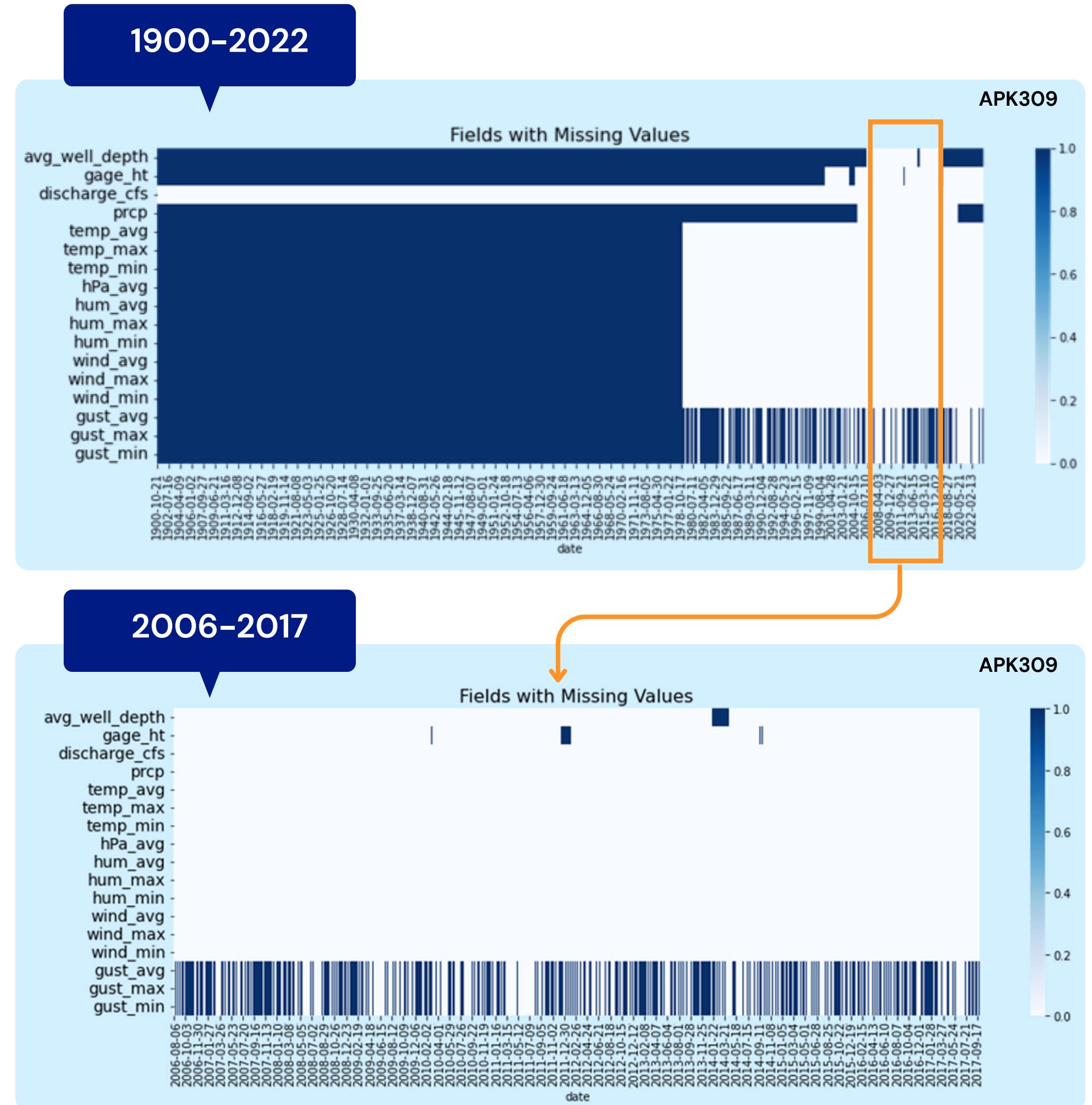
We selected **four** groundwater monitoring wells in the Spokane area with differing lithographies.

Data was gathered from various government and commercial sources.



Dataset Processing

- 1 Data prior to 2006 was dropped due to missing values.
- 2 Wind Gust & Gage Height missing data between 2006 – 2017 were replaced with zero and last non-missing values respectively.
- 3 Key Features includes: **date**, **gage_ht**, **discharge_cfs**, **prcp**, **temp_avg**, **hum_avg**, **wind_avg** and **gust_avg**
- 4 Engineered Precipitation data with 45 days lag for better correlation with target feature: Water level



Models

1 Baseline

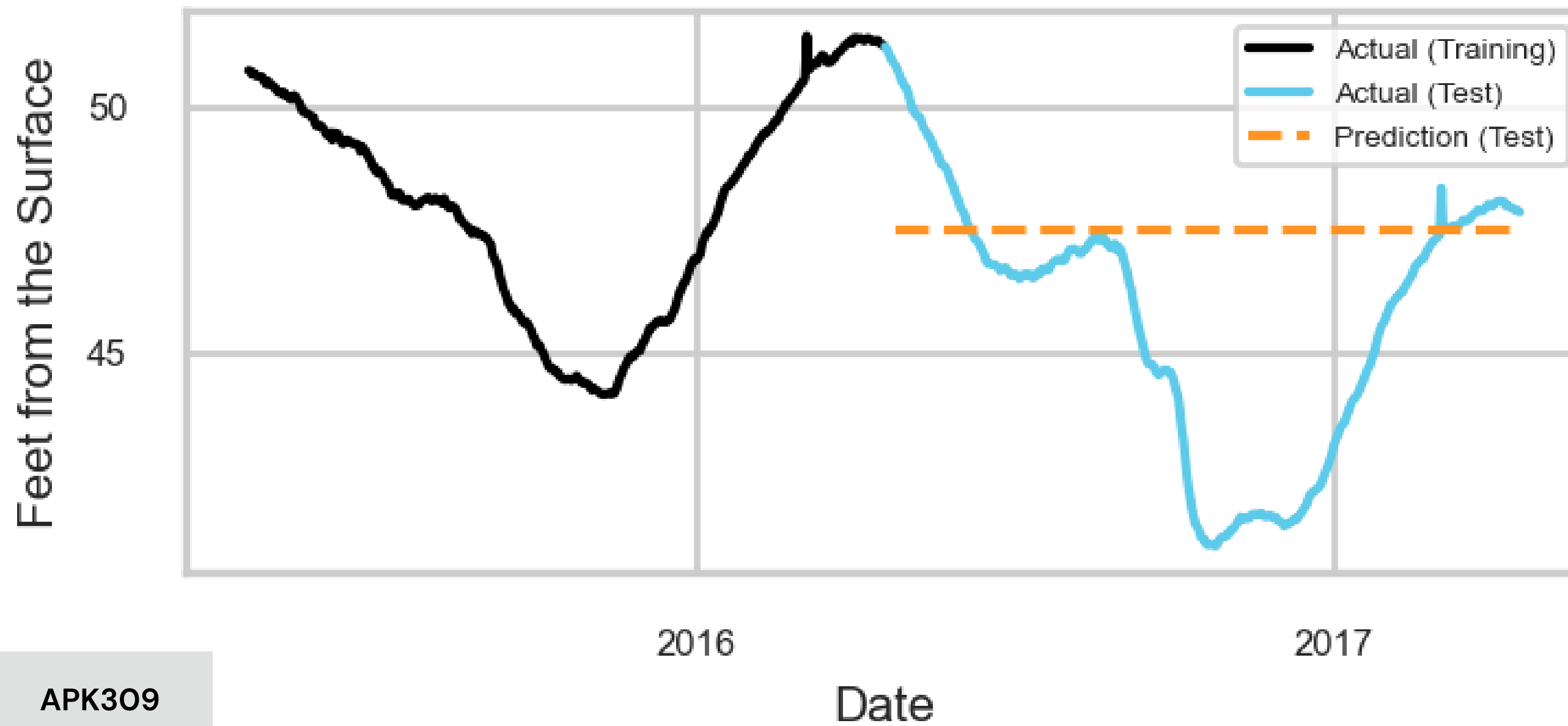
2 Linear Regression

3 Convolutional Neural Network

4 Recurrent Neural Network
(Long Short Term Memory)

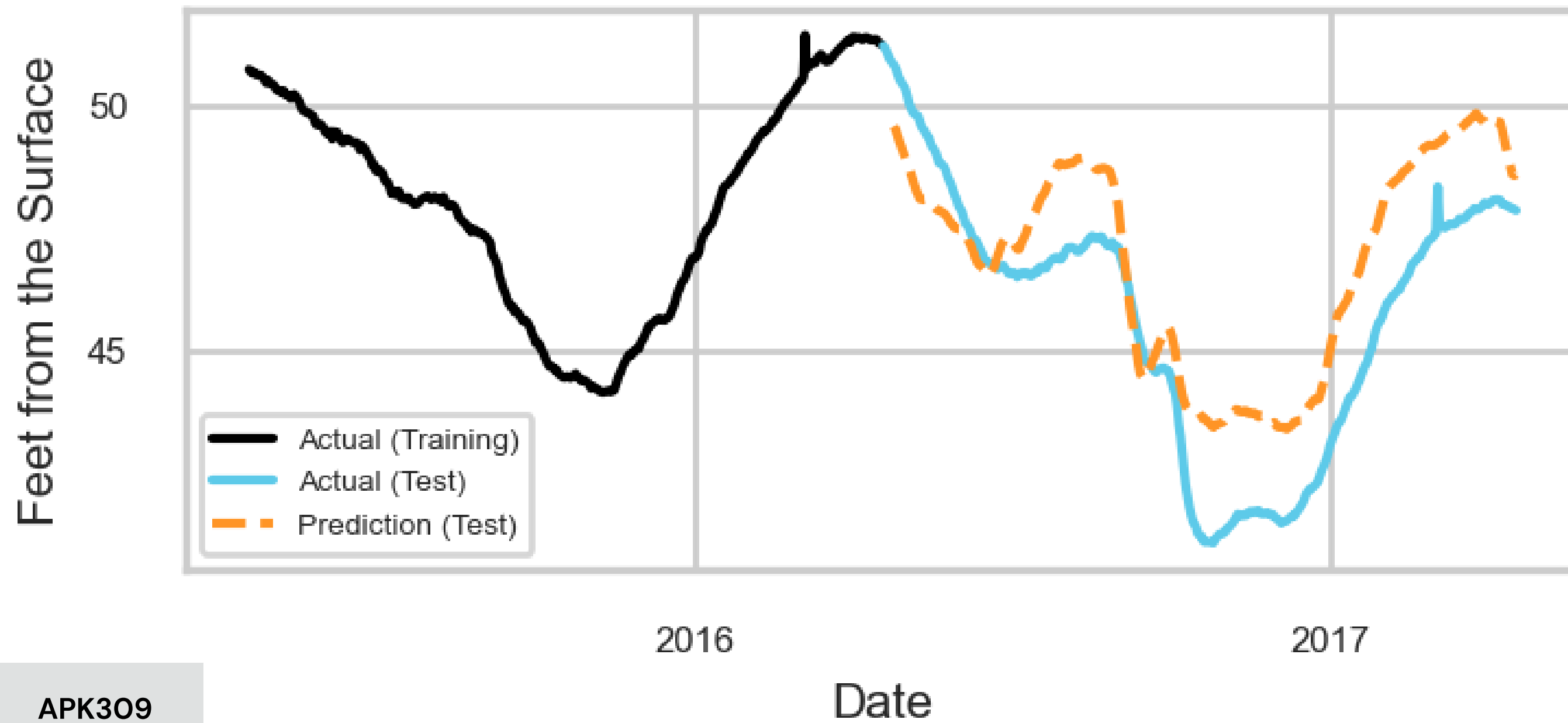
Wrapped in custom Scikit-Learn estimators/transformers

Baseline



APK309

Linear Regression

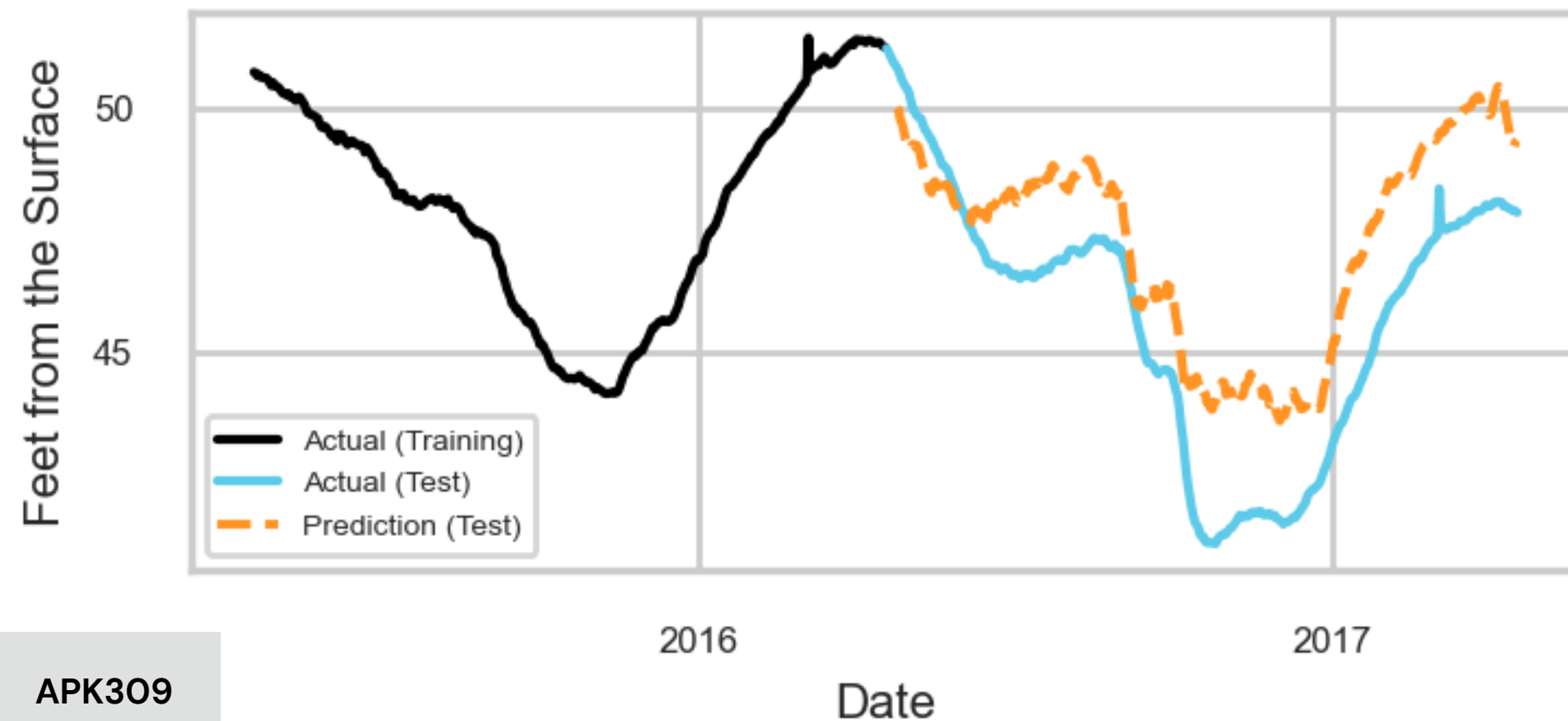


APK309

CNN

Hyperparameters were tuned using grid search to enhance performance.

- Number of convolutional layers.
- Kernel and filter sizes.
- Number of dense layers.
- Units in each dense layer.
- Batch size
- Learning rate
- Early stopping threshold.



Model Structure

Input

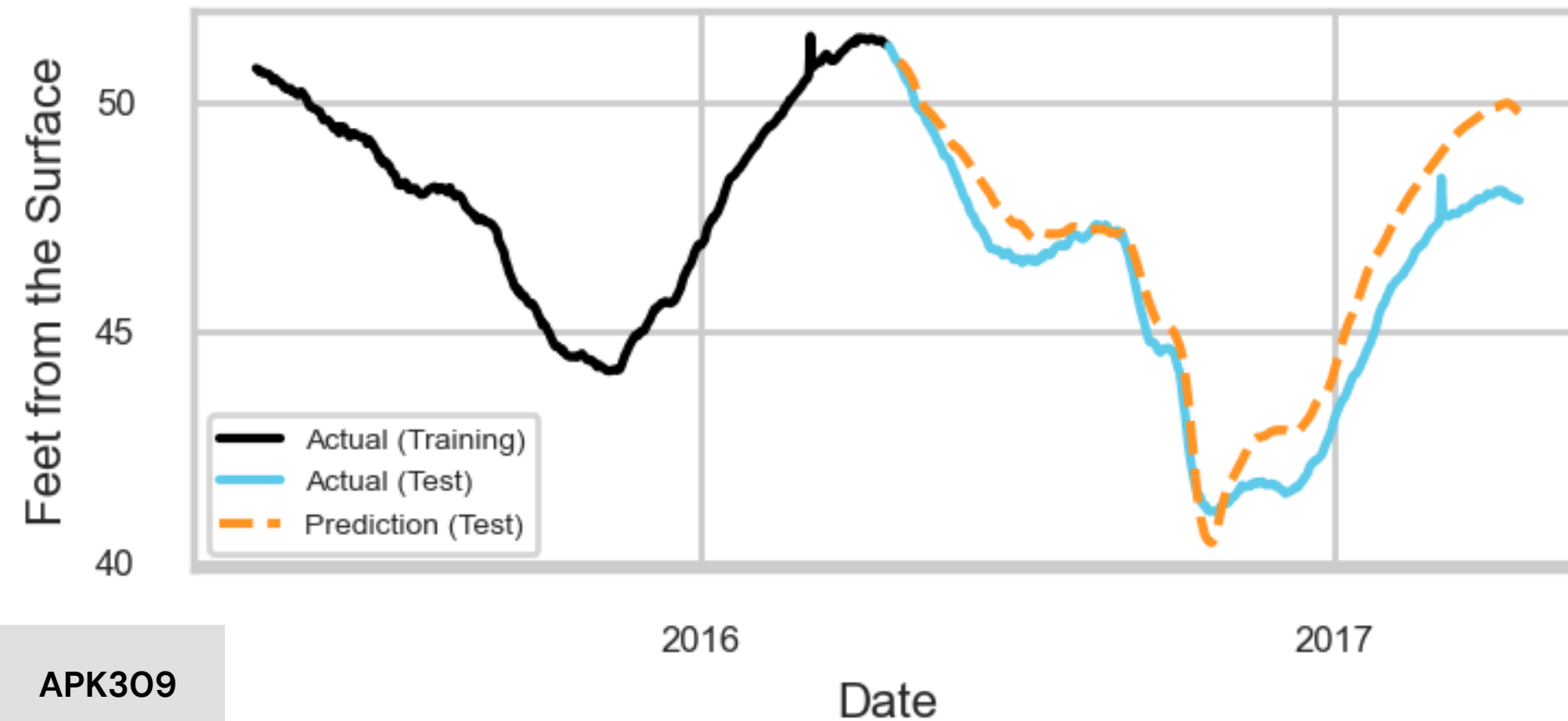
Convolutional

Flattening

Dense

Output

LSTM



Hyperparameters were tuned using grid search to enhance performance.

- Window size
- Units in the LSTM layer and each dense layer
- Number of dense layers
- Learning rate

Model Structure

Input

Long Short Term Memory

Flattening

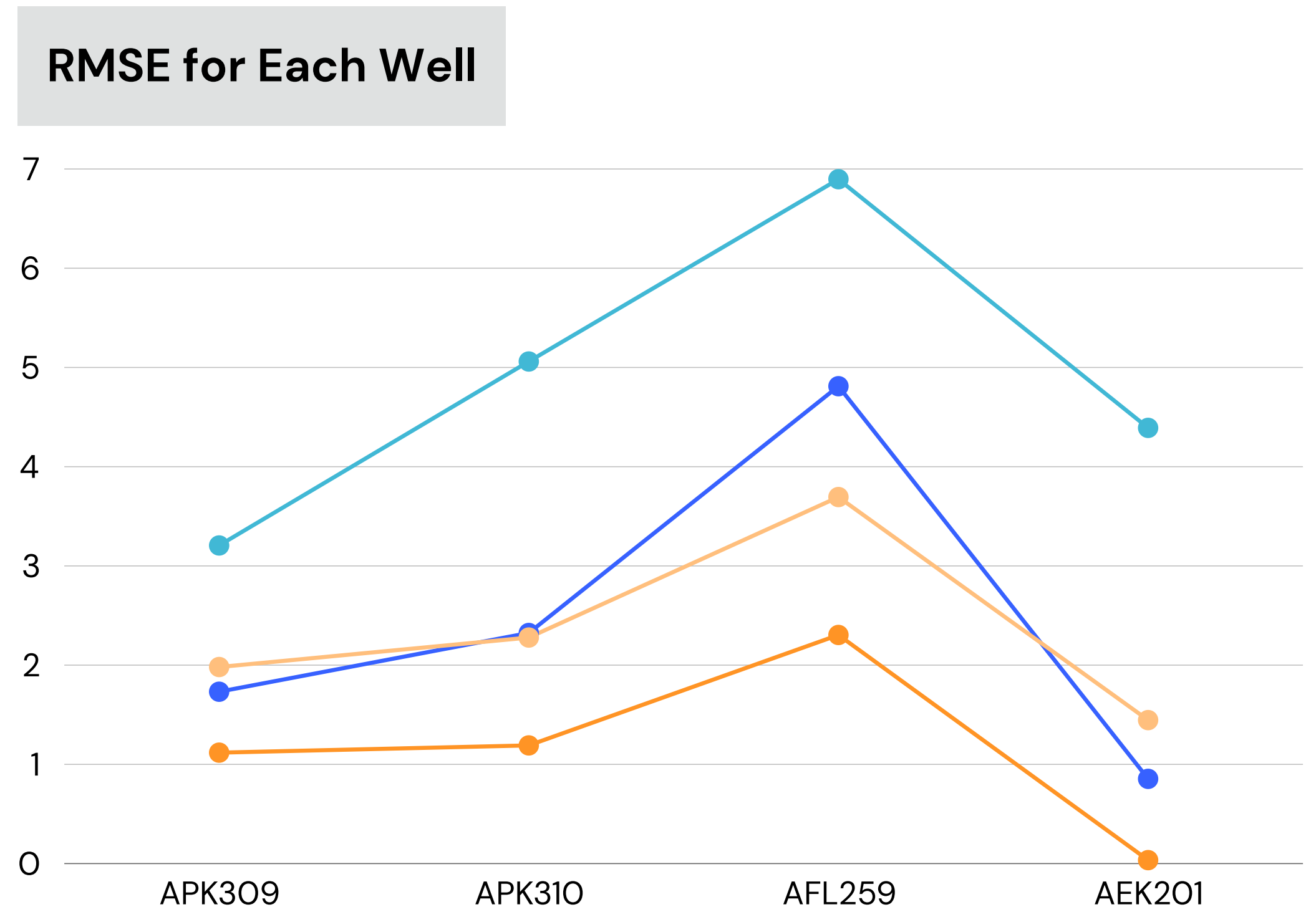
Dense

Output

Results

We held out the last year of groundwater level data for each of our wells. This is how our models performed on that holdout set. (Lower is better.)

	APK309	
	RMSE	MAE
Baseline	3.206	2.372
Linear Regression	1.732	1.587
CNN	1.981	1.756
LSTM	1.118	0.914

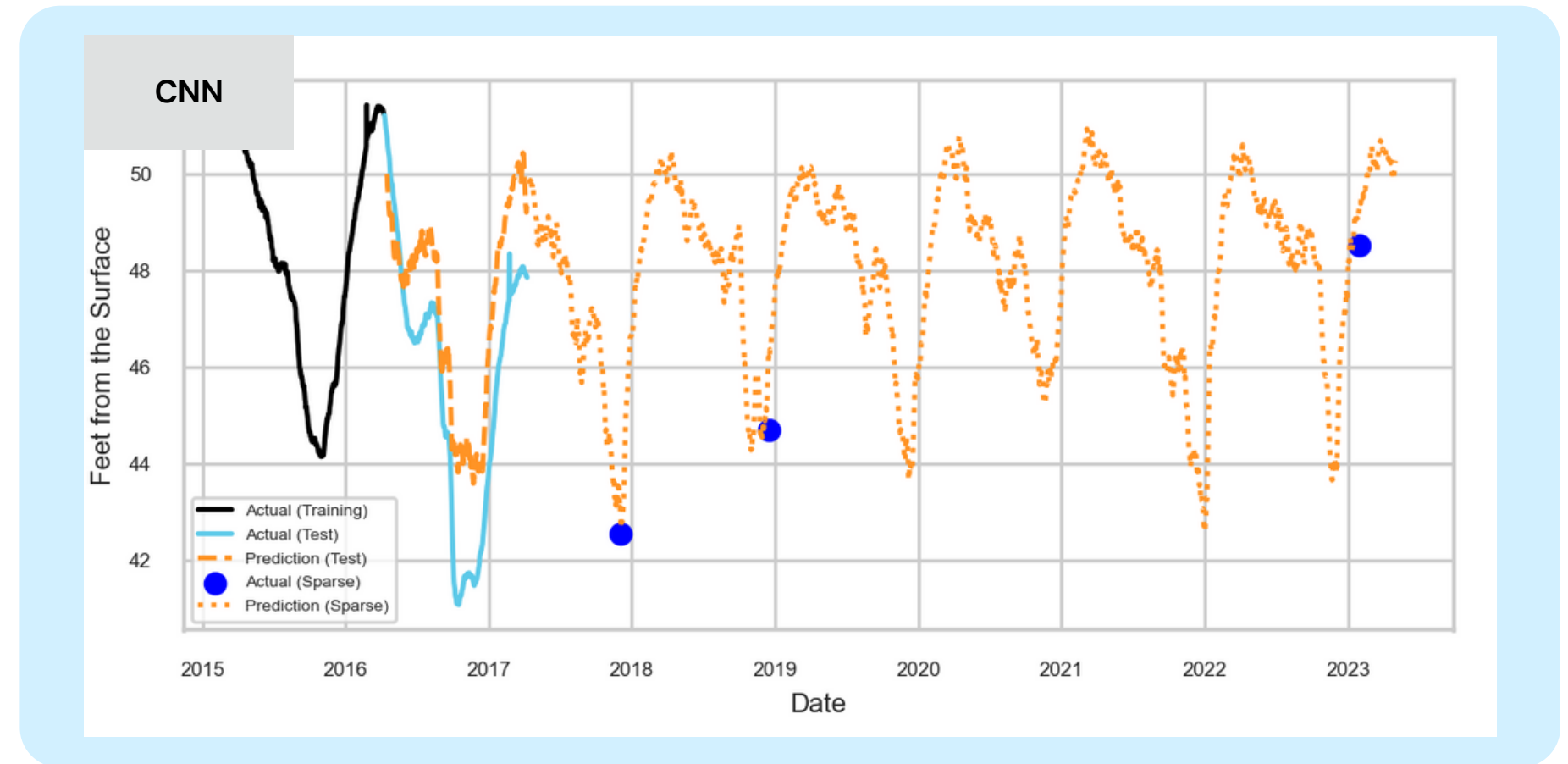


Models

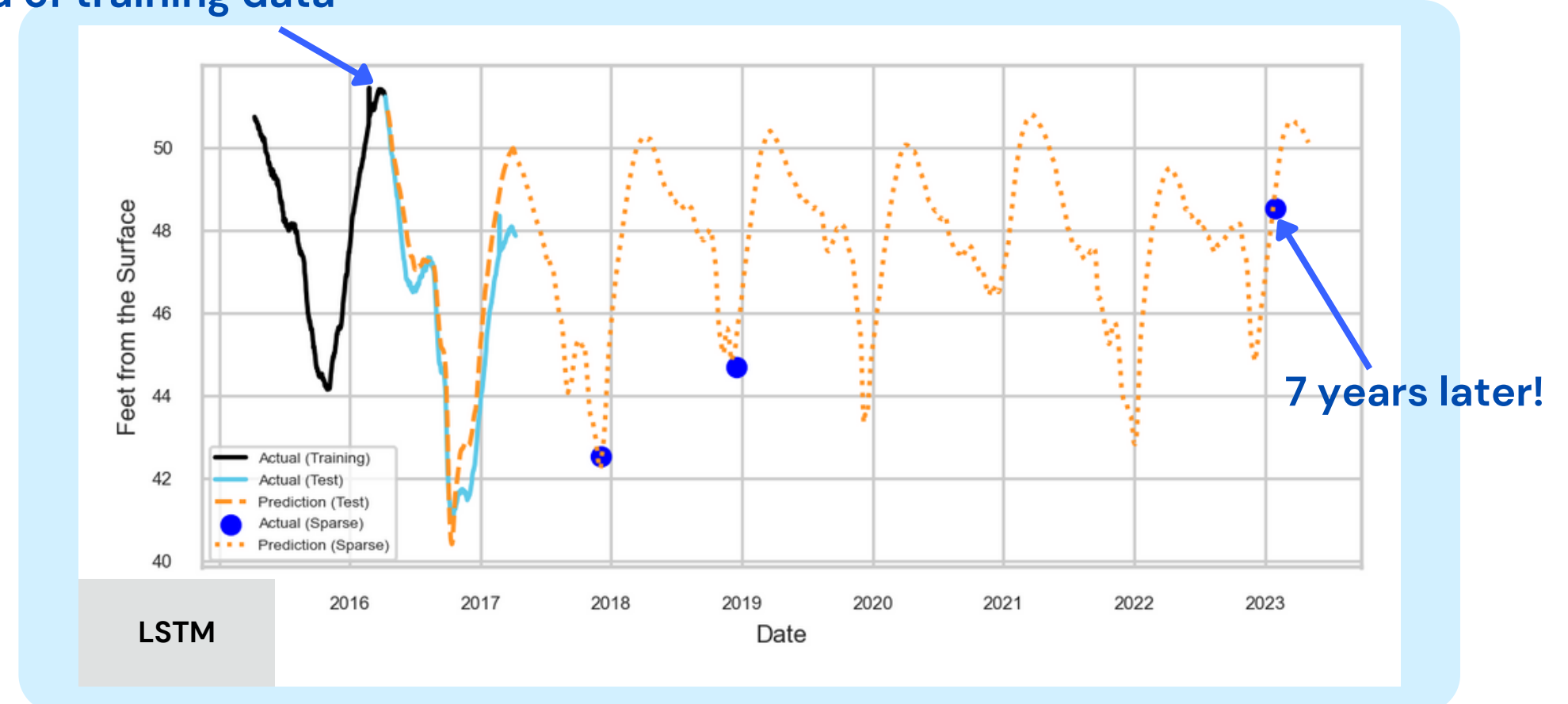
Sparse Data

The availability of weather and river data allows us to make predictions beyond the continuous well data, and we can evaluate them on a limited number of data points.

APK309	Actual	CNN	LSTM
7-20-23	48.53	49.47	49.13



End of training data



Streamlit App

We integrated our models into Streamlit to allow users interactive insights across four selected wells.



Future Work

- 1 Improve the web app
- 2 Wells at other locations
- 3 Additional parameter tuning
- 4 Seasonal ARIMA

Acknowledgements

Thank you to Roman Holowsky, Matt Osborne, Alec Clott and the Erdős Institute for their support throughout the Fall 2023 boot camp.

Thank you Gleb Zhelezov for his mentorship throughout the project.

