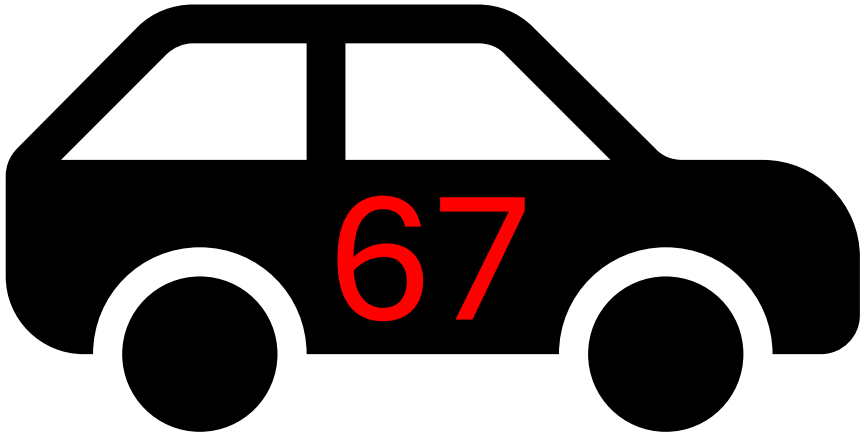
Erdos Data Science Bootcamp



Texas A&M Transportation Institute (TTI)

[Traffic congestion hits record high, but travel patterns are changing, report finds – Texas A&M Stories]

07 January 2026

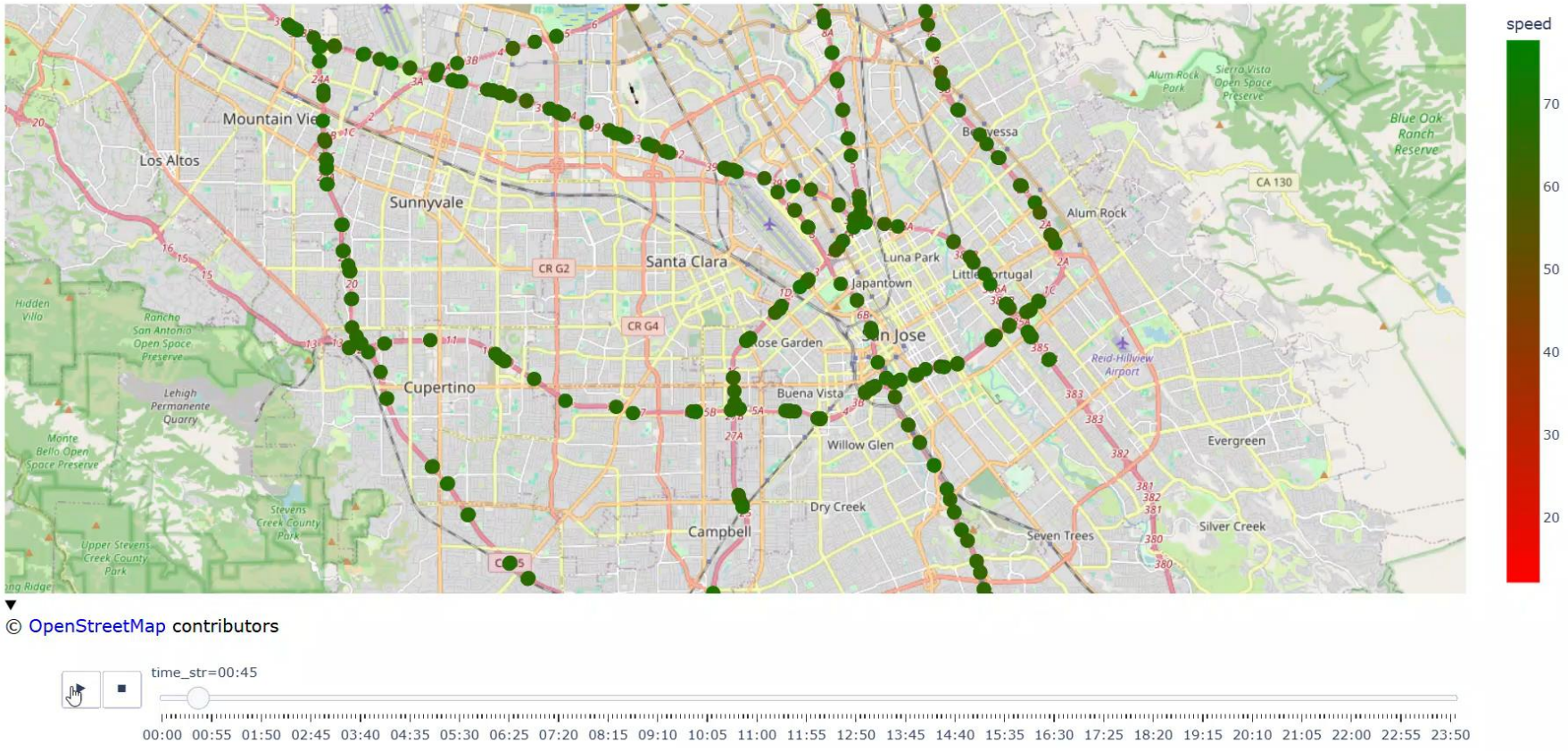
Slava Ivanov

Ifunanya Nwogbaga

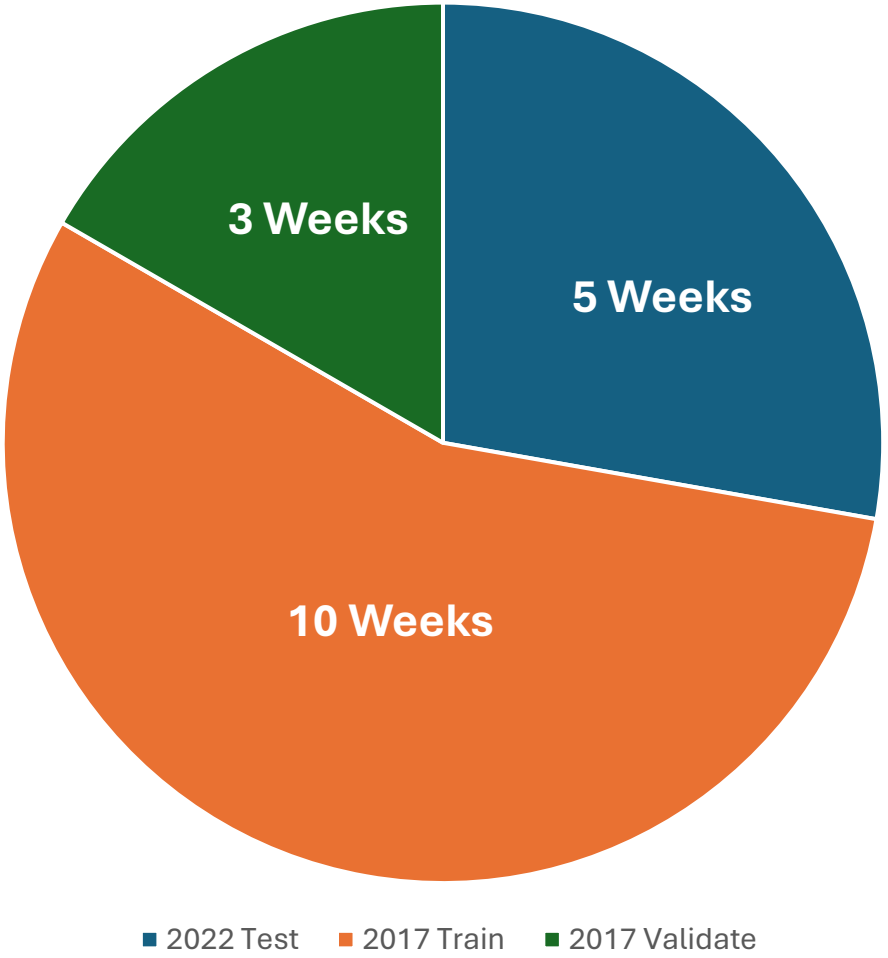
Can we predict the average speed of traffic?

Data Set

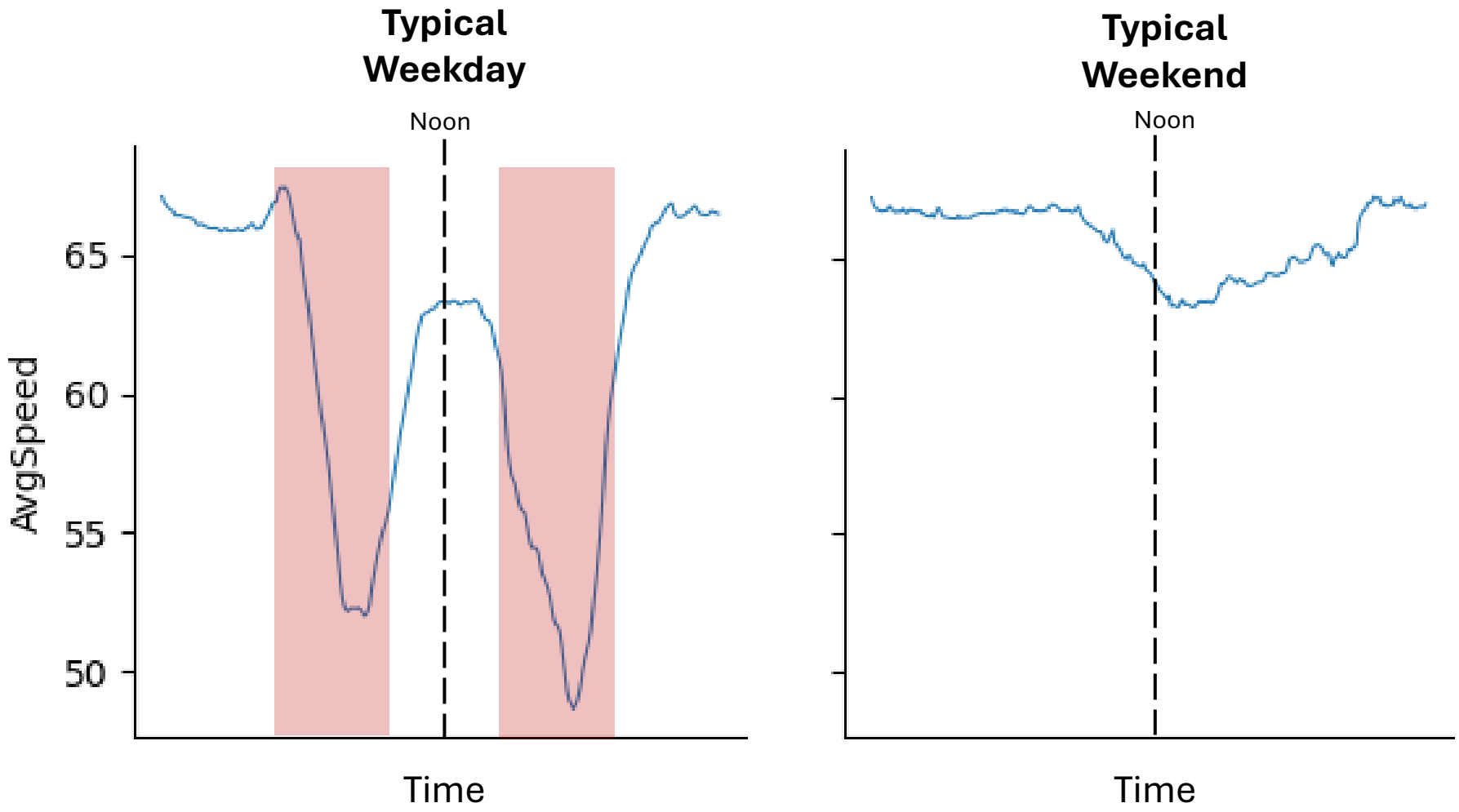
- Traffic data from 2017 and 2022 in the Bay area highways and streets
- 300 sensors are placed at different locations and track the speed of cars every five minutes
- Had to account for day light savings



Train Test Validate Split



Traffic Patterns Throughout the Week



Models

Naive

Forecasts every future value as the last observed value. Good for random-walks or when the last value is the best guess.

Not good at learning structure. It struggles when there is a trend, seasonality, or mean reversion.

ARIMA

Auto Regressive Integrated Moving Average predicts using a linear combination of past values plus recent forecast errors to capture autocorrelation. Good for stationary series with clear autocorrelation patterns.

Not good for nonlinear dynamics or when strong trends do not exist.

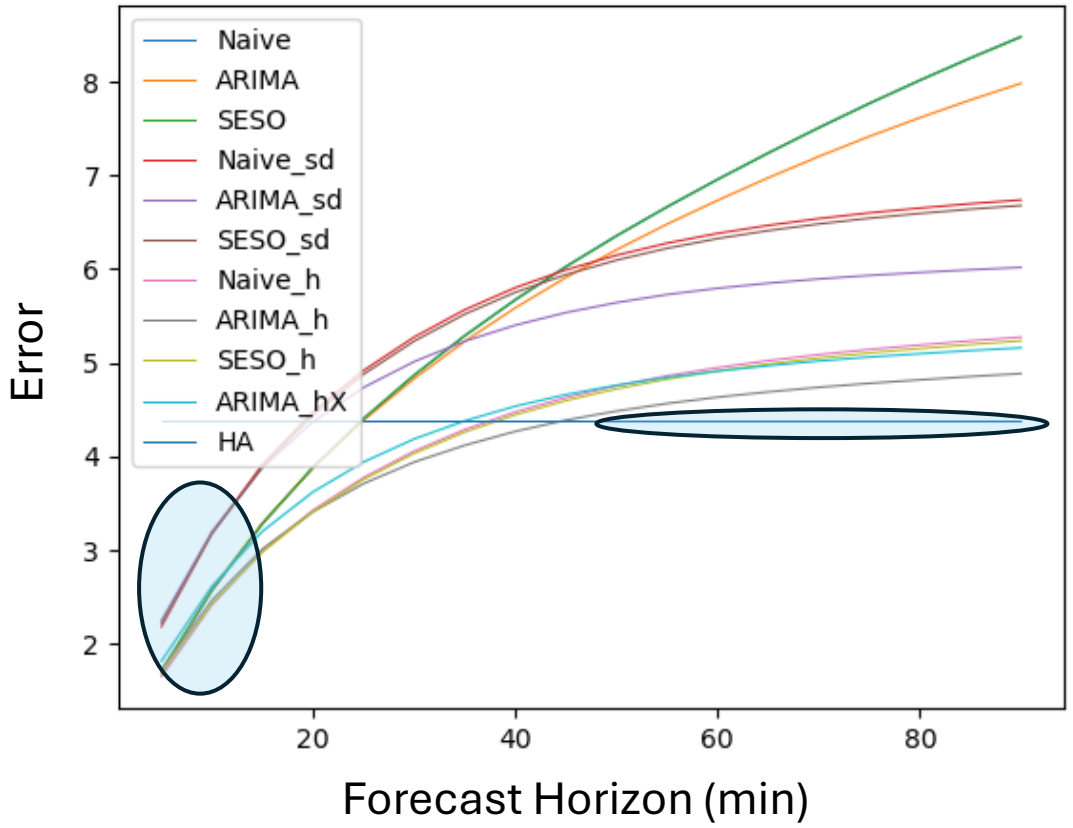
SESO

Simple Exponential Smoothing Optimized forecasts future values according to a smooth level, in which it controls the rate of smoothing. Great for series with no trends or seasonality but a slowly drifting level (adaptive average).

Not good for strong trends or structural changes or shifts.

We train the models to the base data set, seasonal differences, and deviance from the historical average

Naïve model performs just as well for short times, but the historical average becomes the best at long times

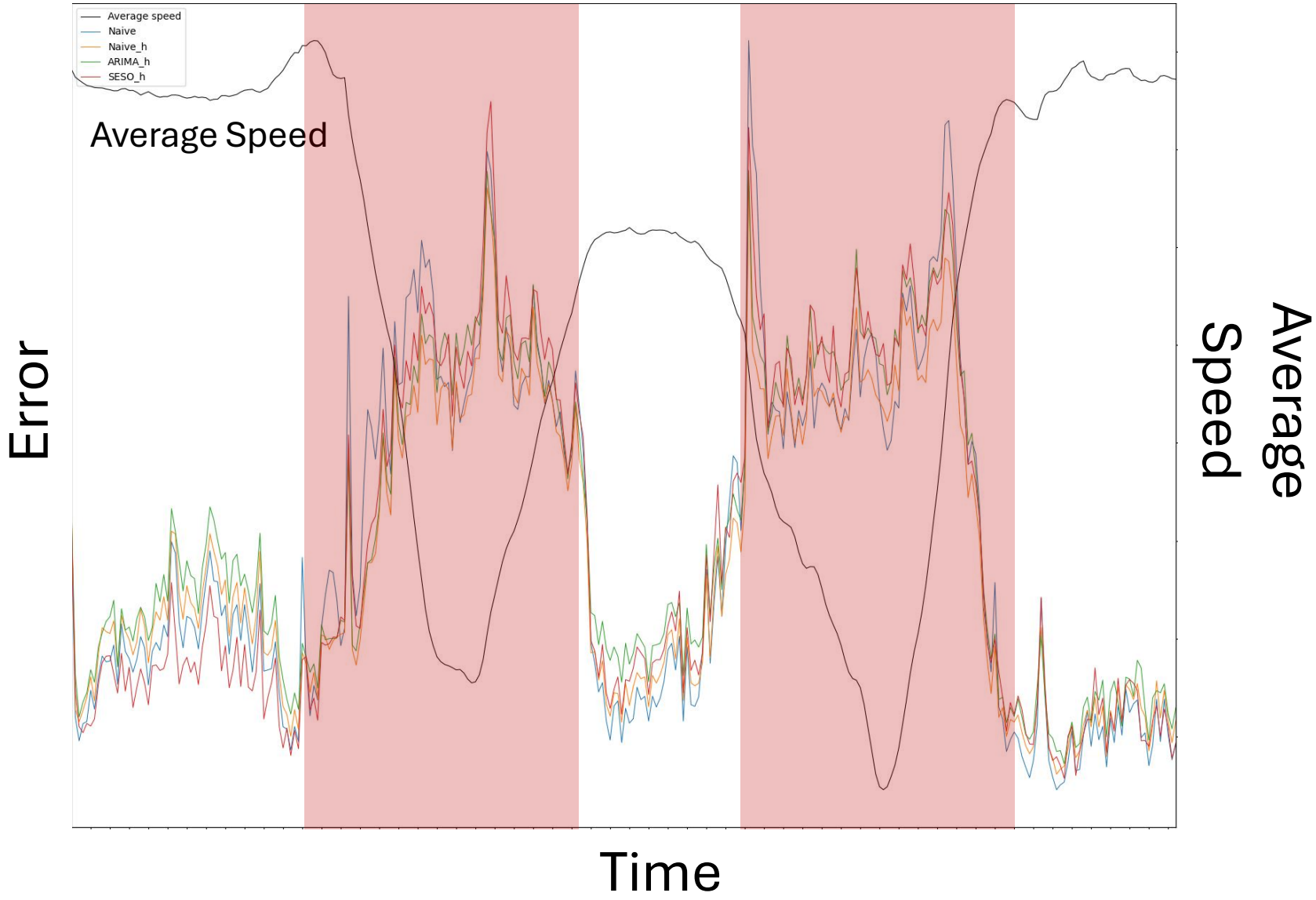


5-minute forecast

Naive

ARIMA

SESO

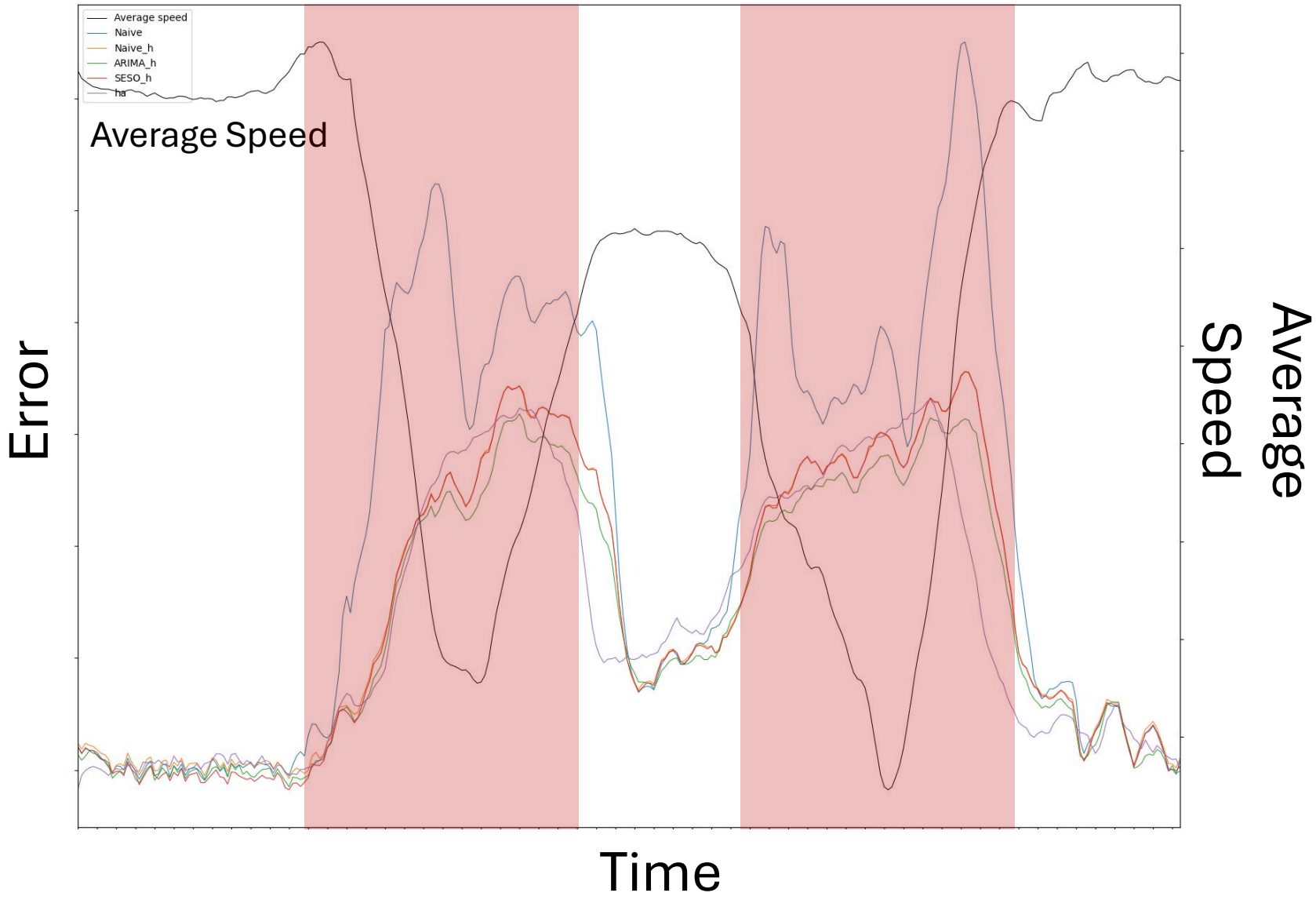


45-minute forecast

Naive

ARIMA

SESO

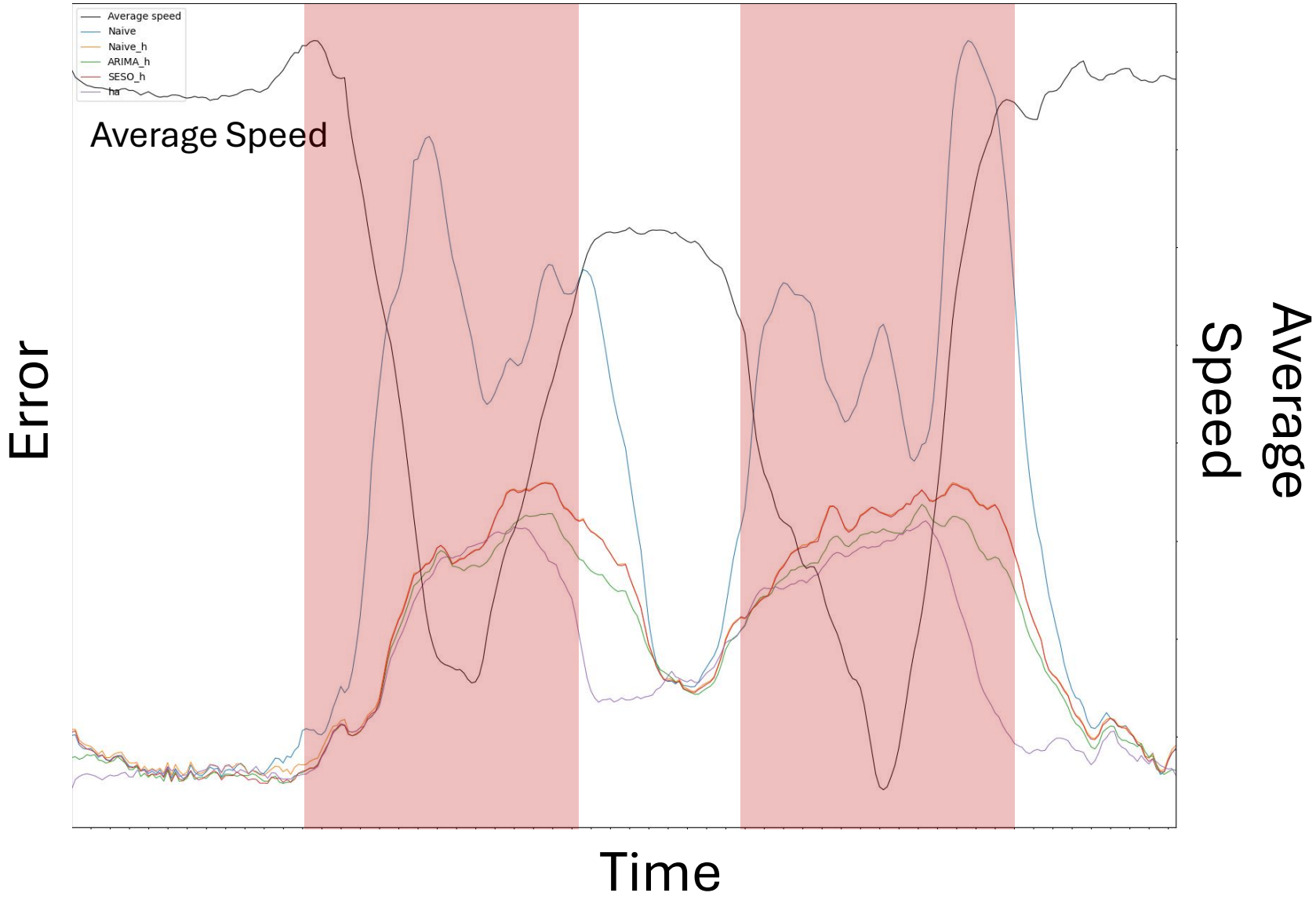


90-minute forecast

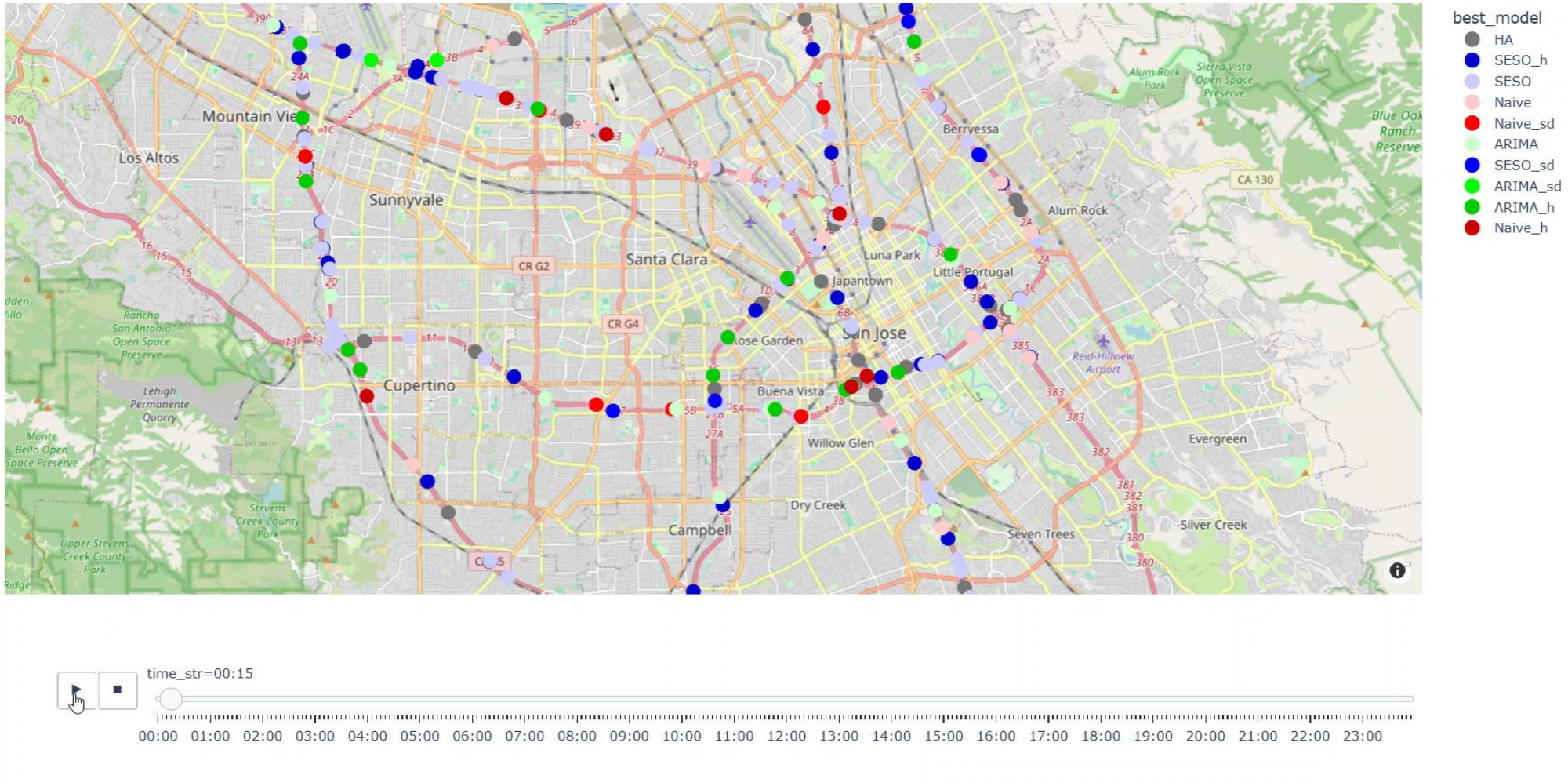
Naive

ARIMA

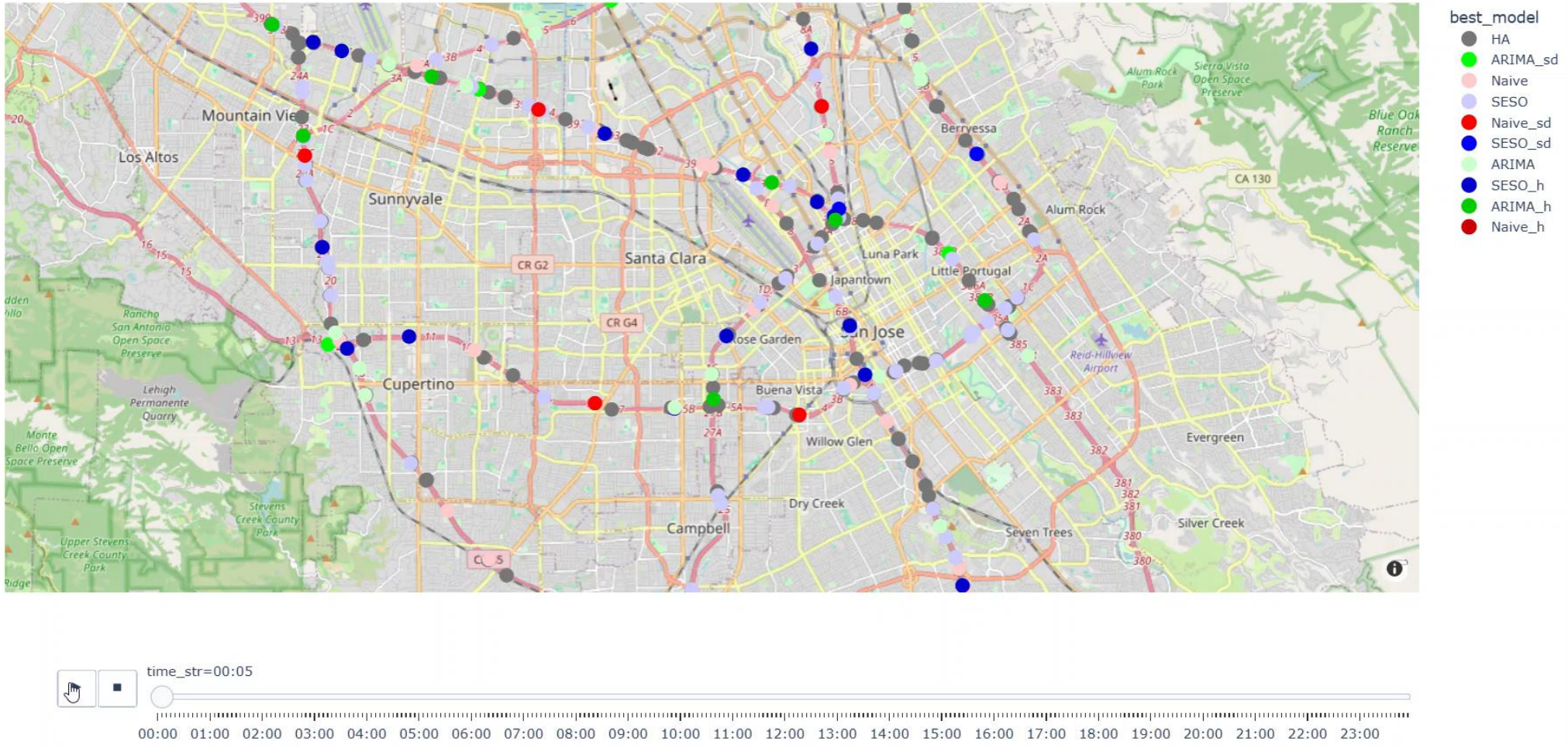
SESO



5-minute forecast of best model



20-minute forecast of best model



Conclusions

- For short forecasts, all models perform similarly, so naïve model is actually a good model and there appears to be little need for more advanced models
- For long time forecasts, naïve model is the worst so more advanced models are better
- For long time forecasts, historical average tends to be the best predictor
- More complicated models appear to be most useful for “intermediate time horizons”

Conclusions