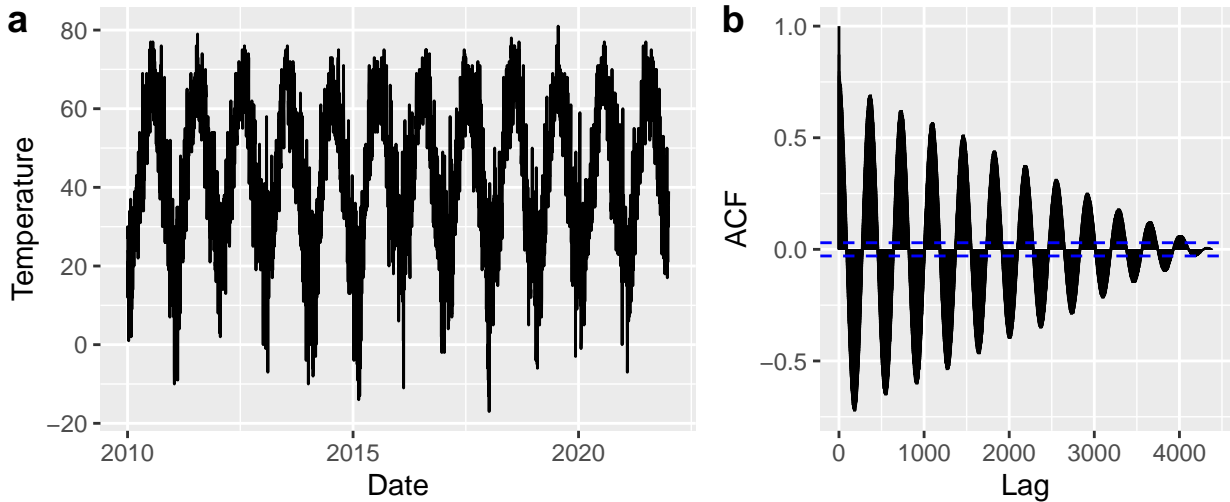


Analysis of the Predictability of Weather Data

Ariane Stark

Temperature is determined by a myriad of factors. This paper seeks to analyze how useful historical weather data is in predicting future temperatures. Specifically the analysis will focus on the weather in Amherst, Massachusetts from January 1, 2010 to December, 31 2021 for a total of 4383 observations. This is a subset of the raw data that started on January 1, 2000 and has been reduced due to missing data instances that all occurred before 2010. The data has a seasonal trend (Figure 1a) with prior observations being heavily correlated with each other seen by looking at the sample auto correlations (Figure 1b). To be able to asses predictive accuracy model fitting and selection were done on the first 90% of the data set. These are the dates from January 1, 2010 to October 18, 2020.

Fig 1: Data and Autocorrelation

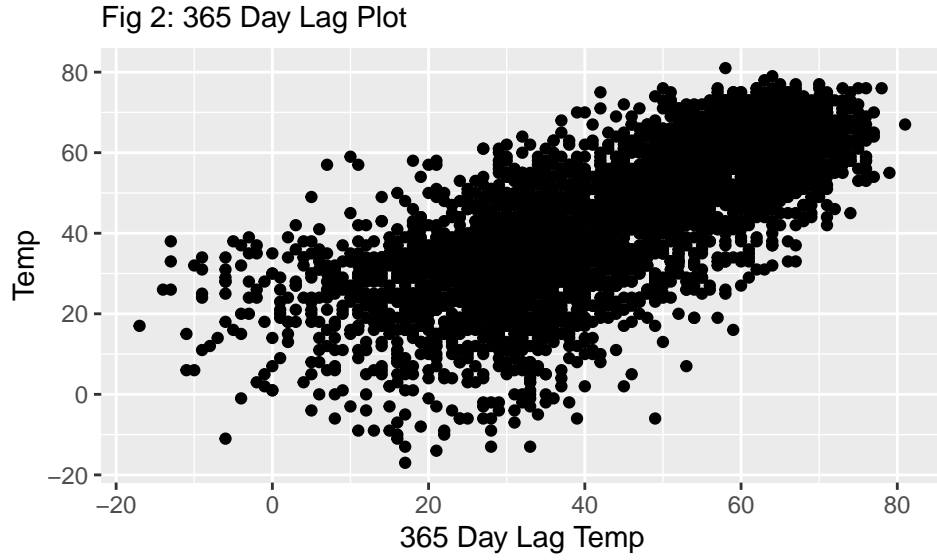


In order to begin the process of fitting an ARIMA model to the data stationarity or the lack thereof had to be addressed. The data without any transformation was not stationary as the mean was time dependent. Due to the patterns seen in the data (Figure 1a) and the 365 day lag values being correlated (Figure 2) a sinusoidal trend that accounted for when a date fell within a calendar year was a good trend to consider to address the non-stationarity

of the data (Eq. 1).

$$\text{Temp.} = X\beta; X = [1, \cos(2\pi\omega), \sin(2\pi\omega)]; \omega = \text{day in year}/\text{num.days in year} \quad (1)$$

To test that this was stationary, the more general Phillips-Perron level test that the residuals from the trend were stationary was conducted. The test with $\alpha = 0.05$ concluded that the residuals were stationary. Note a level test was used as a trend had already been accounted for.

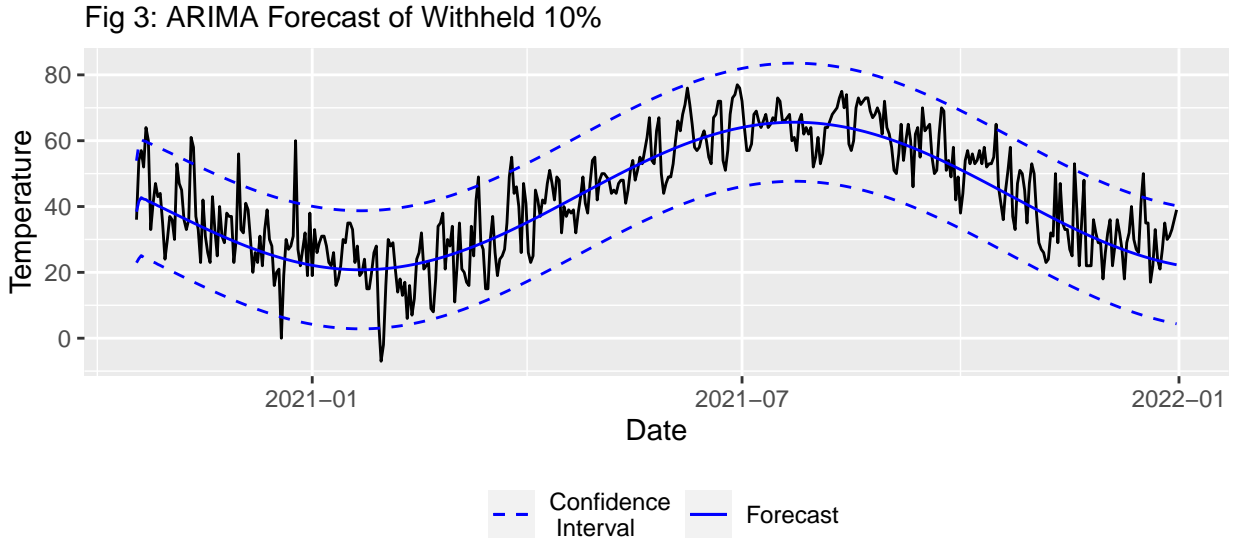


The simpler AR(p) and MA(q) models for deviations from an overall level and day of year ratio means $y_t - \mu - \beta_1 \cos(2\pi\omega) - \beta_2 \sin(2\pi\omega)$ where $\omega = \text{day in year}/\text{num. days in year}$ were considered first. Based on the sample autocorrelations and sample partial autocorrelations a MA(q=2959) model and AR(p=1893) model would be selected respectively.

The AR(p) and MA(q) models alone for modelling the deviation from a level and the sinusoidal terms resulted in extremely high terms. Instead focus turned to ARIMA(p,0,q) models for deviations from an overall level and day of year ratio means. Selection of p and q was done via k-step ahead analysis, as k-step ahead (k=30) model selection is ideal when the goal is forecast accuracy. The 3944 entries in the data frame with 90% of the data got broken up into 10 consecutive sets of 390 observations starting from the first entry. The last

30 observations of the 390 were withheld and were forecasted based on the model fit to the first 360 in that set. The values considered for p ranged from 0 to 16 and for q from 0 to 5. Larger values of p and q resulted in errors. The model with lowest average mean squared error (MSE) over all 10 test sets was ARIMA(4,0,2).

Using the parameter fit to the ARIMA(4,0,2) model, predictions were made for $y_t - \mu - \beta_1 \cos(2\pi\omega) - \beta_2 \sin(2\pi\omega)$ for $t = 3945, \dots, 4383$, the 10% of observation in the initial data frame that were withheld. Inspection of this forecast (Figure 3) does not look great but this is predicting a long time interval (439 days). This forecast had a MSE of 189.61, a mean absolute error of 11, and a 95% prediction interval coverage of 96.81%. The values for point prediction accuracy are mediocre relative to the unit being measured (temperature). A temperature prediction that is on average off by 11 degrees is not great. However, we would expect a 95% prediction interval coverage that is close to 95% which we do get but this is likely because the confidence intervals are wide on average covering a 35.85 degree range.



To assess the model's accuracy over shorter time scales the first 409 days in the last 10% were looked at and on each day forecasts were made for 1, 7, 14, and 30 days in the future. The two point forecast accuracy metrics (MSE and MAE) were reported as were the confidence intervals obtained by taking the 2.5% and 97.5% sample quantiles of the metrics. Overall

when predicting smaller horizons the model had better performance (Table 1).

Horizon	MSE		MAE	
	Value	CI	Value	CI
1	58.39	(0.05, 311.29)	5.92	(0.22, 17.64)
7	70.59	(10.63, 210.32)	6.66	(2.46, 11.45)
14	70.41	(17.2, 159.45)	6.66	(3.02, 9.68)
30	70.83	(24.93, 120.79)	6.7	(3.73, 8.83)
Full Prediction	189.61		11	

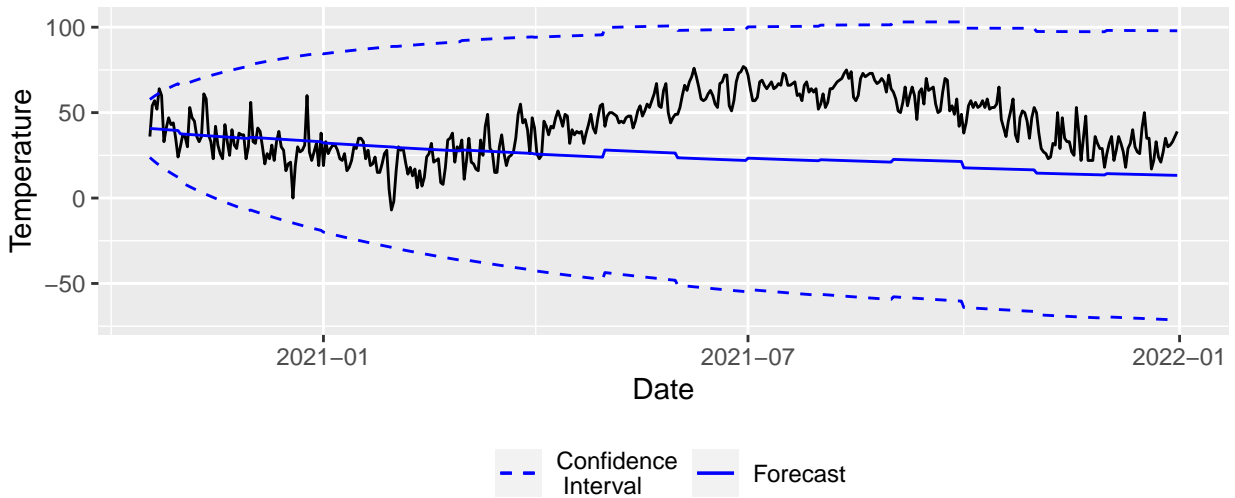
Table 1: Summary of Point Forecast Accuracy By Horizon

Secondary analysis was done on the data frame to predict the temperature using a State-Space model. Initial analysis was done to determine which of three potential state-space model variations was best. The variations considered were the model with no covariates, the model with covariates in the state equation, and the model with covariates in the observation equation. The covariates added in were month indicators of the date. Ideally the sinusoidal trend would have been added in for analysis but that was not compatible with the MARSS package being used for State-Space analysis. Selection between these three methods was done via k-step ahead analysis, as k-step ahead (k=30) model selection is ideal when the goal is forecast accuracy and is identical to the process done for ARIMA model selection. The entries in the training set got broken up into 10 consecutive sets of 390 observations starting from the first entry. The last 30 observations of the 390 were withheld and were forecasted based on the model fit to the first 360 in that set. These are the same 10 subsets were used in the ARIMA model selection process. The model selected was for adding covariates into the observation equation (Eq. 2).

$$\begin{aligned}
y_t &= ax_t + \underline{Z}_t^T \gamma + vt \\
x_t &= \phi x_{t-1} + w_t \\
v_t &\sim_{\text{iid}} N(0, \sigma_v^2); w_t \sim_{\text{iid}} N(0, \sigma_w^2); x_1 = \mu
\end{aligned} \tag{2}$$

Using the state space model with month indicators in the observation equation, predictions were made for the withheld 10% of observations. Inspection of this forecast (Figure 4) looks poor. This forecast had a MSE of 726.34, a mean absolute error of 22.26, and a 95% prediction interval coverage of 99.77%. The values for point prediction accuracy are extremely poor given the unit being measured is temperature. A temperature prediction that is on average off by 22.26 degrees is horrible. However, we would expect a 95% prediction interval coverage that is close to 95% which we do get, but this is because our confidence intervals are extremely wide on average covering a 136.49 degree range which covers the entire range of likely temperature. The prediction for December 31, 2021 (the last data entry in the set) has a lower bound of -59.68 and an upper bound of 109.6 degrees Fahrenheit which are both extreme and potentially implausible temperatures.

Fig 4: State-Space Forecast of Withheld 10%



Considering the two models analyzed, ARIMA and State-Space, it is clear that the ARIMA model works better for this data set and would predict a more accurate temperature than a state space model would. While a temperature prediction that is on average off by 11 degrees is not great, it far exceeds the average error of the state space model of 22.26 degrees. Moreover, short term predictions for the ARIMA model were only off by approximately 6 or 7 degrees (Table 1). Furthermore, visual inspection of the predictions (Figure 3 and 4) have the predictions for the ARIMA model lining up better with the truth than the state-

space model. In conclusion, past temperature and indicator of date in year makes decent predictions of future weather temperature.

References

- Data was acquired from the National Centers for Environmental Information at the National Center for Oceanic and Atmospheric Administration and queried from “here”
- Information about the Amherst, Massachusetts station can be found “here”