Arctic Sea Ice Extent Prediction

Tyler Reid & Paul Tarantino
Department of Aeronautics & Astronautics, Stanford University

INTRODUCTION

Climate change is a pressing and present issue which has come to be at the forefront in the last few decades. As such, understanding climate change is of utmost importance. In addition to environmental impact, these changes will have social and economic ramifications as these changes will affect the way in which we utilize land and sea resources in the future. The minimum annual arctic sea ice extent has been used as a measure of climate change, with record lows being reported in recent years. This reduction in sea ice coverage has given rise to an interest in an increased number of commercial shipping routes through the arctic which have the potential to significantly reduce shipping distance during the summer season.

Computational methods used in the prediction of sea ice extent are typically based on high-fidelity simulations which implement physical models of the environment [1-3]. These involve solving numerous complex differential equations and are computationally expensive. In this project, we would like to take a different approach based on data driven machine learning algorithms. The National Snow and Ice Data Center (NSIDC) has a variety of data sets such as arctic sea ice extent, concentration, temperature, as well as other information derived from the analysis of remote sensing satellites. From 1978 to present, this data has been recorded on a daily basis and is now publicly available for download.

Using machine learning algorithms, the goal of this project is to forecast future sea ice extent and then compare our results to those obtained via complex physics-based climate simulations. We hope to show that a simpler, less computationally expensive, data driven dynamic model can capture the dominant trends as predicted by the complex models used in ice forecasting today.

SEA ICE DATA

The first step in this process was in the acquisition and proper understanding of the sea ice data from the National Snow and Ice Data Center (NSIDC). The data of interest was both spatially and time varying arctic sea concentration [4]. An example data set from October 26, 1978 is shown in Figure 1. Notice that there is a hole missing in the middle of the data set. This is an artifact of the remote sensing satellite orbits not traversing directly over the geographic North Pole. The satellite's orbital inclination is decided based on operational constraints. Here, inclination is chosen so as to give rise to a precession rate which allows the satellite to ascend and descend at the same local longitudes at the same local time every day. This orbit is known as a Sun synchronous orbit.

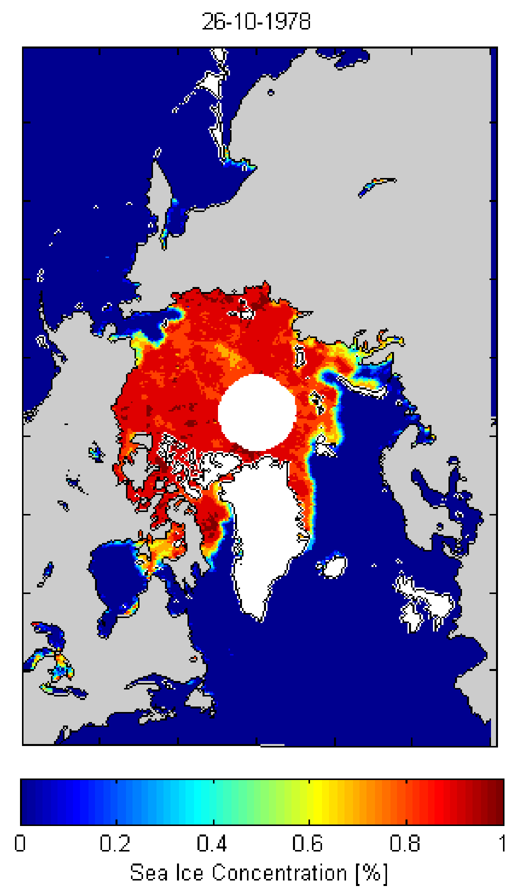

Figure 1: Arctic sea ice concentration

The problem with the data is that different satellites have been used in data collection over the years giving rise to different sized holes in different years. Examining the data set has shown that this missing region is always a part of the sea ice extent since the ice concentration is

15% or greater [3]. We can thus gap fill the data sets to achieve consistency. An example of a post-processed data set is shown in Figure 2. Here, we see only the arctic sea extent in lieu of the finer details of sea ice concentration shown in Figure 1.

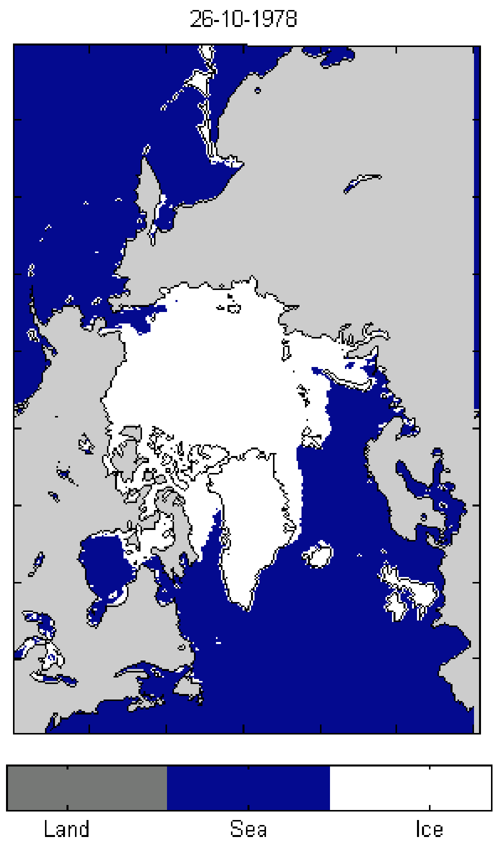

Figure 2: Arctic sea ice coverage

We are ultimately interested in modelling and forecasting the total arctic sea ice extent. It was decided that modeling both the spatial and temporal sea ice extent would be too computational expensive and outside the scope of this project. Instead, we will attempt forecasting of the total arctic sea ice extent as a function of time, a time series which is derived from post-processing of the data shown in Figure 2. This amounts to determining the area represented by each pixel as defined by the NSIDC and adding them together. As a validation of our methodology, we compared our results with those given by the NSIDC. Figure 3 shows the September average arctic sea ice extent from 1979 to 2011 as calculated by the NSIDC. September sea ice extent is typically used as a metric for how much sea ice coverage is diminishing due to the fact that September represents the month with minimum sea ice coverage. Figure 4 shows this plot as calculated by our methods for the years 1979 to 2012. We find that these have nearly identical results. This comparison gives us confidence that we are indeed using the correct time series as our training set in our learning algorithms.

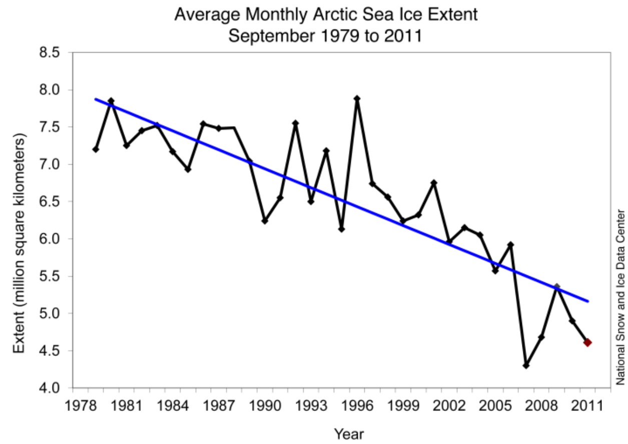

Figure 3: September arctic sea ice extent as computed by the NSIDC [5]

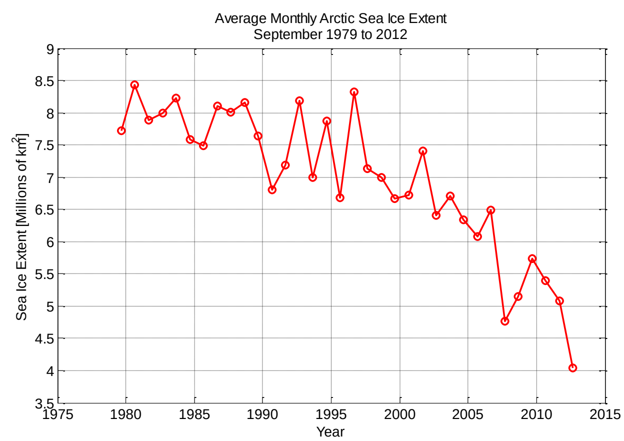

Figure 4 September arctic sea ice extent

SEA ICE FORECASTING

Our goal is to predict arctic sea ice extent using only the time series data described in the previous section. This problem amounts to the forecasting of a stochastic dynamical system based on an observed time series. There have been many methods developed to tackle this class of problem, many of which can be found in [6].

Here, we treat the time series as an unknown dynamic system of the form:

$$y_t = f(y_{t-1},\ y_{t-2},\ldots,\ y_{t-d}) \tag{1}$$

where d is the maximum delay our dynamic system is reliant on to achieve the next output in the time series. Our approach is to use machine learning algorithms, specifically Support Vector Regression (SVR), to determine this function f. SVR has been used successfully in the time series forecasting of financial systems [7], and we feel that due to the similar nature of the problem, this approach would work for this case as well.

The problem was setup assuming m training sets with scalar output data $y^{(i)}$ and input feature vectors $x^{(i)}$. The features are the previous d outputs due to the assumptions on the form of the dynamic model given in Eq. (1). Next we solved the optimization problem associated with SVR [8]:

$$\text{minimize} \quad \sum_{i=1}^{m}(\alpha_i + \alpha_i^*) + C\sum_{i=1}^{m}(\xi_i + \xi_i^*)$$

$$\text{subject to} \quad \begin{cases} y_i - \sum_{i=1}^{m}(\alpha_j - \alpha_j^*)k(x_j, x_i) - b \le \varepsilon + \xi_i \\ \sum_{i=1}^{m}(\alpha_j - \alpha_j^*)k(x_j, x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \alpha_i, \alpha_i^*, \xi_i, \ \xi_i^* \ge 0 \end{cases} \tag{2}$$

where ξ_i, ξ_i^* are slack variables, α_i, α_i^* are Lagrange multipliers, $k(x_j,x_i)$ is the kernel function, b is the offset variable, ε is the soft margin parameter, and C is the variable that determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated. C and ε can be arbitrarily selected before solving the optimization problem.

The optimization problem given in Eq. (2) was solved using the freely available CVX software package, a research tool for Matlab-based analysis of convex optimization problems [9, 10]. Using CVX, this problem is solved as follows:

```
cvx_begin

    % define variables
    variable alpha(m,1)
    variable alphaStar(m,1)
    variable epsi(m,1)
    variable epsiStar(m,1)
    variable b

    % define optimization problem
    minimize(oneMat'*(alpha+alphaStar) + ...
                 C*oneMat'*(epsi+epsiStar))

    % define constraints
    subject to
    Y'-K*(alpha-alphaStar)-b<=epsilon+epsi;
    K*(alpha-alphaStar)+b-Y'<=epsilon+epsiStar;
    alpha>=0;
    alphaStar>=0;
    epsi>=0;
    epsiStar>=0;

cvx_end
```

In order to run this script, we had three main variables to select: the free optimization parameters C and ε as well as the delay d to be used in the model. The training data was processed to be evenly spaced at 30 day intervals, representing approximately a month between data points. The delay d is chosen to be one half the length of the data set in order to reserve data points for tuning of parameters C and ε. This was done by choosing initial conditions based on our selection of the delay d, marching forward in time until the end of the known data set, and comparing how well the forecasted result matches the known data. Sweeping through several values of C and ε allows for the determination of the optimal model. In addition, we found that a linear kernel gave rise to the best results. Therefore, this scheme will produce a linear dynamic model of the system which allows for combinations of sinusoids and exponential functions. Thus, the model can capture several harmonics as well as exponential behavior.

VALIDATION RESULTS

In order to validate the effectiveness of the approach we used 1978 – 2002 data for model training and tuning parameter selection. Next, we validated our forecasted results from 2002 to 2012 against known sea ice extent data. For this simulation, a delay of d = 130 (~11 years) was chosen, and C and ε which minimized the model error were 4.22 and 0.017, respectively. The validation of predicted seasonal sea ice extent for 2002 - 2012 is given in Figure 5. Figure 5 shows that our model predicts seasonal variation as well as yearly sea ice extent decay well.

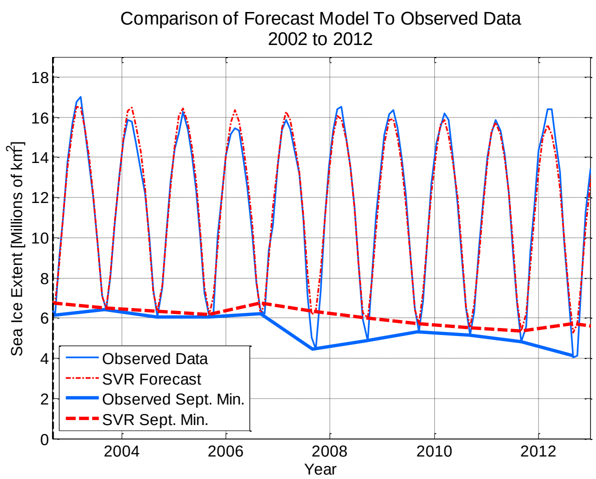

Figure 5: Seasonal arctic sea ice forecasting using model built with 1978 - 2002 training data

FORECASTING RESULTS

Now confident in the algorithm's ability to accurately model the system, we reran the model using all of the data available for training and tuning parameter selection. A delay of d = 200 (~17 years) was chosen and C and ε were found to be 0.089 and 10.67, respectively. Figure 6 depicts the parameter selection trade-off contour. This contour illustrates that an optimal value of the tuning parameters exists.

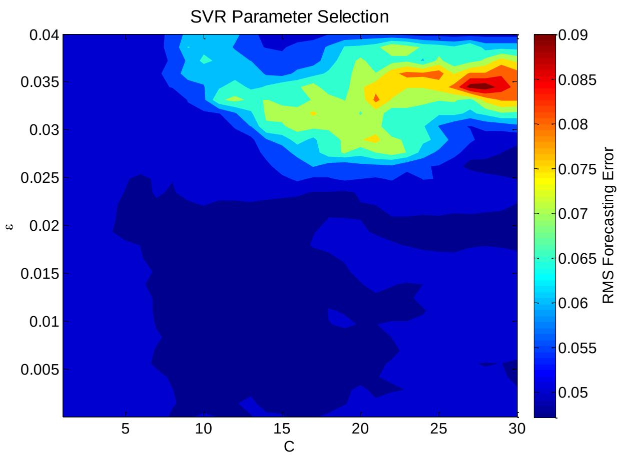

Figure 6: Parameter selection trade-off contour

Figure 7 shows the results obtained from the optimal parameters selected from the trade-off process described earlier. This represents the full seasonal variation of the arctic sea ice as well as the September minimum. The blue curve is the observed data obtained from the NSIDC and the red curve is the result obtained by forecasting with the learned model. The leftmost vertical line represents the initial condition selected for the model corresponding to the delay d. The rightmost vertical line represents the point at which forecasting begins. Even though the model was learned using the full data set, the propagation forward from a single initial condition demonstrates the strength of the model in that no seasonal drift occurs with respect to the observed time series and that it follows the observed data well overall.

Figure 8 shows the arctic sea ice September minimum for both the observed data and that which was forecasted by the model. Running this model into the unforeseen future allowed us to obtain an estimate as to when there would be no arctic sea ice in the September minimum season. Our model predicts that this would occur in the year 2030, which is surprisingly not far in the future.

To get a sense if our estimate was a reasonable one, we looked at results obtained from high-fidelity physics based global climate model. Figure 9 shows the September minimum prediction results obtained from several different organizations and models from around the world. Each thin colored line represents one ensemble member from the model. The thick yellow line is the mean of all ensemble members, and the blue line is their median value. The thick black line represents observations based on data from the NSIDC (previously shown in Figure 3). Here we can see that the climate models predict a zero arctic sea ice September minimum scenario as early as 2020 and as late as 2100. This puts our prediction of 2030 within the range of results obtained from this conglomerate of climate models.

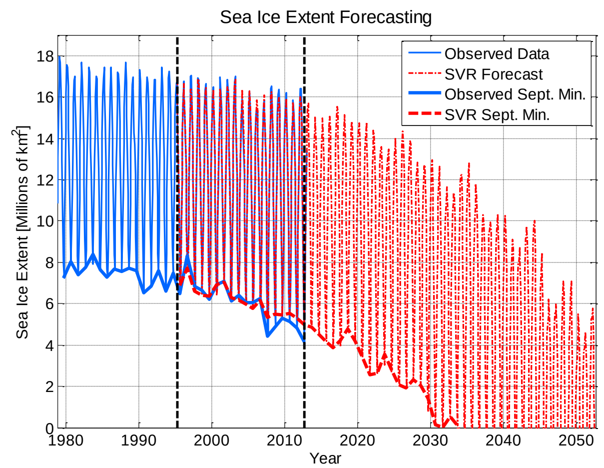

Figure 7: Seasonal arctic sea ice forecasting using model built with all training data

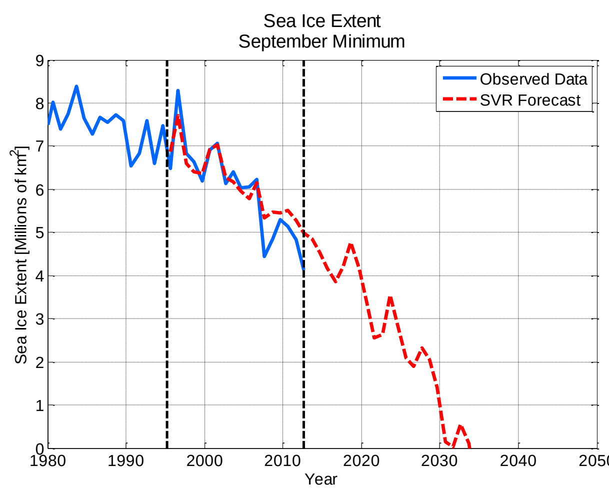

Figure 8: September minimum arctic sea ice forecasting using model built with all training data

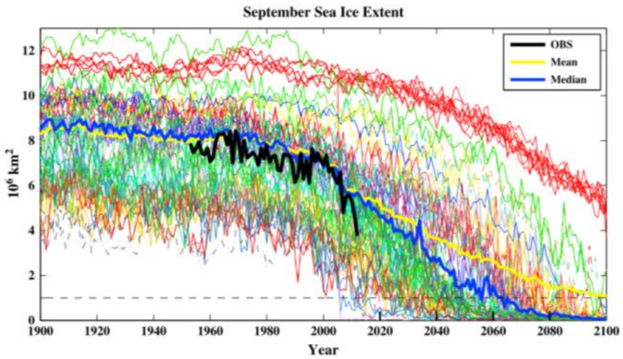

Figure 9: September sea ice extent based on 89 ensemble members from 36 CMIP5 models under the CP8.5 (high) emission scenario. [2]

CONCLUSION

We have used machine learning techniques to develop a linear dynamic model for the forecasting of arctic sea extent based on Support Vector Regression (SVR). This model is not limited to the September minimum ice extent, but also includes the full seasonal variation as well. The results from this model gave forecasts which are within the range of results obtained via complex numerical simulations in the arctic sea ice community. This gives us confidence in the learned dynamic model's ability to capture the dominant trends in the data as well as in the techniques employed to obtain this dynamic model.

REFERENCES

[1] M. Chevallier, D. Salas y Mélia, A. Voldoire, M. Déqué, and G. Garric, "Seasonal Forecasts of the Pan-Arctic Sea Ice Extent Using a GCM-Based Seasonal Prediction System," *Journal of Climate*, vol. 26, pp. 6092-6104, 2013/08/01 2013.

[2] J. E. Overland and M. Wang, "When will the summer Arctic be nearly sea ice free?," *Geophysical Research Letters*, vol. 40, pp. 2097-2101, 2013.

[3] J. C. Stroeve, V. Kattsov, A. Barrett, M. Serreze, T. Pavlova, M. Holland, *et al.*, "Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations," *Geophysical Research Letters*, vol. 39, p. L16502, 2012.

[4] D. J. Cavalieri, C. L. Parkinson, P. Gloersen, and H. Zwally, "Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data," ed. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center, 1996, updated yearly.

[5] NSIDC. (2011). *Summer 2011: Arctic sea ice near record lows*. Available: http://nsidc.org/arcticseaicenews/2011/10/summer-2011-arctic-sea-ice-near-record-lows/

[6] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," *CoRR*, vol. 1302.6613, 2013.

[7] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, pp. 307-319, 2003.

[8] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199-222, 2004.

[9] M. Grant and S. Boyd. (September 2013). *CVX: Matlab software for disciplined convex programming, version 2.0 beta*. Available: http://cvxr.com/cvx

[10] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., ed: Springer-Verlag Limited, 2008, pp. 95-120.