Applications of machine learning to the forecasting of short-term sea lice abundances in British Columbia

Tristan Garry – EEB498 – 2019-2020
Supervisors: Martin Krkosek and Marie-Josée Fortin
https://tristangarry.github.io/sea-lice-forecasting/

Abstract

Wild salmon populations have been negatively affected by salmon farms through the increased prevalence of disease brought on by farming. Sea lice have been well-researched in the Broughton Archipelago of British Columbia, Canada. Through resulting surveys, it was found that sea lice affliction in farmed salmon has been shown to increase sea lice prevalence on wild populations and the ability to predict sea lice outbreaks by identifying the relevant environmental factors would aid conservation efforts. Knowledge of the variability in sea lice occurrence due to the environment is crucial for fishery and conservation efforts, yet the benefit of forecasting has not been fully explored even with a large set of data available. Given that other industries have used machine learning to provide better forecasting based on large data sets, regression, machine learning models, and deep neural networks were used in order to identify the relative importance of climatic factors and standing stock of lice in nearby farms for the prediction of sea lice occurrence. Although observed years were modeled accurately, it was found that the chosen models were unable to consistently and accurately forecast short-term sea lice occurrence for unobserved years. It is recommended that future modelling efforts explore the scale of the underlying processes that affect sea lice abundance as well as including explicit spatial factors and local densities. These findings indicate that further research is required in understanding the factors that would allow us to produce short-term forecasts of sea lice abundances.

Introduction

In a world increasingly marked by big data and easy access computing, scientific disciplines have been able to take advantage of analytical advances from multiple disciplines. Ecology is becoming a more predictive science (Dietze et al. 2018, White et al. 2019) and is moving towards using more computationally intensive methods (Green et al. 2005). With widespread access to large datasets and stronger computing, forecasting is becoming easier to perform and as such our perspective on the subject is changing. Ecology has traditionally lacked short-term iterative forecasts, a form of forecasting performed repeatedly on relatively short timescales with

consistent evaluation (Dietze et al. 2018, White et al. 2019) and that can be used to make more effective types of inferences. Without iterative forecasting on new data, it is difficult for us to evaluate our ability to make causal inferences and our ability to improve our understanding of our forecasting ability is hindered. There is little literature on forecasting in ecology compared to other fields, perhaps due to the perceived difficulty of accurately modelling biological systems (Houlahan et al. 2017).

Forecasting is imperative in many fields and industries such as economics (Teschner and Weinhardt 2015) and energy generation (Suganthi and Samuel 2012). Such fields have seen the use of large amounts of different modelling techniques in order to best accommodate the high demand for forecasts unlike in ecology which has not seen the same magnitude of forecasting efforts. Themes applicable to ecology have been explored in forecasting population behaviour from macroeconomics (Teschner and Weinhardt 2015), and forecasting disease in public health (Shaman and Karspeck 2012). It has been found in numerous instances that repetition of forecasts over time improves their accuracy (Teschner and Weinhardt 2015, Dietze et al. 2018), and fields such as economics and energy have benefited greatly from large amounts of repetition and review. While ecology does not have the same history in performing forecasting as many fields and industries, it would benefit the field of ecology to look to other fields when further developing its own forecasting expertise.

Forecasting has been applied in ecology for analyzing populations in fisheries management, the conservation of threatened species, and examining population responses to perturbations (Ward et al. 2014). The intricacies of ecological data, such as population and spatial structure, has brought on the adoption of some novel modelling techniques in the field. As an example, ecologists have had to develop models able to independently define subpopulations

by using methods such as state-space models (Ward et al. 2010) in order to account for spatial structure. Although this complexity may seem daunting at certain levels of ecological organisation (Ward et al. 2014), many attempts at forecasting populations have been successful and iterating on these forecasting approaches is key in order to move our understanding of the predictability of ecology forward (White et al. 2019).

A set of models that is of interest in forecasting is artificial neural networks (ANN) (Böse et al. 2017). This type of model "learns" relationships present in data in non-parametric ways by creating a graph of connected nodes with weights that fit to the data passed in to them in order to describe a pre-determined output. This can perhaps address concerns that biology is "too complex" to accurately model (Dietze et al. 2018) as no functional form for the data-generating model has to be assumed. ANNs have shown very strong results in many applications such as natural language processing and image recognition (Gu et al. 2018) and have shown strong results in forecasting in certain applications (Kimoto et al. 1990). Combination of neural network and parametric methods have also seen great success (Makridakis et al. 2018) and may even be preferred in some circumstances. Advances in machine learning (Crone et al. 2011, Nevo et al. 2018, Wang et al. 2019) could be used to forecast difficult-to-model ecological relationships and expand on the growing imperative of ecological forecasting.

Many other non-ANN machine learning algorithms have been used in forecasting (Carbonneau et al. 2008, Makridakis et al. 2018). Among these, regression-based methods have proven very effective in certain situations (Carbonneau et al. 2008) and tree-based methods in others (Voyant et al. 2017). The algorithm that will perform the best typically depends on the structure of the data, which is impossible to completely describe (James et al. 2013). However, typically regression-based methods perform better on more linear data and tree-based algorithms

and ANNs perform better on less linear data (James et al. 2013). This means that unless a dataset has been extensively studied before, all modelling approaches should be considered when first creating models.

Forecasting has seen many uses in disease research, where an accurate forecast of an outbreak can be crucial to the health of a population (Petukhova et al. 2018). Disease can have drastic effects on population viability, as has been seen recently in White-nose syndrome in the bats of North America (Wilder et al. 2011). Of current relevance, climate change can have strong effects on the prevalence of disease (Zamora-vilchis et al. 2012) and so attention should be given to the understanding of these relationships. Parametric forecasting methods such as ARIMA models have been applied to the forecasting of many diseases historically (Kaundal et al. 2006, Molento et al. 2018). More recently, ANNs have shown to be very powerful when estimating disease response to climate due to the sometimes difficult to model relationship between disease and its environment (Kaundal et al. 2006).

A disease of particular interest to this project is the parasitism of salmon by sea lice on the west coast of Canada. Sea lice parasites have had negative effects on both wild and farmed salmon populations in the area, but these parasites have shown to be controllable given the proper treatment (Peacock et al. 2013). Abundance of the lice in local fish farms has been shown to have a correlation with occurrence of the lice in wild fish (Marty et al. 2010), perhaps indicating that lice management on farms would have significant effects on wild populations. Critically, the timing of management efforts was found to be important in the control of sea lice abundance on salmon (Bateman et al. 2016) and so being able to forecast outbreaks would be a useful management tool. Improvement in these forecasting methods could also spur further development in the field of forecasting ecological systems.

There are two species of lice that are common on Pacific salmon: *Lepioptheirus salmonis*, a sea louse that is a salmon specialist, and *Caligus clemensi*, a generalist that is less pathogenic to salmon. Sea lice reproduce sexually and produce nauplii that are able to disperse tens of kilometres along ocean currents at the first stage of their life. These nauplii moult into copepodites, with timing depending on temperature. At this stage, the lice are infectious to fish and attach to a host before progressing through immobile chalimus stages and then two mobile pre-adult stages. At the adult stage, sea lice are mobile on the surface of their host and may move between hosts (Peacock et al. 2019). These sea lice have been the subject of extensive research and management efforts (Peacock et al. 2013). Accurate forecasting of sea lice abundance, and specifically outbreaks, could help the control of sea lice infestations in the region and help deter the damage caused by parasites on both wild and farmed animals.

Methods

Data

Wild Juvenile Salmon Data

Forecasting efforts were focused on estimating sea lice abundance on wild juvenile salmon in the Broughton Archipelago, BC. The data consist of a monitoring effort ongoing from 2001 to 2019, specific details on sampling and methods are available in a supporting paper (Peacock et al. 2016). Sampling consisted of visually searching coastal waters, ~2-5m from the shore, and capturing schools of juvenile salmon for examination. Once captured, the salmon were measured along with notes on their health characteristics and species. The salmon were non-lethally visually examined using 16x magnification for sea lice, with life stage and species of lice being noted. As well as fish and lice information, site information was collected for each sampling event including temperature and latitude/longitude.

The forecasted variable (response variable), motile sea lice, is derived from this dataset; motile sea lice for this analysis are defined as pre-adult and adult stage male and female sea lice. Predictor variables (covariates) of interest from this dataset include date and site information, temperature and location.

Farmed Salmon Data

Parasite occurrence in farms is thought to have a significant effect on parasite occurrence on wild fish (Marty et al. 2010), and so data describing salmon farm sea lice counts were analysed. The dataset used consists of Department of Fisheries and Oceans (DFO) mandated counts of monthly averages of motile sea louse occurrence. These data are collected by the farms themselves and averages are submitted to the DFO, however resolution to the individual sampling event level is

not publicly available. Variables of interest from this dataset include date, site location, and average motile sea lice count.

Weather Forecast Data

In order to supplement the months not sampled by the wild salmon sampling efforts, daily weather reports from a nearby weather station, Port Hardy, were analysed. Historical air temperature reports were used to find relationships between off-season trends and the collected wild salmon data that were not sampled by the wild sampling efforts.

Frequency

All data were resampled to a weekly frequency when training models and making predictions to align with the other data and the requirements of the models, using linear interpolation for missing data (Moritz and Bartz-Beielstein 2017).

Predictors and Response Variables

The response variable for all forecasting was the count of all species of motile sea lice on wild salmon. Predictor variables were: surface temperature and salinity of wild salmon sampling sites; juvenile sea lice on wild salmon; industry motile sea lice averages from Sargeaunt Pass, Doctor Islets, Humphrey Rock, Burdwood, Glacier Falls, Sir Edmund Bay, and Wicklow Point salmon farms; and air temperature from the Port Hardy weather station.

Modelling

Long Short-Term Memory (LSTM)

The first model trained for forecasting was a Long Short-Term Memory (LSTM) model (Fig. 1), a type of recurrent neural network (Sainath et al. 2015). This is a type of neural network with feedback as well as feedforward connections, making it suitable for time series predictions, as seen in Fig. 1. Neural networks allow for the description of highly non-linear relationships, but many neural network frameworks do not account for sequential data such as time series. An LSTM model was chosen as inputs are taken to be inherently sequential and therefore can be modelled as a time series adequately.
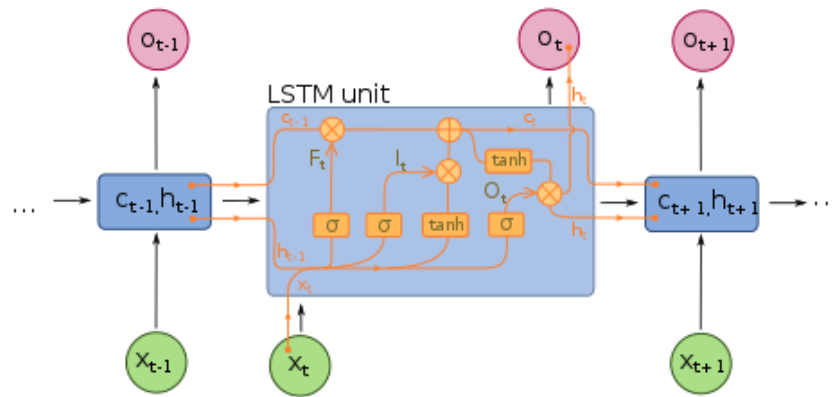

Figure 1. Long Short-Term Memory model unit. For each timestep t, given the inputs x_t, the overall model weights h_t are influenced by the short-term trend c_{t-1} to produce the output o_t and the trend c_t and model weights h_t are updated. Author: fdeloche, Wikimedia Commons. License:
https://creativecommons.org/licenses/by-sa/4.0/legalcode

Before fitting the model, data were normalised and scaled using maximum absolute value scaling. To assess the modelling framework's ability to forecast each year, an LSTM model was trained on all years of data excluding one year. This excluded year would be used to test the viability of the model. This model was then tasked to predict motile sea lice on wild salmon from

March to July under three conditions of available predictor data from that year: all predictor data available, predictor data available up to and including April, and predictor data available up to and including June. Models were trained and assessed using this method on all years of data from 2003-2017.

Models were built using Keras 2.3.1 and Tensorflow 1.13.2. Code for specific configurations can be found at https://github.com/TristanGarry/sea-lice-forecasting/tree/master/lstm.

AutoML

Microsoft Azure Auto Machine Learning (AutoML) (Barnes 2015) was used in order to assess the effectiveness of other machine learning methods. AutoML is a Microsoft service that allows several specified machine learning models to be assessed on a dataset. This was used to fit Elastic Net, Decision Tree, Random Forest, and Gradient Boosting models as well as voting and stack ensemble models of all the fitted models. Using AutoML, additional features were constructed, and each model was tried with both maximum absolute value and regular standardisation. Three classes of models were applied using AutoML: Linear regression, tree-based methods, and ensembling methods.

Linear Regression

Linear regression methods of the general form $Y = \sum_{i=0}^{p} \text{ß}_i x_i$ were one class of models fit using this framework. The algorithm used to fit linear regression models is known as Elastic Net, which is a regression method that linearly combines the penalty terms of lasso and ridge

regression methods (Zou and Hastie 2005). The parameters, ßi, are constrained to only be large if they contribute sufficiently to the fit of the model, as seen in Fig. 2.
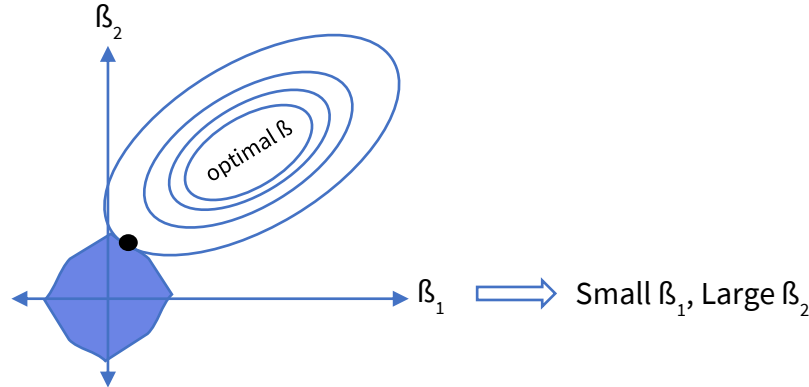

Figure 2. Elastic Net constraint visualisation in two dimensions. Ordinary least squares methods would choose a parameter in the centre of the elipses around the optimal ß. Elastic Net is constrained to choosing parameters on the surface of the shape around the origin in order to ensure parameters provide sufficiently to the fit.

Tree-Based Methods

Tree-based methods consist of splitting the data based on predictor threshold values in order to best explain it, as seen in Fig. 3. At each split, the sample is split into two (or more) sub-samples based on the most significant predictor. The number of splits to explain a dataset can be chosen as an arbitrary integer, using a stopping criterion (e.g. R_2 threshold value), or dynamically by using another algorithm. Three different tree-based methods were evaluated using AutoML: decision trees, random forest, and gradient boosting.

Decision trees consist of splitting the data based on threshold values of the predictors sequentially until the data is sufficiently explained or a stopping criterion is reached (James et al. 2013).

Random forest consists of fitting several decision trees to the same data and ensembling these models' predictions for the random forest's predictions. Individual decision trees are ensured to not be correlated by randomly choosing a subset of the original predictors that can be fit on, as well as bootstrapping each tree's data (James et al. 2013).

Gradient boosting is similar to random forest except that sequential decision trees correct for the previous one's errors (James et al. 2013).
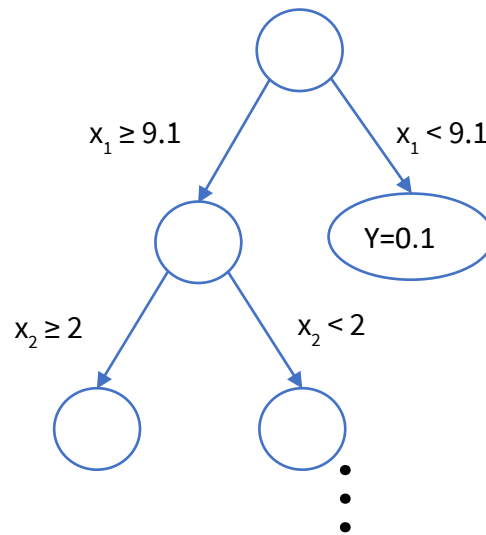

Figure 3. Visualisation of a tree-based learning model. In this case, the first split on predictor X_1 splits the sample into two subsamples, one where X_1 is greater than or equal to 9.1and another where X_1 is less than 9.1. This would continue for all predictors until some stopping criteria is reached.

Ensembling Methods

Ensembling methods take output from other models and create a meta-model with them in order to create models that can perform better than any single algorithm, as seen in Fig. 4. These outputs can be combined together to produce a new output or can be processed by another machine learning algorithm to create a model. The ensembling methods used were: voting ensemble and stack ensemble.

Voting ensemble (Wang et al. 2011) consists of allowing all models to make a prediction. Using these predictions, the voting ensemble output is the most agreed-upon prediction based on soft voting with weighted averages.

Stack ensemble (Wang et al. 2011) consists of a first layer using the trained models to recreate a training dataset based on the trained models' predictions, and a second layer training an Elastic Net model on these outputs.
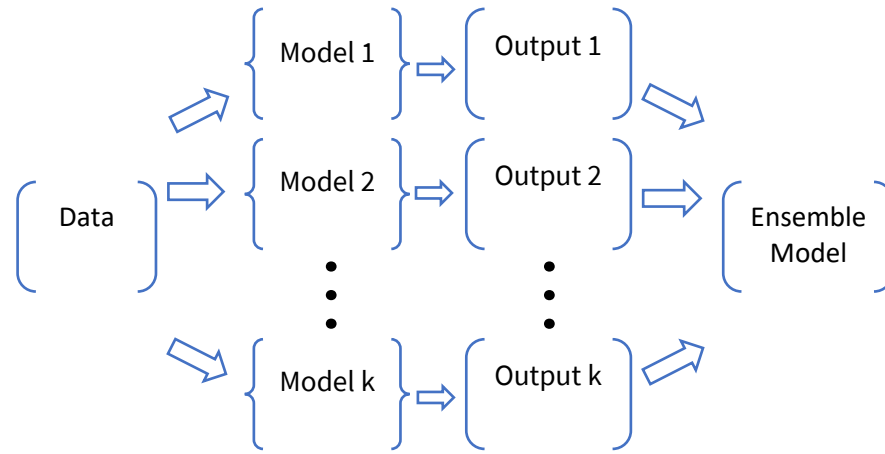

Figure 4. Visualisation of how models come together to form an ensemble model. The ensemble model takes outputs from each of the previously fit models and can either use these outputs directly to form its own output or use the models' outputs to parametrise a new model.

Similar to the LSTM evaluation, AutoML models were trained on all years of data excluding one year. These models were then tasked to predict motile sea lice on wild salmon from March to July under three conditions of available predictor data from that year: all predictor data available, predictor data available up to and including April, and predictor data available up to and including June. The best performing model, the stack ensemble model, was the model selected by this method. Models were trained and assessed using this method on all years of data from 2003-2017.

Models were built and trained using AzureML 1.0.83. Code for specific configuration can be found at https://github.com/TristanGarry/sea-lice-forecasting/tree/master/automl.

Naïve Seasonal Bayesian

A naïve seasonal Bayesian model was used as a null model. Given a date of the year, this model outputs the average of all previous years' outputs on that date.

This model was built and trained using Numpy 1.17.3 and Pandas 0.25.2. Code for specific configuration can be found at https://github.com/TristanGarry/sea-lice-forecasting/tree/master/bayes.

Results Interpretation

The predictions generated from the LSTM, AutoML, and Naïve Bayesian methods were assessed for accuracy based on root mean squared error (RMSE) and mean absolute error (MAE).

Programming environment

Analyses and data were handled using Python 3.6, Pandas 0.25.2, and Numpy 1.17.3. All code can be found at https://github.com/TristanGarry/sea-lice-forecasting, with specific package details in /setup/environment.yml.

Results

Model Diagnostics

Based on the outputs achieved, models performed very well on training data. Both the LSTM and AutoML models outperformed the Naïve Bayesian model in fitting the training data. Training scores saw very little variance in performance year-to-year as seen in the small spread in training scores in Figure 5. Testing scores showed very high variance, with some testing years having a high degree of accuracy and many having a low degree of accuracy, as seen in Figure 1. AutoML models in particular showed a very high variance year-to-year in test performance.
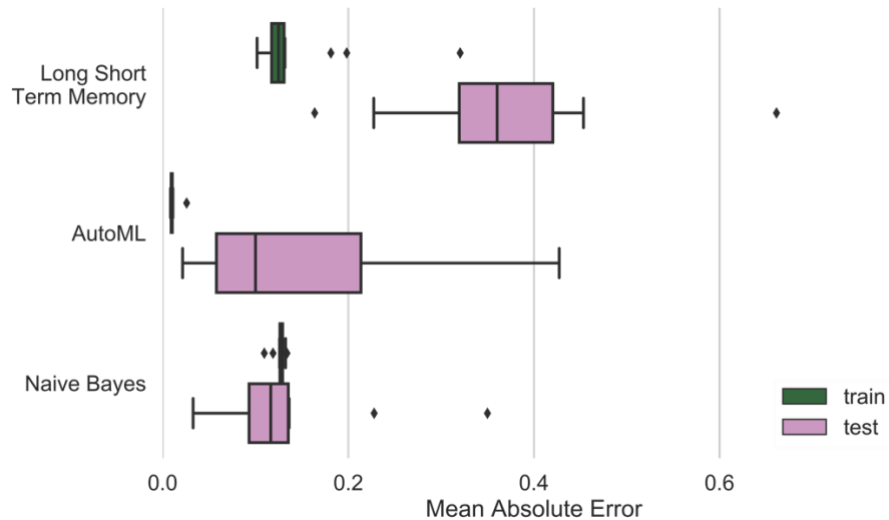

Figure 5. Model training diagnostic results. Individual data points indicate individual training datasets.

Forecasts

Full test data set

Forecasts produced by both the LSTM and AutoML models did not produce sufficiently accurate forecasts, as evidenced in Figure 6. Although some years such as 2006 and 2007 had good forecast accuracies and low variability in the methods' predictions, most years had a large variation in the forecasts produced. The large spread in any given years' predictions between the three models indicates a high uncertainty in predictions.
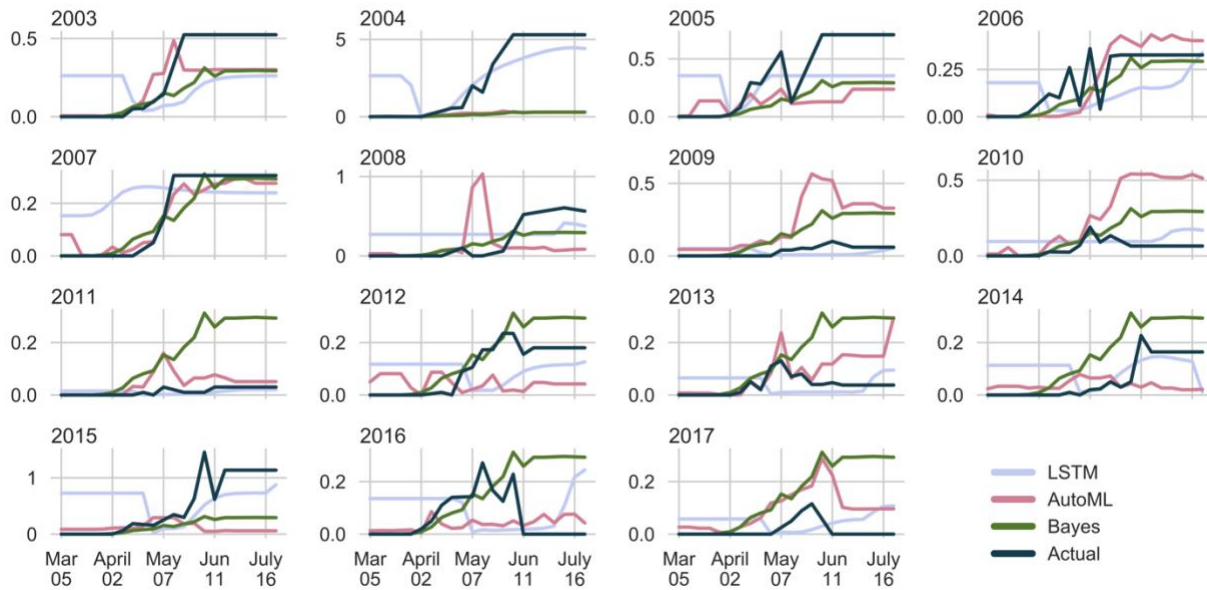

Figure 6. Forecasts from LSTM and AutoML methods when provided with full test data against the actual values and Naïve Bayes model.

Test data set up to and including April

Forecasts produced by both the LSTM and AutoML models did not produce sufficiently accurate forecasts, as evidenced in Figure 7. The models performed more accurately when given this test data compared to the full test data in 2006, 2008, and 2009.
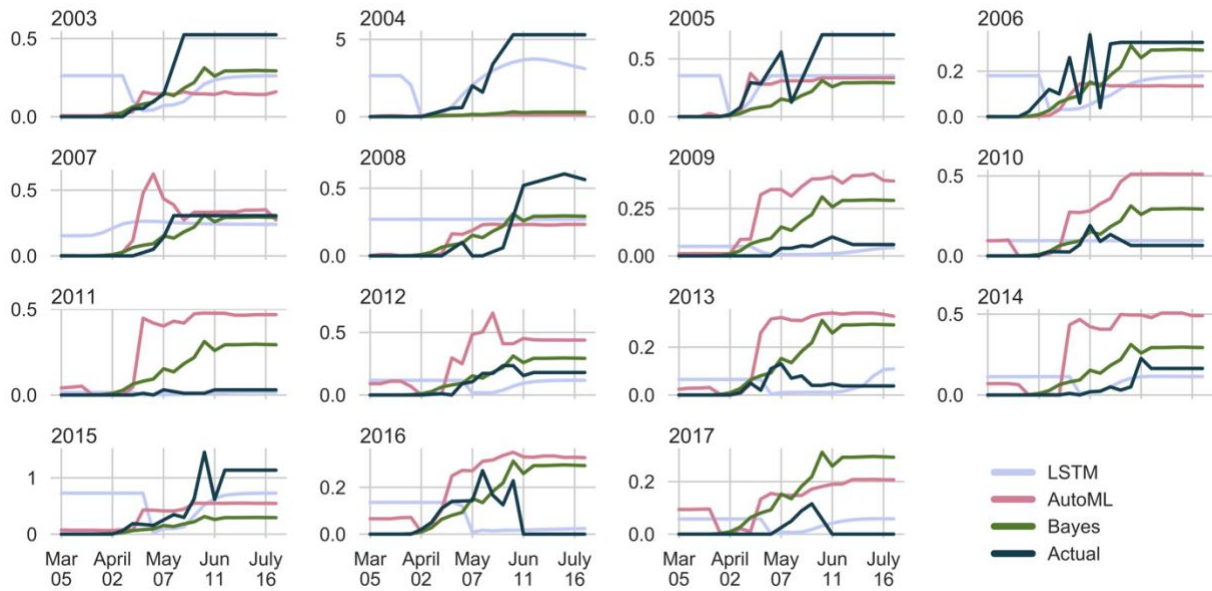

Figure 7: Forecasts from LSTM and AutoML, methods when provided with data from January to April against the actual values and Naïve Bayes model.

Test data set up to and including June

Forecasts produced by both the LSTM and AutoML models did not produce sufficiently accurate

forecasts, as evidenced in Figure 8. Similarly, to the April test data, some years had lower

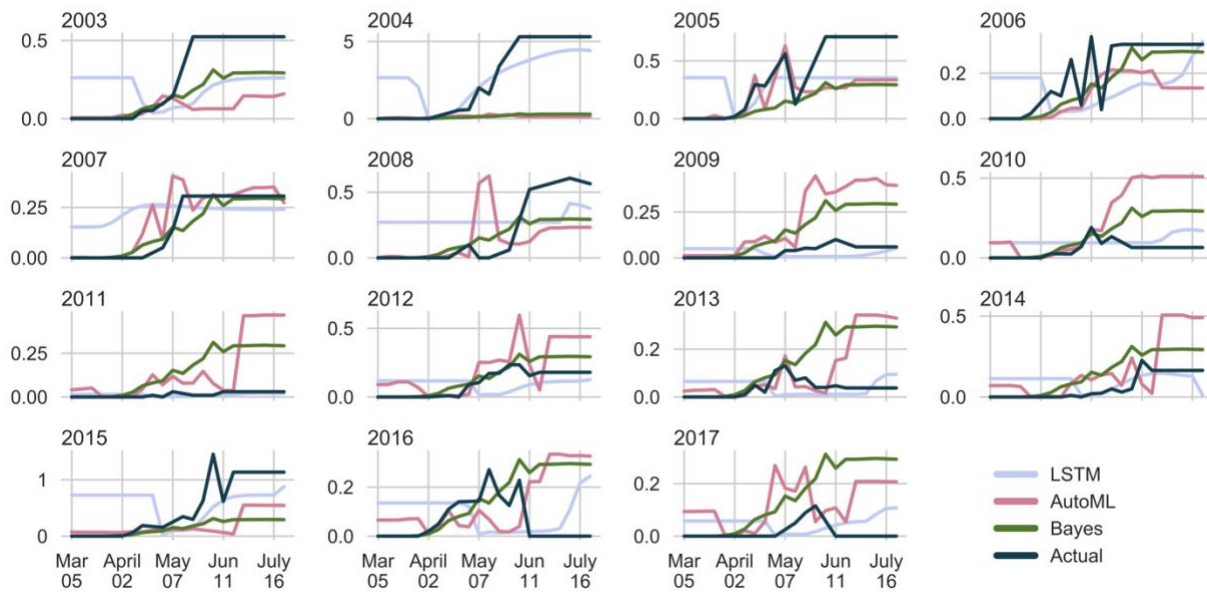spreads in predictions compared to the full test data.

Figure 8: Forecasts from LSTM and AutoML methods when provided with data from January to June

against the actual values and Naïve Bayes model.

Discussion

Overall, the initial goal of the project was not achieved. Although the fit models described observed years well, they could not forecast unobserved years. Progress was made in the definition of methods to be used for the application of Machine Learning models to ecological datasets and understanding the implications and follow-ups required to fulfil the goal. This is not unusual in the science of forecasting as highly accurate forecasts that we now take for granted, such as weather forecasting, had long histories of trial and error preceding the now-reliable forecasts that are easily and widely accessible (Serafin and Wilson 2000). In order to ensure that future trials of forecasting sea lice will build on these results, specific areas that require more research should be identified. Before attempting further forecasting on this particular system, some characteristics of the system and data should be explored such as the scale of the processes, the relevance of the collected variables in forecasting, and the system's inherent potential to be forecasted.

A concern throughout the modelling of this system was that of the frequency scale of the forecast and data. Input frequencies in the data sources ranged from daily to monthly, thus forecasting on a weekly basis was done to allow for this variance. It was understood that the down-sampling and imputation of values could have had unintended consequences on the trends in the data, but the main concern regarding timescale is of the forecasted processes. The natural history of sea lice is well described in terms of generation time, with generation times ranging from eight to sixteen weeks depending on temperature (Peacock et al. 2019). Because of this, the weekly frequency of forecasts was assumed to be suitable to the system. However, within this generation time the sea lice undergo several life stages. The major distinctions between life stages that were used for this study were between motile and non-motile sea lice, which is a

distinction that is made in much of the sea lice literature, but perhaps distinction between more life stages would have uncovered interactions that were not able to be described given the current variables.

Another concern in scale comes from the physical scale at which the system was modelled. When making forecasts, only the farms around the Broughton Archipelago were included as covariates in order to reduce the number of predictors used to fit the data. These farms acted as a standing stock of disease pressure, as past studies have shown a high correlation between local farms' sea lice abundance and wild sea lice abundance (Krkošek et al. 2011). Future modelling efforts should also explore the possibility of a larger regional scale effect. Sea lice can disperse on very large scales, and so models predicting the abundance of sea lice should also explore explicitly modelling the dispersal of lice on larger scales than that explored in this study.

Machine learning methods were chosen for this particular system because relative to many other wildlife disease datasets, the dataset available for this system is quite large and comprehensive. Bringing together all the data sources, the data used in this project described the abundance of sea lice in wild salmon populations for over 15 years and the abundance of sea lice in farmed salmon for over 7 years. Typically, machine learning performs best on large amounts of data and so this system seemed like a good fit for the methods used. Given the relative amount of data available here, one might think that machine learning is the obvious choice for describing this system, but this may be too narrow a point of view.

Many forecasting efforts in ecology do not rely on machine learning for their predictions, but on mathematical process models (Houlahan et al. 2017, Dietze et al. 2018). Machine learning, when found in the disease forecasting models, is mostly used to estimate parameters

and ensemble predictions from mathematical models rather than to produce predictions on its own. In the context of disease, it is interesting to make the comparison in the suitability of machine learning algorithms in modelling systems versus mathematical process models.

It is easy to see how a mathematical model is well-suited for describing a disease such as SARs (or more relevant in 2020, COVID-19); it is conceptually simple to break up the population in an area into susceptible, exposed, infected, etc. and from there to model these different groups. External variables might be less important than the contact between individuals and size of groups when making predictions on the disease's outcome and these kinds of transmission models have been extensively studied. However, in a system such as sea lice, the suitability of a mathematical model is less obvious. Macroparasites such as sea lice are much less studied and modelled and so a robust process model such as SEIR is not available for sea lice. The modelling done in this project was done with the assumption that external variables such as temperature and salinity would have a major impact on the abundance of sea lice in a given year, but the results from this study did not find these factors to have predictive power. Perhaps given the same data, a mathematical model could have been parametrised to produce accurate forecasts, however it is not possible to answer this question without speculation as it would require going back to the beginning of the project and exploring a different route of modelling. The suitability of mathematical versus machine learning models for this particular system is still undetermined, and further research should explore the advantages and disadvantages that each method provides.

Overall, more research should be done into the forecasting of sea lice. Starting with this project as a base it would be reasonable to add more data from different sources, explore the

scale of the system further, and compare different modelling approaches in order to determine if

useful insights can be achieved.

References

Barnes, J. 2015. Azure Machine Learning Microsoft Azure Essentials.

Bateman, A. W., S. J. Peacock, B. Connors, Z. Polk, D. Berg, M. Krkošek, and A. Morton. 2016. Recent failure to control sea louse outbreaks on salmon in the Broughton Archipelago, British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 73:1164–1172.

Böse, J. H., V. Flunkert, J. Gasthaus, T. Januschowski, D. Lange, D. Salinas, S. Schelter, M. Seeger, and Y. Wang. 2017. Probabilistic demand forecasting at scale. Proceedings of the VLDB Endowment 10:1694–1705.

Carbonneau, R., K. Laframboise, and R. Vahidov. 2008. Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research 184:1140–1154.

Crone, S. F., M. Hibon, and K. Nikolopoulos. 2011. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. International Journal of Forecasting 27:635–660.

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers, and E. P. White. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences of the United States of America 115:1424–1432.

Green, J. L., A. Hastings, P. Arzberger, F. J. Ayala, K. L. Cottingham, K. Cuddington, F. Davis, J. A. Dunne, M.-J. Fortin, L. Gerber, and M. Neubert. 2005. Complexity in Ecology and Conservation: Mathematical, Statistical, and Computational Challenges. BioScience 55:501.

Gu, J., Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen. 2018. Recent advances in convolutional neural networks. Pattern Recognition 77:354–377.

Houlahan, J. E., S. T. Mckinney, T. M. Anderson, and B. J. Mcgill. 2017. The priority of prediction in ecological understanding. Oikos:1–7.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning with Applications in R. Springer, New York.

Kaundal, R., A. A. Kapoor, and G. P. S. Raghava. 2006. Machine learning techniques in disease forecasting: A case study on rice blast prediction. BMC Bioinformatics 7:1–16.

Kimoto, T., K. Asakawa, M. Yoda, and M. Takeoka. 1990. Stock market prediction system with modular neural networks. Pages 1–6 1990 IJCNN international joint conference on neural networks.

Krkošek, M., B. M. Connors, A. Morton, M. A. Lewis, L. M. Dill, and R. Hilborn. 2011. Effects of parasites from salmon farms on productivity of wild salmon. Proceedings of the National Academy of Sciences of the United States of America 108:14700–14704.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos. 2018. The M4 Competition : Results , findings , conclusion and way forward The M4 Competition : Results , findings , conclusion and way forward. International Journal of Forecasting.

Marty, G. D., S. M. Saksida, and T. J. Quinn. 2010. Relationship of farm salmon, sea lice, and wild salmon populations. Proceedings of the National Academy of Sciences of the United States of America 107:22599–22604.

Molento, M. B., S. Bennema, J. Bertot, I. C. Pritsch, and A. Arenal. 2018. Bovine fascioliasis in Brazil: Economic impact and forecasting. Veterinary Parasitology: Regional Studies and

Reports 12:1–3.

Moritz, S., and T. Bartz-Beielstein. 2017. imputeTS: Time series missing value imputation in R. R Journal 9:207–218.

Nevo, S., V. Anisimov, G. Elidan, R. El-Yaniv, P. Giencke, Y. Gigi, A. Hassidim, Z. Moshe, M. Schlesinger, G. Shalev, A. Tirumali, A. Wiesel, O. Zlydenko, and Y. Matias. 2018. ML for Flood Forecasting at Scale. Pages 2–5 32nd Conferene on Neural Information Processing Systems.

Peacock, S. J., A. W. Bateman, B. Connors, M. A. Lewis, and M. Krkošek. 2019. Ecology of a marine ectoparasite in farmed and wild salmon. Pages 544–573 Wildlife Disease Ecology. Cambridge University Press, Cambridge, UK.

Peacock, S. J., A. W. Bateman, M. Krkošek, B. Connors, S. Rogers, L. Portner, Z. Polk, C. Webb, and A. Morton. 2016. Sea-louse parasites on juvenile wild salmon in the Broughton Archipelago, British Columbia, Canada. Ecology 97:1887.

Peacock, S. J., M. Krkosek, S. Proboszcz, C. Orr, and M. A. Lewis. 2013. Cessation of a salmon decline with control of parasites. Ecological Applications 23:606–620.

Petukhova, T., D. Ojkic, B. Mcewen, R. Deardon, and Z. Poljak. 2018. Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario , Canada. PLOS ONE:1–17.

Sainath, T. N., O. Vinyals, A. Senior, and H. Sak. 2015. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings 2015-Augus:4580–4584.

Serafin, R. J., and J. W. Wilson. 2000. Operational weather radar in the United States: Progress

and Opportunity. Bulletin of the American Meteorological Society 81:501–518.

Shaman, J., and A. Karspeck. 2012. Forecasting seasonal outbreaks of influenza. Proceedings of the National Academy of Sciences of the United States of America 109:20425–20430.

Suganthi, L., and A. A. Samuel. 2012. Energy models for demand forecasting — A review. Renewable and Sustainable Energy Reviews 16:1223–1240.

Teschner, F., and C. Weinhardt. 2015. A macroeconomic forecasting market. Journal of Business Economics 85:293–317.

Voyant, C., G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy. 2017. Machine learning methods for solar radiation forecasting: A review. Renewable Energy 105:569–582.

Wang, G., J. Hao, J. Ma, and H. Jiang. 2011. A comparative assessment of ensemble learning for credit scoring. Expert Systems with Applications 38:223–230.

Wang, Y., A. Smola, D. C. Maddix, J. Gasthaus, D. Foster, and T. Januschowski. 2019. Deep Factors for Forecasting.

Ward, E. J., H. Chirakkal, M. González-Suárez, D. Aurioles-Gamboa, E. E. Holmes, and L. Gerber. 2010. Inferring spatial structure from time-series data: Using multivariate state-space models to detect metapopulation structure of california sea lions in the gulf of California, Mexico. Journal of Applied Ecology 47:47–56.

Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: A meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos 123:652–661.

White, E. P., G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, and S. K. M. Ernest. 2019. Developing an automated iterative near-term forecasting system for an

ecological study. Methods in Ecology and Evolution 10:332–344.

Wilder, A. P., W. F. Frick, K. E. Langwig, and T. H. Kunz. 2011. Risk factors associated with mortality from white- nose syndrome among hibernating bat colonies. Biology Letters 7:950–953.

Zamora-vilchis, I., S. E. Williams, and C. N. Johnson. 2012. Environmental Temperature Affects Prevalence of Blood Parasites of Birds on an Elevation Gradient : Implications for Disease in a Warming Climate. PLOS ONE 7.

Zou, H., and T. Hastie. 2005. Erratum: Regularization and variable selection via the elastic net (Journal of the Royal Statistical Society. Series B: Statistical Methodology (2005) 67 (301-320)). Journal of the Royal Statistical Society. Series B: Statistical Methodology 67:768.