TUM Data Innovation Lab

Munich Data Science Institute

Technical University of Munich

&

Chair of Digital Agriculture (TUM School of Life Sciences)

Final report of project:

Soybean yield estimation with Machine Learning and Remote Sensing in Brazil

Authors	Ahmet Saltik, Leoni Kaiser, Vivian Haller, Xaver Wangerpohl
Mentor(s)	M.Sc. Malte von Bloh, M.Sc. Rogerio de Souza Noia Junior
Project Lead	Dr. Ricardo Acevedo Cabra (MDSI)
Supervisor	Prof. Dr. Massimo Fornasier (MDSI)

Aug 2022

Abstract

Brazil is the largest producer of soybean in the world and the year-to-year variation of its national production affects global food prices. The goal of this study was to develop a regional and national soybean wheat yield estimation system for Brazil. Twenty years (2001-2020) of readily available global gridded monthly climate and remote sensing indices across the most important soybean cultivated areas were used to build statistical models to estimate regional and national soybean yield, including six different machine learning-based approaches: *Random Forest, Gradient Boosted Trees, Artifical Neural Network, Long Short Term Memory, Bi Directional Long Short Term Memory, and Transformer.* The best performance to estimate national soybean yield were obtained by ANN and Bi-LSTM approaches, with an leave-one-year-out cross-validation rRMSE of less than 2%. As the statistical models employed monthly climate and remote sensing data from within a season, accurate national yield predictions are possible during the cropping season prior to the harvest. This predictability of the national soy production improves approaching the harvest. Estimating soybean yields in Brazil before harvesting can help to plan ahead for production failures due to adverse weather.

Contents

1   Introduction

Soybean is an important source of food, protein and oil and is considered as the primary world source of vegan protein [**Abraham et al.**, **2020**]. In the recent years, Brazil has become the world's largest soybean producer exceeding the United States. Brazil has doubled its annual production over the last decade, from 68 million tonnes (Mt) in 2010 to 138 Mt in 2021 [**Conab**, **2022**]. Soybean is one of the most important crops for Brazilian national economy as it is a top export commodity reaching its highest value in 2021 with 86 millions tons that generated almost 40 billion dollar value. Soybean cropland in Brazil is expected to increase more than 27% in the next 10 years by occupying lands with degraded pasture. With the new areas, the national soybean production is expected to exceed 150 Mt by 2029 [**Statista Research Department**, **2022**], [**Colussi and Schnitkey**, **2021**]. With continuous growth of human population, the global demand for food and particularly for protein are expected to increase by the end of the century, pressuring agricultural systems to produce up to 60% more food. Brazil contributed with 7% of global cereal exports in 2020, and is an important player in meeting current and future demand for food [**FAO**, **2022**]. However, climate variability which is mainly driven by the El Niño phenomenon [**Noia Junior et al.**, **2019**], has strongly affected soybean production in Brazil. In 2022, the country expected to produce $140Mt$ of soybeans, but severe droughts in southern Brazil reduced its production to less than $125Mt$ [**Conab**, **2022**]. Similar national soybean production failures due to widespread droughts also occurred in 2012 and 2016 [**Conab**, **2022**]. These extreme events affected the global trade of agricultural commodities, causing food price spikes. Soybean yield estimation systems could be useful to anticipate national soybean production failures, which are usually only known after harvest. These systems could help policy makers and food traders to plan ahead any disruption of the food market caused by climate variability, stabilizing food security [**Voora et al.**, **2020**], [**Toloi et al.**, **2021**].

Crop yield prediction is a challenging task as crop yield depends on many different factors such as climate, weather and soil. Methods based on machine learning emerge as an alternative to crop simulation models as they can determine patterns and gain knowledge from several different features using historical data. Various machine learning algorithms have been used including regression models, Random Forests and Gradient boosting trees [**VanKlompenburg et al.**, **2020**]. The most used model for crop yield prediction are Artificial Neural Networks (ANN). [**Barbosa dos Santos and Ferreira dos Santos**, **2021**] for example, used a a Multi-Layer Perceptron to predict soybean yield with seasonal average temperature, relative humidity and solar radiation. Also, Cai et al. [**Cai et al.**, **2019**] integrates satellite and climate data to predict wheat yield at regional scale using different machine learning approaches.

Accordingly, this study aims to estimate regional and national soybean yield and production in Brazil, using twenty years (2001-2020) of climate and remote sensing data. Several machine learning approaches including Linear Regression Model(LM), Random Forest(RF), XGBoost(XGB), Long Short-Therm Memory(LSTM), Artifical Neural Network (ANN) as well as a transformer based approach are compared for this time series prediction.

First an exploratory data analysis on the climate and remote sensing data was applied to select which climate variables and vegetation indices should be included. In section 3 the different model approaches are introduced and given further information on their setup. In section 4 The models are compared by their performance on estimating soybean yield, their spatial performance on state level as well global level using different metrics including coefficient of determination (r^2), Root Mean Squared Error (RMSE) and Root Relative Mean Squared Error (rRMSE). The models with the best performance were tested for an in-season analysis to also identify their soybean yield estimation performance during the soybean season in Brazil.

2 Material and Methods

2.1 Soybean crop, climate and remote sensing data

Brazil is divided into 27 federative units (including 26 states and the federal district), of which 20 produce soybean. This study was carried out for the six largest soybean producing states in Brazil, which together account for 82% of national production. The states studied are: Mato Grosso (MT, which accounts for 33% of the national soybean production in 2022), Goias (GO, 13%), Parana (PR, 10%), Rio Grande do Sul (RS, 7%), Mato Grosso do Sul (MS, 7%), Minas Gerais (MG, 6%) and Bahia (BA, 5%) (figure 1). Historical soybean yield and harvested area data from 2001 to 2020 (totaling more than 26,000 yield observations) were collected from the Brazilian Institute of Geography and Statistics [**IBGE**, **2022**] for over 1400 municipalities. Around 3.5% of the yield data was missing. The missing data were filled using a simple variant of the k-nearest-neighbour approach (described in the subsection 2.1.1 Data Completion). Long term daily weather data from 2001 to 2020 consisting of maximum and minimum temperature, solar radiation and rainfall were obtained from the NASA POWER service (Prediction Of World-wide Energy Resources) [**NASA**, **2021**]. Weather data were collected for each of the studied municipalities based on their central geographical coordinates (centroid). The satellite data originated from NASA's Moderate-resolution-Imaging-Spectroradiometer (MODIS) satellites, which have been available since February 2000 with a resolution of 500m-1000m. Only the bands 1-16 and 31-32 were used. The data was obtained from the Google Earth Engine platform [**Gorelick et al.**, **2017**]. The following remote sensing indices were used: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Chlorophyll Vegetation Index (CVI) and Green Leaf Index (GLI). A landcover mask was applied to the satellite imagery for each individual county and only those pixels classified as "cropland" were used for further processing. The landcover mask was created in 2016 and is based on the Sentinel-II satellites from the European Copernicus mission [**Buchorn et al.**, **2020**]. The remaining pixels for each county were then averaged.
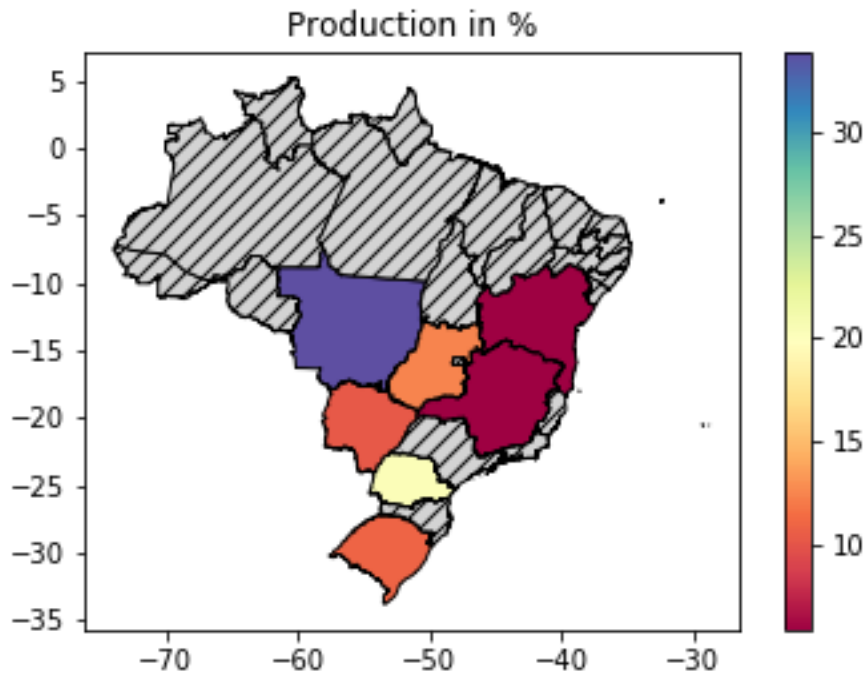
Figure 1: National Soybean Production Share for the different states of Brazil
.

The soybean yield data in each municipality were trend-corrected with a cubic spline function as suggested by [**Guarin et al.**, **2020**] to remove the long-term trend typically due to technological improvements within the study period.

2.1.1 Soybean crop yield and remote sensing data completion

The spectral data from MODIS as well as the soybean yield data contained missing values. The climate data had no missing values. Different approaches to fill the missing spectral and soybean yield data, were used as follow:

Non-yield data: Climate/Spectral data has been generated by sampling from suitably fitted distributions. Missing values in the spectral data were in turn filled by values sampled from these distributions. The features of the spectral data followed either a normal distribution or simple bimodal distributions. The shape of these distributions varied notably for different climate zones and years. As a consequence, 2-component Gaussian mixture models have been used as target distributions for each combination of an input feature other than precipitation, year and climate zone, while for precipitation an exponential distribution was chosen.
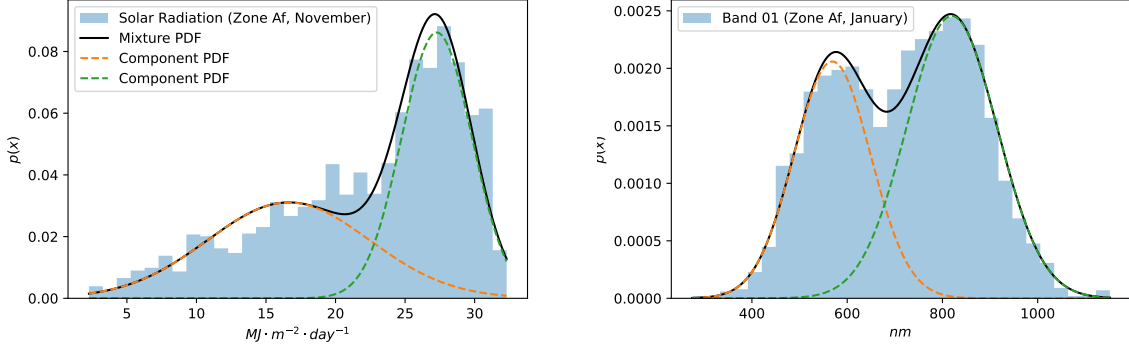
Figure 2: Two distributions resulting from the fitting process used for data completion and seasonal forecasting, both 2-component Gaussian mixture models. The first corresponds to solar radiation in November while the second corresponds to the frequency measurements of the first MODIS band in January, both restricted to climate zone *Af* of the Köppen classification.

Yield data: A different approach has been used for completing the missing soybean yield data. It requires less compute time than the previous one. The approach is based on the following case distinction for each combination of year i and county c, going from 2001 to 2020:

1. If $i > 2001$ and a proper yield value $y_{c,i-1}$ exists, let $\mathcal{N}_{c,i}$ denote the set of at most k counties $d \neq c$ with proper yield values $y_{d,i}$ in the same climate zone as c which are closest to c in terms of the 2-norm of the respective climate features (the climate data has been normalized prior to determining the neighbourhood). If there are fewer than k such neighbours in $\mathcal{N} := \mathcal{N}_{c,i} \cap \mathcal{N}_{c,i-1}$, apply step 2 instead. Form the average ratio

$$\gamma := \frac{1}{|N_{c,i}|} \sum_{d \in \mathcal{N}} \frac{y_{d,i}}{y_{d,i-1}}$$

   of the counties' in years i and $i-1$ and set the missing value $y_{c,i} := \gamma \cdot y_{c,i-1}$.

2. Define $y_{c,i}$ as the average of the yields $y_{d,i}$ in $\mathcal{N}_{c,i}$.

2.2   Climate and Vegetation Indices

Oceanic Niño Index   Climate anomalies and volatility (e.g. abnormally low precipitation) regularly cause crop yield losses all over the world. Water deficit and extreme temperatures are of particular significance in limiting soybean yields in Brazil [**Noia Junior et al., 2019**], and their year-to-year variability is mainly caused by a phenomenon known as *El Niño Southern Oscillation* (ENSO) [**Noia Junior et al., 2019**], which has two main phases the *El Niño* and *La Niña* phases, defined as follow [**NOAA, 2022**]:

- *El Niño* brings above average temperature waters to the central and eastern tropical Pacific, sometimes all the way to the coast of South America. At the surface, the

prevailing easterlies (trade winds) slow down or sometimes even reverse. Rainfall increases over the warm waters in the central-east tropical Pacific, but decreases over Indonesia and the western Pacific. During *El Niño* events there is an increased likelihood for floods in southern Brazil and droughts in the northern regions of the country [**Noia Junior et al., 2019**]

- *La Niña* results in an influx of below average temperature waters to the same region while the prevailing easterlies intensify. Its effects are adverse to the effects attributed to *El Niño* events, i.e., floods in northern Brazil and droughts in the south of the country.

For illustration, the large-scale *El Niño* event of 2015/2016 resulted in off-season yield reductions of 4% for soybean and 31% for maize, respectively. These losses were mainly associated with states in the northeast of Brazil (e.g. Bahia). The central part of Brazil does generally not seem to be as affected by the effects of ENSO [**Noia Junior et al., 2019**].

The *Oceanic Niño Index* aims at quantifying the drasticity of the ENSO phases by means of the running 3-month average of sea surface temperatures in the east-central tropical Pacific between 120°-170°W (close to the International Dateline) and whether they are warmer or cooler than average [**NOAA, 2022**]. The corresponding data obtained from [**CPC, 2022**] contains one such measurement for each month, independent of county and state.

2.2.1 Weather based Indices

Due to the climate in Brazil, the most restricting factors for the yield are heat and drought. The following indices were used to help the models to detect when such events occur:

Hot Days Hot days are defined as a day where the highest recorded temperature exceeds $32°C$ and the sum of these days per month was used as an additional input feature.

Lowrain Days Similar to the hot days, this index counts the number of days with less than $0.5mm/m^2$ of precipitation per month.

Evapotranspiration Evapotranspiration (ET) is the simultaneous transfer of water to the atmosphere by evaporation (i.e. the process by which liquid water is converted into water vapor and removed from the evaporation surface) and transpiration (i.e. the process which consists of the vaporization of liquid water in plant tissues and its vapor removal into the atmosphere) [**Allen et al., 1998**]. It is measured in mm/m^2 .

Following [**Noia Junior et al., 2019**], the *Heargreves-Samani* method was used to to estimate the *potential evapotranspiration (ETP)*. Furthermore using the *water soil balance* proposed by [**Thornwaithe et al., 1957**], the *actual evapotranspiration (ETA)* was computed. Together, the Agricultural Reference Index for Drought, also known as ARID, described by [**Woli et al., 2012**], was added to the data.

Standardized Indices In contrast to the indices shown above, these indices use past weather data to quantify how far present data deviates from the ordinary.
Three standardized indices were calculated:

- *1-month Standard Precipation Index*

- *3-month Standard Precipation Index*

- *1-month Standard Temperature Index*

Past monthly precipation and monthly average Temperature data form the basis for the computation. The indices are constructed as follows:

1. A *gamma* distribution is fit to the historic data (for the 3-month index the average of the past three months is used).

2. The *gamma* distribution is fit to a *standard normal* distribution.

3. According to the above steps, find a transformation that maps the historic data such that it follows a standard normal distribution.

4. Use this transformation on the new data to get the index

A value of -2 for the *1-month Standard Precipation Index* would roughly correspond to a month drier than approximately 95% of the measured months.

2.2.2 Spectral Based Indices

The following four related indices based on the spectral MODIS data have also been included as input features to the models:

- Normalized Difference Vegetation Index (NDVI),

- Enhanced Vegetation Index (EVI),

- Chlorophyll Vegetation Index (CVI) and the

- Green Leaf Index (GLI)

The computation of these indices is based on the first four spectral bands recorded by MODIS. All of them can be regarded as measures for the existence of vegetation and individually can be used to estimate more intricate features such as the plant biomass, the vegetation density and so forth.
Consider the NDVI as an example: It is computed from measurements of the first two bands $\frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}}$ and has values between -1 and 1 if these measurements are positive and finite. Areas harbouring dense vegetation will typically have a positive NDVI between 0.3 and 0.8.
The idea for this is as follows: Live green plants absorb solar radiation in the PAR spectral region (photosynthetically active radiation). Leaf cells re-emit solar radiation in the near-infrared spectral region corresponding to roughly 50% of the incoming solar energy in order to avoid overheating. As a consequence, such plants appear rather dark in the PAR spectral region and quite bright in the near-infrared spectral region, which is why the difference in the nominator is going to be large.

2.3 Statistical Data Exploration and In-season analysis

2.3.1 Metrics for soybean yield modeling performance

The performance of the soybean yield estimating models was tested with a leave-one-year-out cross validation. First a model was built using each of the seven machine learning aproaches, with 19 of the 20 years to select the best subset of variables, and then it was tested on the excluded years (Cross-validation). This process was repeated for each year for a total of 20 iterations. The mean absolute percentage error (MAPE) and RMSE were then calculated based on the predicted yield of the test year and the corresponding observed national soybean yield. The relative root mean square error (rRMSE) was also calculated, as the RMSE divided by the average observed trend correct yield in each country.

2.3.2 Mapping the Yield to the Production

The provided data includes the yield (kg/ha) and soy cropping area (ha). Since the yield is a relative measurement, it provides a good value to use as the response. After the estimation a simple multiplication with the area provides the soy production values

2.3.3 In-Season Analysis

The Soybean cropping season in Brazil starts in September with sowing, and finishes in March with the harvest. Additionally to the simulation task, an in-season analysis was performed to identify the soybean yield estimation performance of the models during the soybean season in Brazil. This was carried out at the end of each month during the cropping-season from August (no live data) to March (data from the whole growing cycle).

There are two basic approaches to generating results with missing data:

- Generating (forecasting) the missing data and using the original model (Compare the data generation of the *Data Completion* (2.1) section).

- Truncating the regressors, such that only the known values remain, i.e. drop all the values from March and train a new model with only the truncated data matrix as regressors for the February forecast,

Both methods were applied and will be discussed in the corresponding parts of the report.

2.4 Explainable AI

Deep Neural Networks are a blackbox approach by nature. In other words, it is hard to explain how all the individual neurons work together to arrive at the final output. In general, it isn't even clear what any particular neuron is doing on its own. That means it is basically ignored how it works and just giving it input and getting output.

SHAP Values (SHapley Additive exPlanations), proposed by Scott M. Lundberg and Su-In Lee [**Lundberg et al. (2017)**], is one of the most used ways of explaining the model and understanding how the features of the data are related to the outputs. The

original idea behind the SHAP values are based on the Shapley Values in Game Theory which was proposed by the Nobel laureate mathematician Lloyd Shapley. It is a method derived from coalitional game theory to provide a way to distribute the "payout" across the features fairly. To each cooperative game it assigns a unique distribution (among the players) of a total surplus generated by the coalition of all players [**Shapley (1951)**].

More clearly, Shapley values are based on the idea that the outcome of each possible combination (or coalition) of players should be considered to determine the importance of a single player. *"What Shapley does is quantifying the contribution that each player brings to the game. What SHAP does is quantifying the contribution that each feature brings to the prediction made by the model"* [**Mazzanti (2020)**].

The following SHAP formula reported in the article [**Lundberg et al. (2017)**] can be obtained:

$$SHAP_{feature}(x) = \sum_{set:feature \in set} \left[|set| \times \binom{F}{|set|} \right]^{-1} \left[Predict_{set}(x) - Predict_{set \setminus feature}(x) \right]$$

where x is the data observations, F is the total number of feature, set is the all possible combinations of feature subsets that contains the corresponding feature for the calculation of the SHAP value, Predict($\cdot$) is the prediction function of the corresponding model.

3 Soybean yield estimation models description

This section covers the approaches used to estimate soybean yield (t/ha). Each model consists of 20 submodels with 1 year as test set and the other 19 years split in train and validation sets (leave one year out cross validation LOYOCV). This approach aims to get good generalisation properties, since the yields from one year could be strongly correlated. The predictions are made on the county level, i.e. a prediction is made for every county, which are then accumulated to estimate at the state and national level. The trend correction is applied beforehand and readded after the raw prediction to get the final values.

3.1 Baseline Models

The baseline Models are simple models, that help to show advantages of the more sophisticated models. All results should be compared to the baseline in order to get meaningful statements about each models performance.

Farmer's Approach

A very natural and simple approach to estimate yield is to take an average of the yield over the last five years. One special property of the Farmer's Approach is, that it can be calculated before sowing and thus will also serve as a baseline model for the *In-Season Forecasting* section of the report. Note that since it is also used as a regressor, trend correction is applied.

Linear Models

The next to simplest approach consisted of fitting a linear model to the yield observations. Initially, all features were included with no regard to colinearity. To take state or climate zone information into account, two variants have been tested:

- The first variant uses a single ('global') regressor in which the corresponding information is included via a one-hot-encoding.

- The second variant uses multiple regressors, i.e. one per climate zone or state.

During testing it became apparent that at least for the simpler models a single global regressor would produce less noisy predictions. Here the imbalanced distribution of the number of observations per state or climate zone plays a major role. Since some of the features are strongly correlated, a reduction of the feature space seemed appropriate. Furthermore it was noticed that the absolute values of the coefficients for the one-hot encoded state/zone variables in case of the single global regressor tend to be several orders of magnitude larger than the remaining ones when using standard multilinear regression. It is for these reasons that Ridge regression has been used instead, which should address both issues at the same time. In case of the single global regressor on states, choosing the regularization parameter $\lambda = 1230$ yielded lowest average RMSE.

3.2   Decision Tree based Models

A decision tree is a type of model which can both be applied to classification and regression problems. The underlying idea involves repeatedly subdividing the feature space by axis-parallel hyperplanes (splits) such that the resulting partition of the input data has low error with regard to some measure (e.g. impurity in case of classification). Decision trees are typically grown according to some heuristic criteria as the problem of determining an optimal tree for arbitrary input is $\mathbb{NP}$-hard. Such criteria can e.g. be:

- Grow the tree until a maximum depth has been reached.

- Only perform a split as long as the error decrease at the corresponding node is large enough.

A random forest is an ensemble model consisting of multiple decision trees to be trained in parallel while each tree was trained on a varying subset of training data and represents different learned patterns. Their individual predictions on a given input are aggregated in some way (e.g. averaging in case of regression or taking the majority vote in case of classification) in order to obtain a single prediction.

It is a common initial approach to use a random forest regressors when predicting yield (see e.g. [**Cai et al.**, **2019**] and [**Barbosa dos Santos and Ferreira dos Santos**, **2021**]) next to a standard multilinear regressor, as decision trees are usually better at handling non-linearity. Similarly to the linear models, a single global random forest regressor as well as individual ones for states/zones were trained and the comparison between the two variants yielded more or less the same results as for the linear models. 100 trees were used for the single global regressor, while each of the state- and zonewise regressors consisted of 50 trees. The average RMSE over all years encountered during cross validation actually turned out to be larger for the random forest than for the results of the Ridge regression.

3.3   Gradient boosting based Models

This Introduction roughly follows the one from the XGBoost documentation [**XGBoost**]. This package was also used to implement the model.

Like for random forests, the underlying structure of gradient boosted trees is a decision tree ensemble. The difference lies in the way that the trees are generated. Tree generation in gradient boosting somewhat resembles a gradient descent method, where $\beta_t = \beta_{t-1} - s\nabla f(\beta_{t-1})$. Analogously a new tree is generated, such that its addition to the ensemble decreases the loss.

Hyperparameter Selection

Since gradient boosted trees are prune to overfitting, a good selection of hyperparameters is crucial to archive reasonable results. The tuning was done with the package *hyperopt* [**Bergstra2013**]. Every submodel got the same hyperparameters and a custom objective function based on the accumulated loss for all the models was implemented to achieve similar performance between the models (the train data was split again to get independent results). The best parameters were:

- $max\_depth = 4$

- $n\_estimators = 70$

- $learning\_rate = 0.1$

- $\alpha = 3$

- $\gamma = 4.5$

- $\lambda = 1.5$

max_depth is the maximum depth of each tree, $n\_estimators$ the amount of trees and α, γ, λ regularisation parameters similar to the parameters from ridge and lasso regression.

Model Improvements

The model performed well with regard to the yield prediction, but significantly worse in terms of estimating the production. Weighing the samples according to their cultivation areas was the first approach to improve upon this, but this proved too severe of a weighing, since the areas differ by several orders of magnitude for some countys. Weighing according to the $log(area)$ proved to increase the performance for the production while getting similar results for the yield.

One other approach was to include the *Farmer's approach* as a regressor to somewhat measure differences between countys, since differences in soil, soy, and infrastructure like irrigation also affect the yield. This also improved the performance of the model.

3.4 Neural Network based Models

3.4.1 LSTM & Bi-LSTM

As a state-of-the-art approach, a Long Short-Term Memory (LSTM) was proposed to capture the dependencies between the time steps. A LSTM can learn to bridge minimal time lags in excess of 1000 discrete-time steps by using input, output and forget gates [**Hochreiter and Schmidhuber, 1997**]. It can be applied from one way to another (from left to right) or bidirectionally for a sequential data. Two different LSTM models were set up, i.e., from left to right and bidirectional which can capture the information for both directions to make yield predictions.

To do so, the data were reshaped by converting it to a 3-d tensor in order to make it compatible with the input shape of the corresponding LSTM models. Since 19 years, 32 features, and 1243 counties were available to train the model, the input shape and the output shape look $\mathbb{R}^{1243x19x31}$ and $\mathbb{R}^1$ respectively. Two hidden layers were implemented each having a hidden size of 128 and 64 respectively. The layer numbers have been decided after some experiments to get the best result. To improve the model stability and performance, standardization, i.e. standard scaling, was used for both training and test data by using *scikit-learn* library. For the model implementation, Keras API for the TensorFlow library was used.The model structure of the LSTM and Bidirectional LSTM can be seen in the figure (3). The only difference between these two models is that normal

LSTM layers for the LSTM model were used, whereas bidirectional LSTM layers for the Bidirectional LSTM.
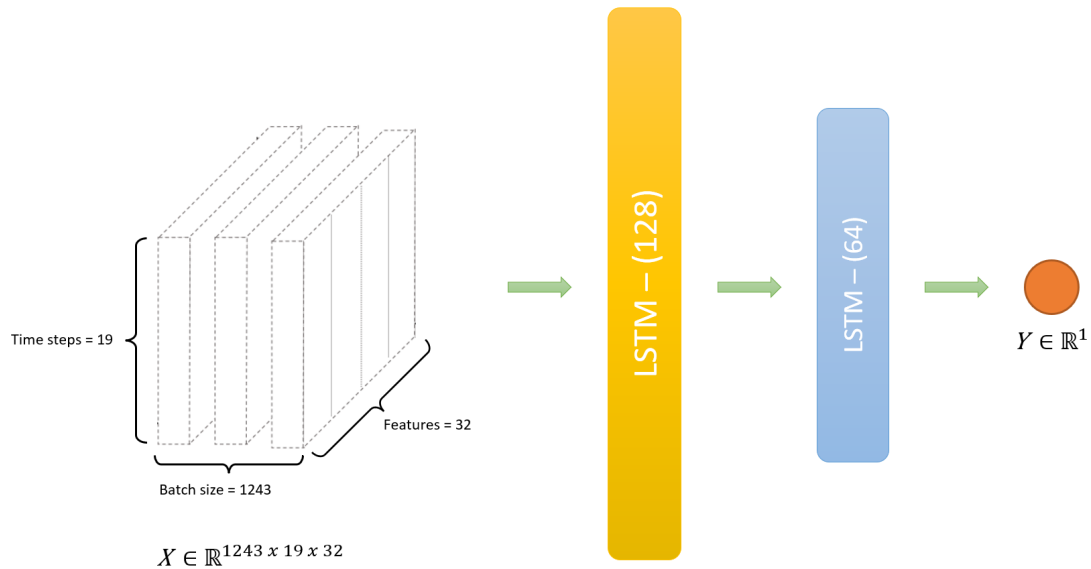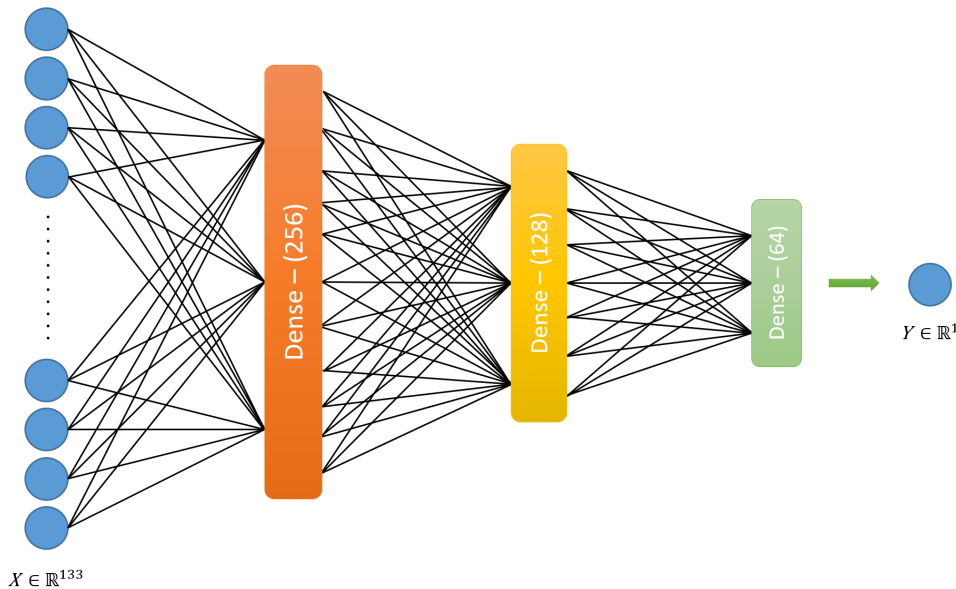

Figure 3: Model structure of Global LSTM and Global Bidirectional LSTM. *Global Bidirectional LSTM has exactly the same structure by just changing the LSTM layers with the Bidirectional ones.*

To validate the model, training set was split by 20% for each year. To get rid of overfitting, there was used a 20% Dropout rate for each hidden layer as well as 10 training step as an early stopping. The details regarding the chosen hyperparameters and the loggings of the loss curves are displayed in table 1, and in the Appendix 5.

Loss:	MSE
Optimizer:	Adam
Learning rate:	10^{-3}
Batch size:	19
Metric:	Validation loss
Early Stopping:	10
Validation split:	20 %
Dropout:	20 % (for the hidden layer)

Table 1: Hyperparameters for Global LSTM and Global Bi-LSTM

3.4.2 ANN

As an alternative, a fully connected Artificial Neural Network (ANN) [**LeCun et al.**, **2015**] was tested under the assumption of independence between the time steps. The data were reshaped by casting to be able to extract monthly information for climate and satellite image data. The input shape and the output shape look $\mathbb{R}^{133}$ and $\mathbb{R}^1$ respectively. For the model structure, two hidden layers were used with 256 and 128 units respectively. However, the model was not sufficiently able to learn in a proper time, that has been detected by the validation loss plots. Afterwards one more hidden layer was added which has 64 units to make the model more robust. To improve the model stability and performance, standardization, i.e. standard scaling, was used for both training and test data by using *scikit-learn* library. For the model implementation, Keras API for the TensorFlow library was used. The model structure of the ANN can be seen in the figure (4).

Figure 4: Model structure of Global ANN

To validate the model, the training set was splitted into 80% of the data for training and 20% for validation for each year. To reduce overfitting, Dropout with a rate of 30% for each hidden layer was implemented as well as 10 training step as an early stopping. All the details about the hyperparameters can be seen in the table 2. Another important hyperparameter that dramatically affected the model performance is the learning rate. The first used learning rate of 10^{-3} - as in the LSTM - showed fluctuation in the validation loss curve. For that reason the validation loss was not able to converge to the training loss as seen in the figures 5. Therefore, the learning rate was decreased from 10^{-3} to 10^{-4} to reduce the fluctuation effect on the validation loss. Several loss plots for comparison can be seen in the Appendix 5.

Loss:	MSE
Optimizer:	Adam
Activation:	ReLu
Learning rate:	10^{-4}
Batch size:	32
Metric:	Validation loss
Early Stopping:	10
Validation split:	20 %
Dropout:	30 % (for each hidden layer)

Table 2: Hyperparameters for Global ANN

3.5 Transformer based Models

This model architecture is based on the transformer approach of [**Vaswani et al.**, **2017**]. This novel approach of a transformer model entirely relies on a self-attention mechanism that computes a representation of the sequence by relating different positions to search the most relevant input timesteps. To apply this approach to timeseries, the structure needs to be adapted as shown in figure 5.
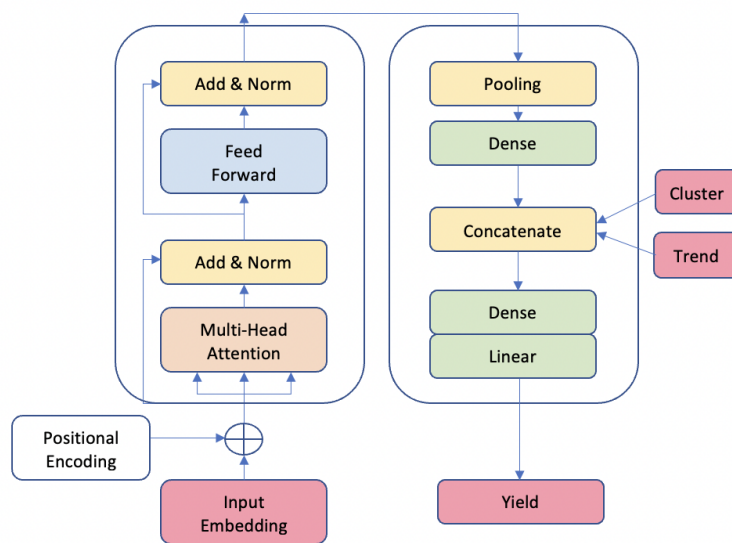
Figure 5: Model Structure of the Transformer based model. *Combines the Transformer encoder with an perceptron classification head.*

Here, only the transformer encoder part is used followed by a (Multi-) layer perceptron classification head. The transformer encoder block maps an input sequence to a sequence of continuous representations and can be stacked multiple times. It includes a Multi-Head Attention that performs an attention mechanism several times in parallel where the single outputs are concatenated and linearly transformed. This is followed by an fully connected feed-forward network position wise and for each of these layers exists a residual connection followed by normalization. As only the yield for a particular input sequence is predicted, instead of the decoder a set of dense layer is used. To apply the model for specific climate clusters, this information is incorporated as well as a yield trend.

As in the LSTM /ANN model before, for validation of the model, the overall training data was splitted into a 80% training set and 20% validation set. To prevent overfitting, a dropout of 0.2 was selected after the Multi Head Attention as well as in the following Feed Forward network. Also, in the Multi -Layer Perceptron head a Dropout of 0.1 was established. Furthermore, 4 attention heads were used with a size of 256. 4 transformer encoder blocks are stacked together and the whole model is trained on batches of size 64 on 100 epochs, with the option of Early Stopping if the validation loss is not decreasing in 5 steps. The details regarding the chosen hyperparameters and the loggings of the loss curves are displayed in Figure 28, and in the Appendix 5.

4   Results

4.1   Performance of Machine learning approaches to estimate soybean yield in Brazil

Farmers usually estimate expected yield by averaging the obtained yield of the last 5-years (farmer-method). When applying this farmers approach to estimate national soybean yield, the rRMSE was of 12.6%:
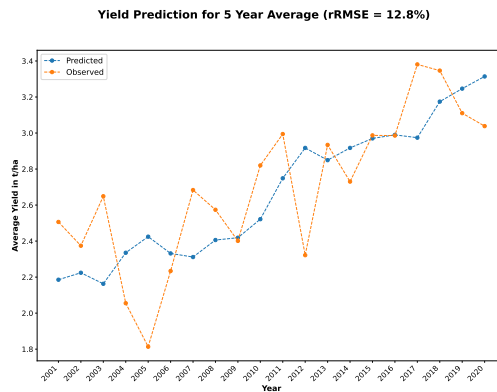

Figure 6: 5-years average prediction

The farmer-method will not capture extreme soybean yield losses caused by extreme climate events. Here, different machine-learning approaches were tested to estimate soybean yield, particularly in years with extreme low yields.

The different machine-learning approaches showed different performance in estimating soybean yield, as presented in figure 7. For the Linear Model the values are clustered around the actual yield value with some significant outliers. Here, the RMSE over all years of the predicted yield counts 465 kg/ha. Also, around half of the observed variation can be explained by the linear model, as indicated by the r^2. The linear model presented a rRMSE value of 16.7%. Random forest model and XGBoost model presented a better performance for estimating soybean yield, compared to the Linear Model. Random forest model showed an rRMSE of 16.1% and an r^2 of 0.61, whereas XGBoost model showed an rRMSE of 15.3% and an r^2 of 0.64. The Time-Series approaches of LSTM and Bidrectional-LSTM, performed poorer compared to the above-mentioned models, with an rRMSE of around 18%.

With the ANN model over than 65% of the observed yield was explained. The ANN model showed a RMSE value of 420 kg/ha.

The best estimations of soybean yield in Brazil were obtained with the transformer based model. This model presented the highest r^2 (76%) and the lowest rRMSE (12.6%), among all the machine-learning approaches tested. However, this model over or under-estimate soybean yield according to the year of cultivation.
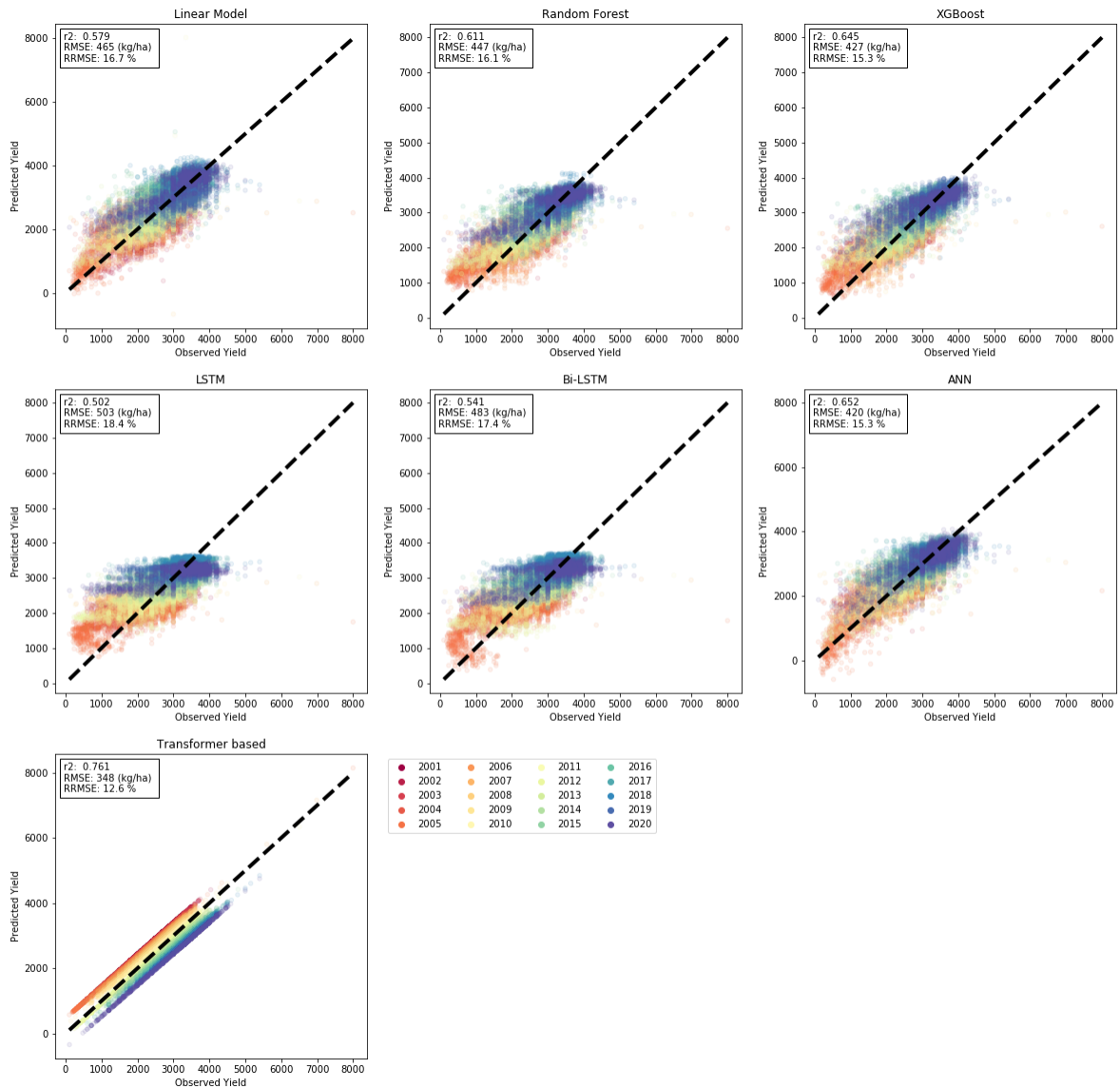
Figure 7: Comparison of the actual observed yield(kg/ha) and the predicted yield(kg/ha) of each model over all years. r^2, RMSE and rRMSE are used as metrics.

4.2   Spatial Performance

The models performance were also evaluated on state level, as shown in figure 8. In order to facilitate the visualization of the results, only the best tree models of ANN, Transformer based and bi-lstm model together with the linear model are shown. In Bahia state, the models presented their worst performance in estimating soybean yield, with an rRMSE varying from 11.2% to 16.2% according to the modeling approach used. Overall, the best model performance states were obtained in Parana and Rio Grande do Sul states. In these states the ANN approach showed an rRMSE of around 5%, with RMSE of 150 kg/ha and r^2 of 0.87. The transformer based model captures the soybean yield year-to-year variability but it overestimates yield from 2001 o 2010 and underestimate it from 2011 to 2020. The linear model does not capture some of the soybean yield year-to-year

variability (both low and high yields), as from 2001 to 2003 in Mato Grosso and Mato Grosso do Sul states. Similar patters are shown by the bi-lstm model from 2001 to 2007 in Minas Gerais and Mato Grosso states.
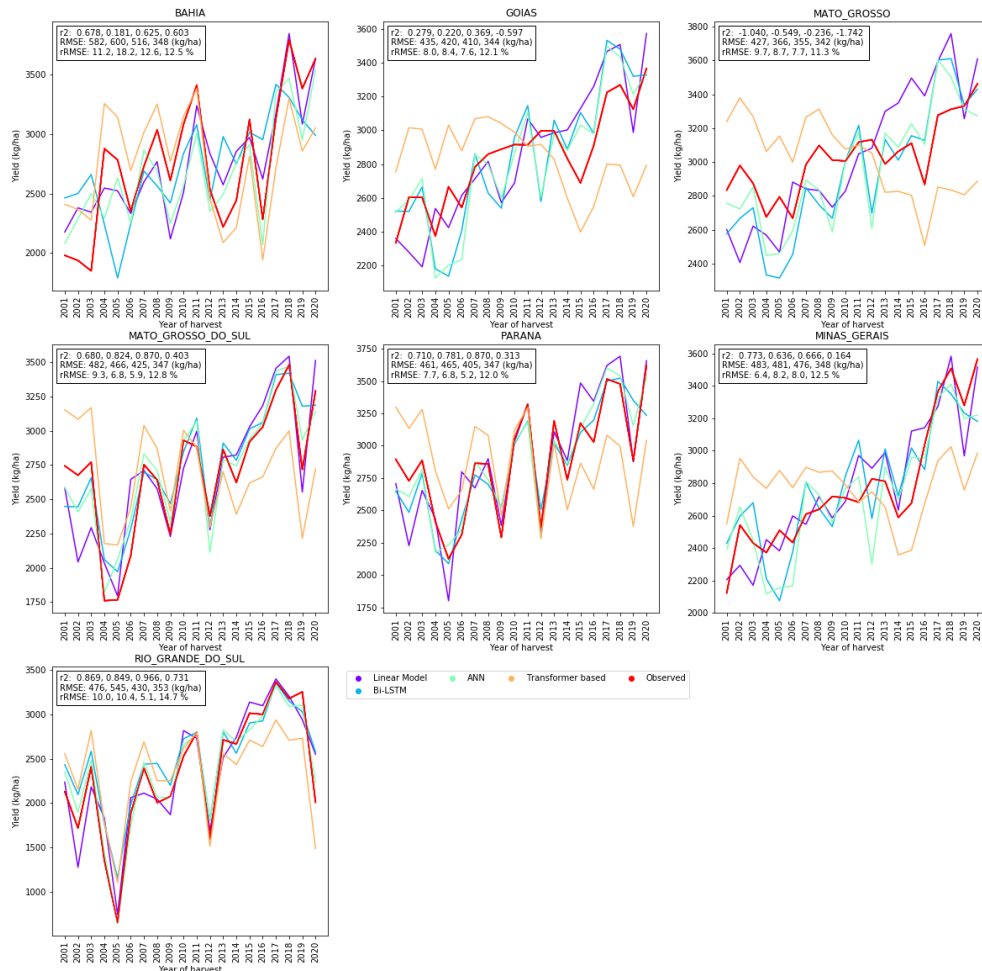
Figure 8: Statewise Comparison of the observed soybean yield(red lines, kg/ha) and the estimated yield (kg/ha) of selected models from 2001 to 2020. *The three best performing models of ANN, Transformer based and bi-lstm in relation to the linear model.*

The rRMSE obtained from each model varied according to the state (figure 9). The best soybean yield predictions were obtained in the states of central Brazil, including Mato Grosso, Goias and Mato Grosso do Sul. From all models, the transformer model presented the smallest rRMSE variation between the states, with rRMSE of around 10% in all states. In comparison, the rRMSE varied from 10% to more than 20% between the states for all the other models tested.
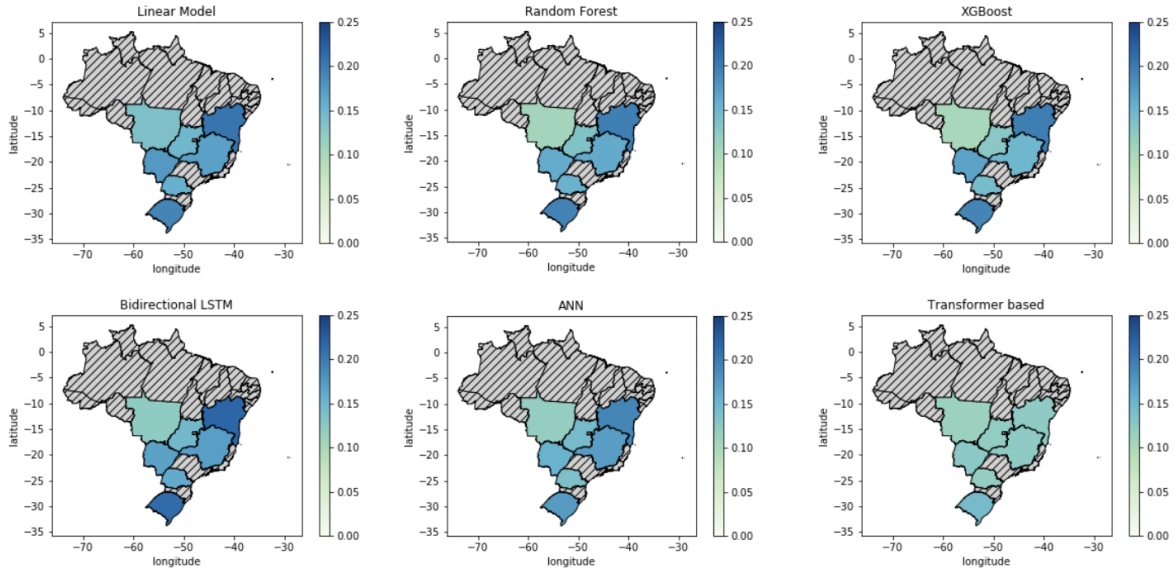
Figure 9: graphical interpretation of the Root Relative Mean Squared Error per State

To analyze the models performance in estimating extremes high or low soybean yields, the models performance were compared during 2015/ 2016 cropping season, which was a very strong El Niño year, and 2010/2011 a strong La Niña year [**Golden Gate Weather Services**, **2022**]. Here, the LM, RF and XGB overestimate soybean yield by around 450 kg/ha, particularly in Mato Grosso and Parana. LSTM/ Bi-LSTM and ANN also overestimate soybean yield but by around 150 kg/ha. The transformer based model totally under-predicts this year with around 400 kg/ha. In contrast, for the La Niña season of 2010/2011 all of the models overestimate soybean yield for most of the states, but not for the high yields of Bahia and Parana.
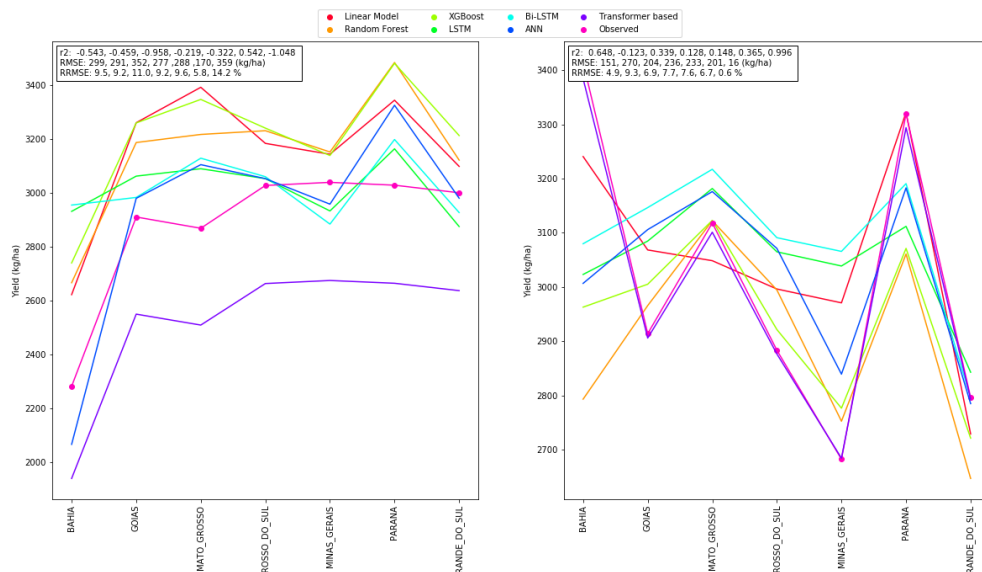
Figure 10: Comparison of model's prediction in climatic extreme years. *left: very strong El Niño season 2015/2016, right: strong La Niña season 2010/2011*

4.3 Evaluation of production on state and national level

The observed and estimated soybean yield were combined with observed cropping areas to evaluate the models performance at both state and national levels. For the state level, the models are compared based on their prediction for Mato Grosso (MT) as it is, stated in section 2.1, accountable for largest amount on the national soybean production, shown in figure 11. For this state, the Bi-LSTM and ANN predict the production of MT with an rRMSE value of around 2.5%.
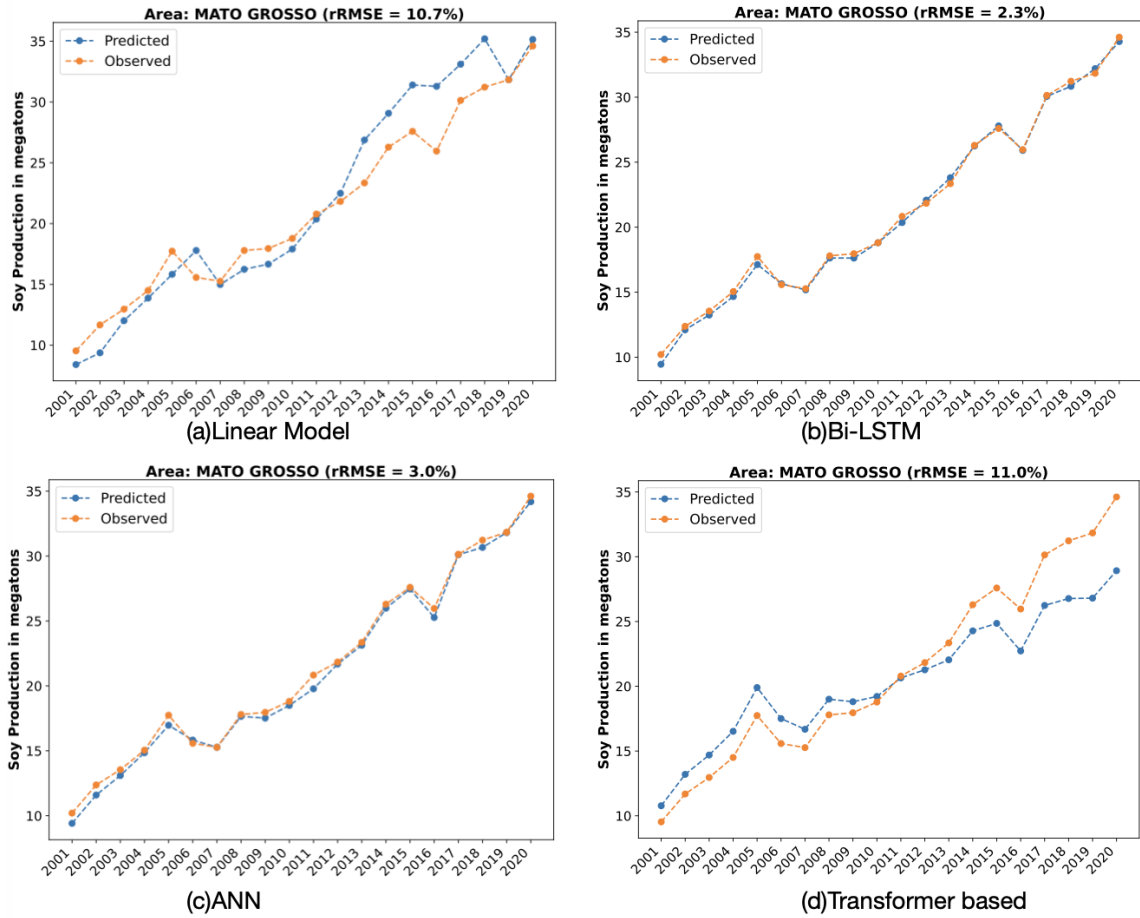
Figure 11: Production on state level on the example Mato Grosso (MT). *MT accounts for 33% of the national soybean production in 2022.*
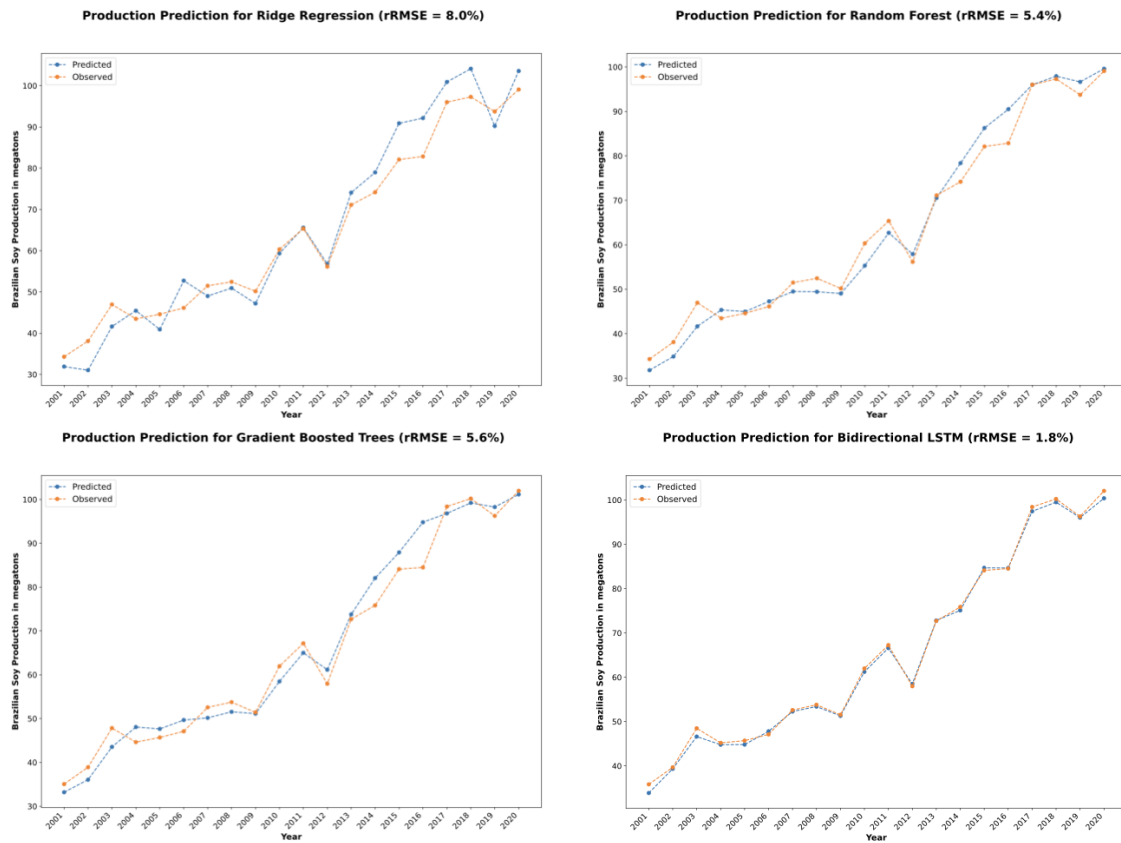
Figure 12: Evaluation of Production on national level comparing all approaches. *LM, RF, XGB and Bi-LSTM.*

On national level, The ANN and Bi-LSTM outperform the other approaches with an rRMSE less than 2%, as shown in figure 12 and 13. The linear model, random forest and XGBoost model were also able to capture most of the significant points with an rRMSE of 5-6%. The transformer based model overestimates soybean yield in the first years from 2001 to 2010 and underestimates from 2011 to 2020, with an rRMSE of 12%.
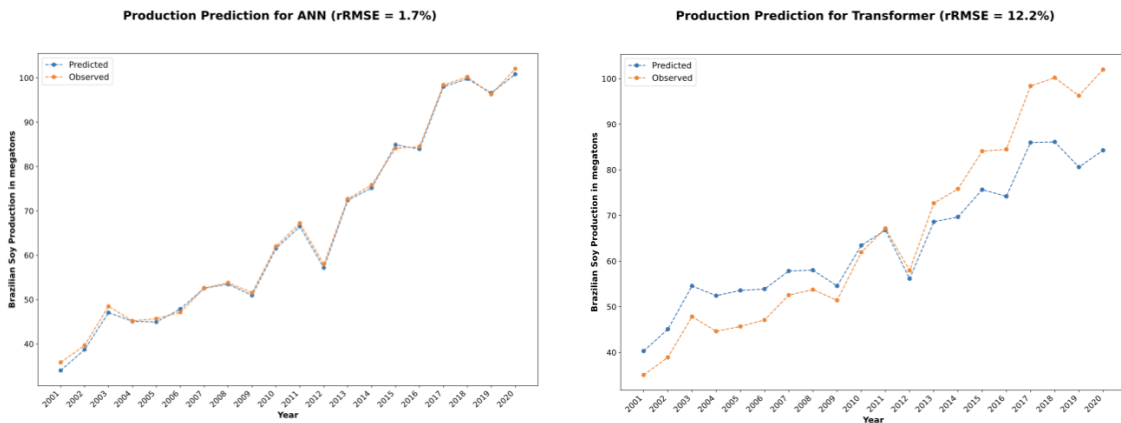
Figure 13: Evaluation of Production on national level comparing all approaches. *ANN and Transformer based.*

4.4   In-season forecast

For the *Gradient Boosted Trees* model, both approaches introduced in the *In-season Analysis* section were modeled. To evaluate the results, the observed and predicted Production values for each year of the LOYOCV were accumulated on the whole dataset i.e. for Brazil. Then the rRMSE was calculated over the years to measure the deviation in the accumulated models, which is shown in the following figure.
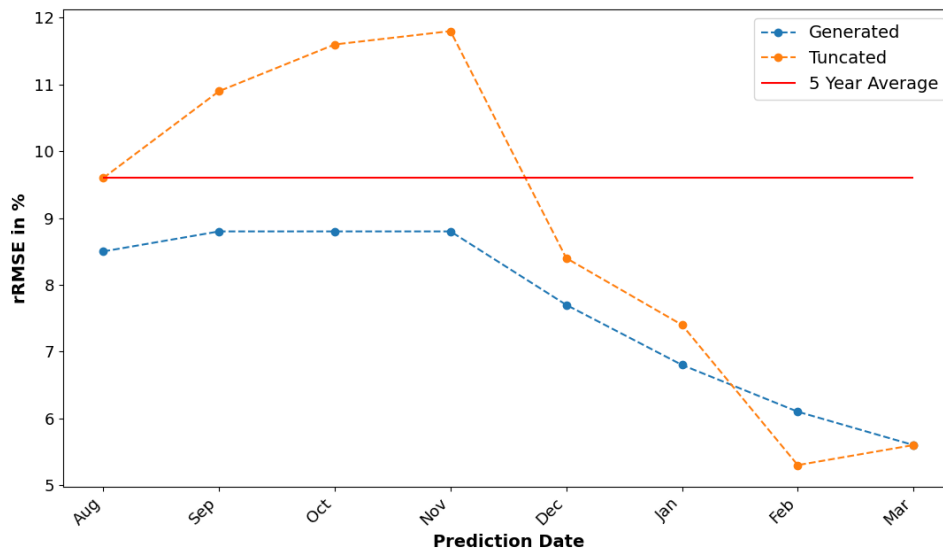
Figure 14: Evaluation of the relative root mean square error of the soy production in Brazil for the above compared approaches of generating predictions at a given time (end of month) with the *5 Year average (Farmer's approach)* as a baseline.

As a quick recap, the *Generated* label in the figure describes the performance of the standard model using generated data inplace of the missing values. For the *Truncated* label, new models were trained (with adjusted hyperparammeters) at each timestamp with the truncated dataset corresponding to the until then collected data. The results of this analysis were compared to the trend corrected average of the last 5 years (Farmer's-approach). The rRMSE didn't decrease for both approaches until December, but even in August, the Model with the generated values was able to outperform the 5 year average by approximately 1 percentage point and thus offers strictly better predictions at any given time. Curiously the overall best performance was observed from the truncated model in February.

4.5   SHAP Values

By applying SHAP approach, as stated in the section (2.4), to our global ANN model for each year, the following figure (15) results have been produced by taking the average of each year.
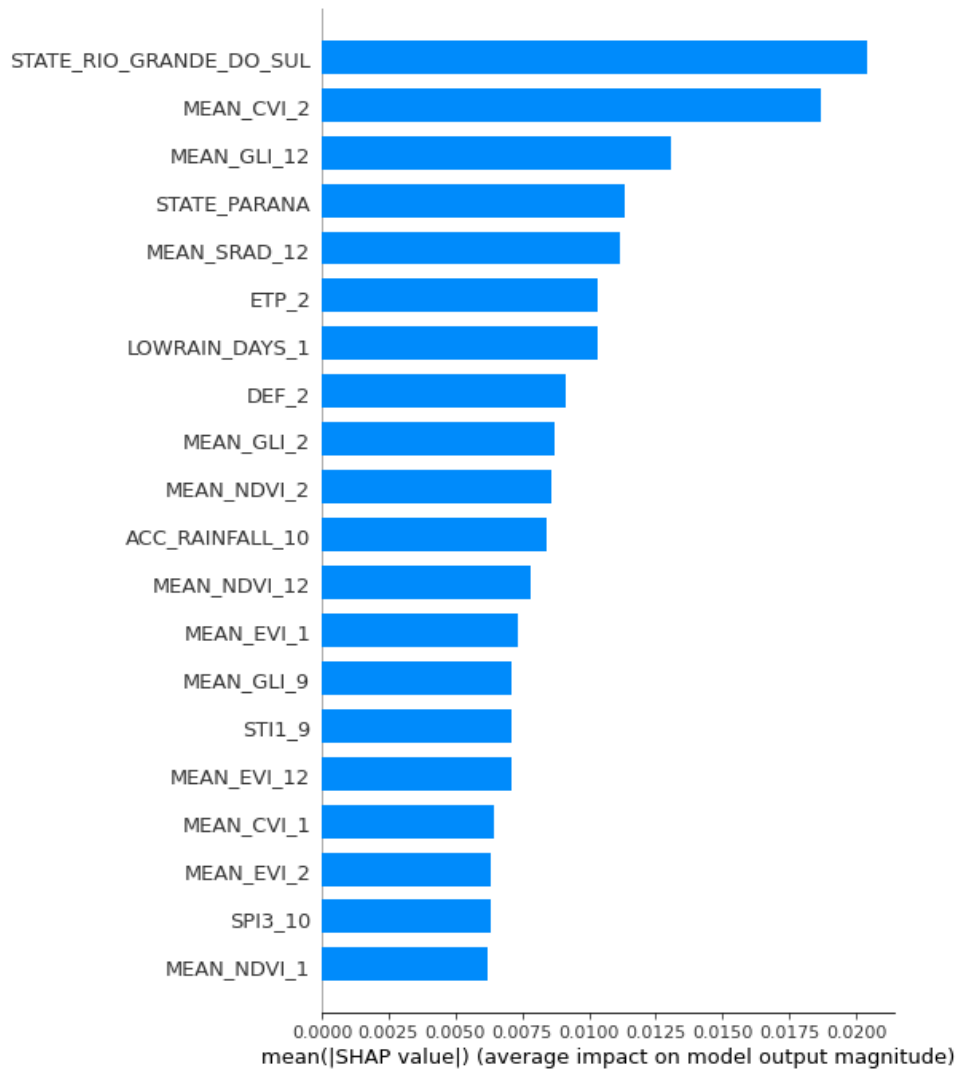
Figure 15: The absolute mean of the SHAP values for the Global ANN model for each year.

The ANN model tends to use the features state Rio Grande Do Sul and state Parana as the most important. This is consistent with the data since it contains the highest number of observations on these two states. Another remarkable point is the SHAP values are mostly higher for the months December, January and February. This is observed from the produced results that CVI, GLI, NDVI and EVI indices are the most important features for the corresponding months mentioned in the previous sentence.

5   Discussion

In this project, several modelling approaches were tested to estimate soybean yield in the seven states with the highest soybean production in Brazil. These states contribute 82% to the total Brazilian soybean production and make up 28% of the global production. The modelling approaches include both simple and widely used machine-learning approaches such as linear and tree-based models, as well as recurrent neural networks and more complex approaches like transformer models. Most of these models, such as linear and tree-based models have been extensively tested for crop yield estimation ([**Jeong et al.**, **2016**]). However, machine learning approaches such as the transformer models have been rarely used for grain crops yield estimation.

All these models were built based on features extracted or aggregated from publicly available datasets, most notably climate-related and spectral measurements. With that, this study can be replicated for any crop in any country of the world.

To simulate soybean yield, climate and remote sensing data was combined from over 26,000 yield observations, across more than 1200 municipalities from 2001 to 2020. Despite the large number of observations, the model performance was still limited by the size of the dataset, particularly in some regions. The states of Bahia and Minas Gerais had the worst predictability from the models (figure 9), because they have the least yield observations and consequently the smallest training sets. Overall, the statewise rRMSE ranges from 1.7% to 12%, according to the modelling approach.

Once built,the time required for training the yield prediction models in the prescribed leave-one-year-out cross validation (LOYOCV) ranged from a few seconds (linear, XG-Boost) to 30-45 minutes (random forest, ANN) to 3-4 hours (Bi-directional LSTM, Transformer). When testing the models at national level, the lowest error was achieved by the ANN (0.7% rRMSE), followed by the bi-directional LSTM (1.1% rRMSE). The tree-based models obtained an rRMSE of 5.3% for Random Forest and of 6.1% for XGBoost. Ridge regression presented an rRMSE of 4.9%. The highest rRMSE value of 13.3% was obtained for the transformer-based model. In comparison, the simple five-year average produced an rRMSE of 12.8%.

At state level, the performance of the different modelling approaches in estimating soybean yield varies more strongly. In the following, the maximum rRMSE over all states will be abbreviated by max-rRMSE. The ANN still shows the best performance (max-rRMSE $\leq$ 2.3%), followed by the bi-directional LSTM (max-rRMSE $\leq$ 2.3%). These models also presented a good performance when applied to simulating yield of other crops ([**Zhu et al.**, **2021**]). Although the performance of *Random Forest* models for estimating crop yields has been extensively tested at regional, national and global levels [**Jeong et al.**, **2016**], it resulted in a high max-rRMSE of $\leq$ 25.7% for the given setting.

Regarding the in-season forecasting results (figure 14), a reliable predicition and thus a significant improvement over the Farmer's approach is given from December onwards. Nevertheless are the increasing errors from March to November and peak performance in

February for the truncated approach indicators, that the impact of some regressors was not accurately identified (the Farmer's approach was included as a regressor and should theoretically provide an upper bound). Together this indicates that either careful selection of the regressors or a larger dataset might provide a way to improve on the models.

Many studies have compared different approaches to estimate crop yield at regional and national level [**Nigam et al.**, **2019**], [**Lobell "& Asseng**, **2017**]). However, comparing seven different machine-learning approaches as a key basis for a soybean yield estimation system in Brazil is a novelty. The biggest difficulties encountered when building an estimation model are the size, format and availability of reliable datasets for agriculture. Despite of this, some of these models are rarely used in the agricultural context which creates a barrier for new users. However, in this project the knowledge of different professionals in agriculture, climate and machine-learning were brought together and successfully implemented and tested all these models to estimate soybean yield.

Multiple regional and national soybean yield forecast models have been introduced using available monthly climate as well as remote sensing data. Besides, the potential for timely in-season yield prediction was also explored and found that decent prediction performance can be obtained about two months before harvest in Brazil. Forecasting soybean production in Brazil before harvest can contribute to some extent to mitigating the instability of potential global food disruptions due to adverse weather.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*. p.. 5998–6008).

[2] Gangopadhyay, T. (2019). Deep Time Series Attention Models for Crop Yield Prediction and Insights.

[3] Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. 9 (8): 1735 – 1780. https://doi.org/10.1162/neco.1997.9.8.1735

[4] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep Learning. *Nature*. 521, 436 - 444. https://doi.org/10.1038/nature14539

[5] Abraham, E.R., Mendes dos Reis, J..G, Vendrametto, O., Oliveira Costa Neto, P.L., Carlo Toloi, R., Souza, A.E., Oliveira Morais, M. (2020). Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production.*Agriculture*. 10(10):475. https://doi.org/10.3390/agriculture10100475

[6] Statista Research Department (2022). Brazil: soybean export volumn 2006-2021. Retrieved July 27, 2022 from https://www.statista.com/statistics/721215/soybeans-export-volume-brazil/

[7] Colussi, J. & Schnitkey,G. (2021). Brazil Likely to Remain World Leader in Soybean Production (2021). *farmdoc daily* (11):105, Department of Agricultural and Consumer Economics,University of Illinois at Urbana-Champaign.

[8] Fukase, Emiko & Martin, Will (2020). Economic growth, convergence, and world food demand and supply. *World Development*. Vol. 132.

[9] Voora, V., Larrea, C., Bermudez, S. (2020). Global Market Report: Soybeans. *Sustainable Commodities marketplace series 2019*. Retrieved July 27, 2022 from https://www.iisd.org/system/files/2020-10/ssi-global-market-report-soybean.pdf

[10] Toloi, M.N.V., Bonilla, S.H., Toloi, R.C., Silva, H.R.O., Naas, I.d.A (2021). Development Indicators and Soybean Production in Brazil. *Agriculture*. 11, 1164. https://doi.org/10.3390/agriculture11111164

[11] Van Klompenburg, T., Kassahun, A., Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*. 177, 105709.

[12] Guarin, J.R., Asseng, S., Martre, P. & Blinznyuk, N. (2020). Testing a crop model with extreme low yield from historical district records. *Field Crops Research*. 249, 107269.

[13] NASA (2021). The POWER Project. *NASA Prediction Of Worldwide Energy Ressources*. Retrieved July 28, 2022 from https://power.larc.nasa.gov

[14] IBGE (2022). Producao Agricola Municipal. Retrieved July 28, 2022 from https://sidra.ibge.gov.br/tabela/1612

[15] Conab (2022). Soja. Retrieved July 28, 2022 from https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras/itemlist/category/911-soja

[16] NOAA (2022). Climate Variability: Oceanic Nio Index. Retrieved July 28, 2022 from https://www.climate.gov/news-features/understanding-climate/climate-variability-oceanic-ni%C3%B1o-index

[17] CPC (2022). Historical El Nio / La Niña episodes (1950-present). Retrieved July 29, 2022 from https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

[18] de Souza Noia Junior, R., Fraisse, C., Karrei, M. & Perondi, D. (2020). Effects of the El Niño Southern Oscillation phenomenon and sowing dates on soybean yield and on the occurrence of extreme weather events in southern Brazil. *Agricultural and Forest Meterology*. 290, 108038.

[19] FAO (2022). FAOSTAT: Crops and livestock products. Retrieved July 28, 2022 from https://www.fao.org/faostat/en/data/TCL

[20] Gorelick, N., Hanche, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 202

[21] Buchhorn, M., Lesiv, M., Tsendbazar, N., Herold, M., Bertels, L. & Smets, B. (2020). Copernicus Global Land Cover Layersâ"Collection 2. *Remote Sensing*.https://doi.org/10.3390/rs12061044

[22] Thornthwaite, C.W. & Mather, J.R. (1957). Instructions and tables for computing potential evapotranspiration and the water balance. *Laboratory in Climatology*. 10, No. 3.

[23] Woli, P., Jones, J.W., Ingram, K.T. & Fraisse, C.W. (2012). Agricultural Reference Index for Drought (ARID). *Agronomy Journal*. 104: 287-300. https://doi.org/10.2134/agronj2011.0286

[24] Noia Junior, R. S., Fraisse, C. W., Cerbaro, V. A., Karrei,M. A. Z., Guindin,N. (2019). Evaluation of the Hargreaves-Samani Method for Estimating Reference Evapotranspiration with Ground and Gridded Weather Data Sources. *Applied Engineering in Agriculture*. 35(5): 823-835. https://doi.org/10.13031/aea.13363

[25] Allen, R.G., Pereira, L.S., Raes, D. & Smith, M. (1998). Crop evapotranspiration guidelines for computing crop water requirements. *FAO Irrigation and drainage paper* 56. Food and Agriculture Organization.

[26] Cai Y., Guan K., Lobell D., Potgieter A., Wang S., Peng J., Xu T., Asseng S., Zhang Y., You L., Peng B (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches, *Agricultural and Forest Meteorology*, https://doi.org/10.1016/j.agrformet.2019.03.010.

[27] Barbosa dos Santos, V. & Ferreira dos Santos, A., (2021).Machine learning algorithms for soybean yield forecasting in the Brazilian Cerrado. *Journal of the Science of Food and Agriculture*.https://doi.org/10.1002/jsfa.11713

[28] Golden Gate Weather Services (2022).El Niño and La Niña Years and Intensities. Retrieved July 28, 2022 from https://ggweather.com/enso/oni.htm

[29] de Oliveira Aparecido, Lucas Eduardo et al. (2022). Predicting coffee yield based on agroclimatic data and machine learning. *Theoretical and Applied Climatology*. http://hdl.handle.net/11449/230415.

[30] Diennevan Souza Barbosa, B. et al. (2021). UAV-based coffee yield prediction utilizing feature selection and deep learning. *Smart Agricultural Technology*. https://doi.org/10.1016/j.atech.2021.100010

[31] Li, X., Metsis, V., Wang, H. & Ngu, A. (2022). TTS-GAN: A Transformer-based Time-Series Generative Adversarial Network. https://doi.org/10.48550/arXiv.2202.02691

[32] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*

[33] Lundberg, S. & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *arXiv*. https://doi.org/10.48550/arxiv.1705.07874

[34] Shapley, Lloyd S. (1951 August 21). Notes on the n-Person Game – II: The Value of an n-Person Game. *Santa Monica, Calif.: RAND Corporation*. https://www.rand.org/content/dam/rand/pubs/researchmemoranda/2008/RM670.pdf

[35] Mazzanti, Samuele (2020, Jan 4). SHAP Values Explained Exactly How You Wished Someone Explained to You. *Towards Data Science*. https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30

[36] Srivastava, A.K., Safaei, N., Khaki, S., Lopez, G.M., Zeng, W., Ewert, F., Gaiser, T., Rahimi, J. (2021). Comparison of Machine Learning Methods for Predicting Winter Wheat Yield in Germany. *ArXiv*, abs/2105.01282.

[37] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PloS* one, 11(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

[38] Prasad, N.R., Patel, N.R., & Danodia, A. (2021). Cotton Yield Estimation Using Phenological Metrics Derived from Long-Term MODIS Data. *Journal of the Indian Society of Remote Sensing*. 49, 2597 - 2610.

[39] Zhu, C., Tian, J., & Li, P. (2021). Research on Grain Yield Prediction Model Based on Contribution Multiplier and Bidirectional LSTM Neural Network. *2021 2nd International Conference on Artificial Intelligence and Information Systems (ICAIIS 2021)*. Association for Computing Machinery. Article 77, 1 â"- 7. https://doi.org/10.1145/3469213.3470278

[40] Nigam, A., Garg, S., Agrawal, A. & Agrawal, P. (2019). Crop Yield Prediction Using Machine Learning Algorithms. *2019 Fifth International Conference on Image Information Processing (ICIIP)*. pp. 125-130. https://doi.org/10.1109/ICIIP47207.2019.8985951.

[41] Lobell, D.B. & Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*. Vol 12,1.

[42] Feng, L., Wang, Y., Zhang, Z. & Du, Q. (2021). Geographically and temporally weighted neural network for winter wheat yield prediction. *Remote Sensing of Environment*. Vol. 262, p. 112514. https://doi.org/10.1016/j.rse.2021.112514

Appendix

Table S1 − Loss Plots for ANN with learning rate 10^{-4}
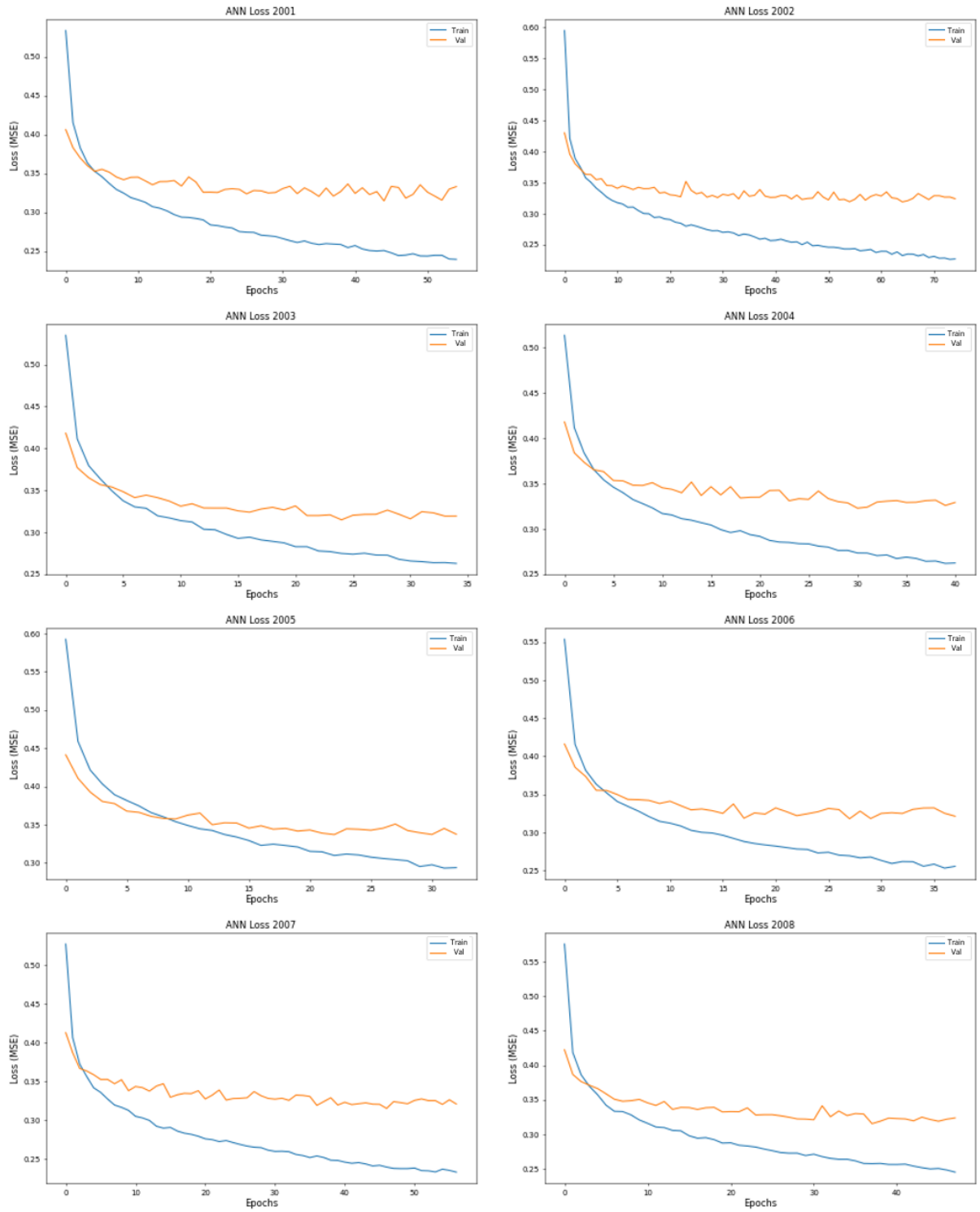
Figure 16: Training and validation loss plots of the Global ANN model with learning rate 10^{-4} for 2001 − 2008.
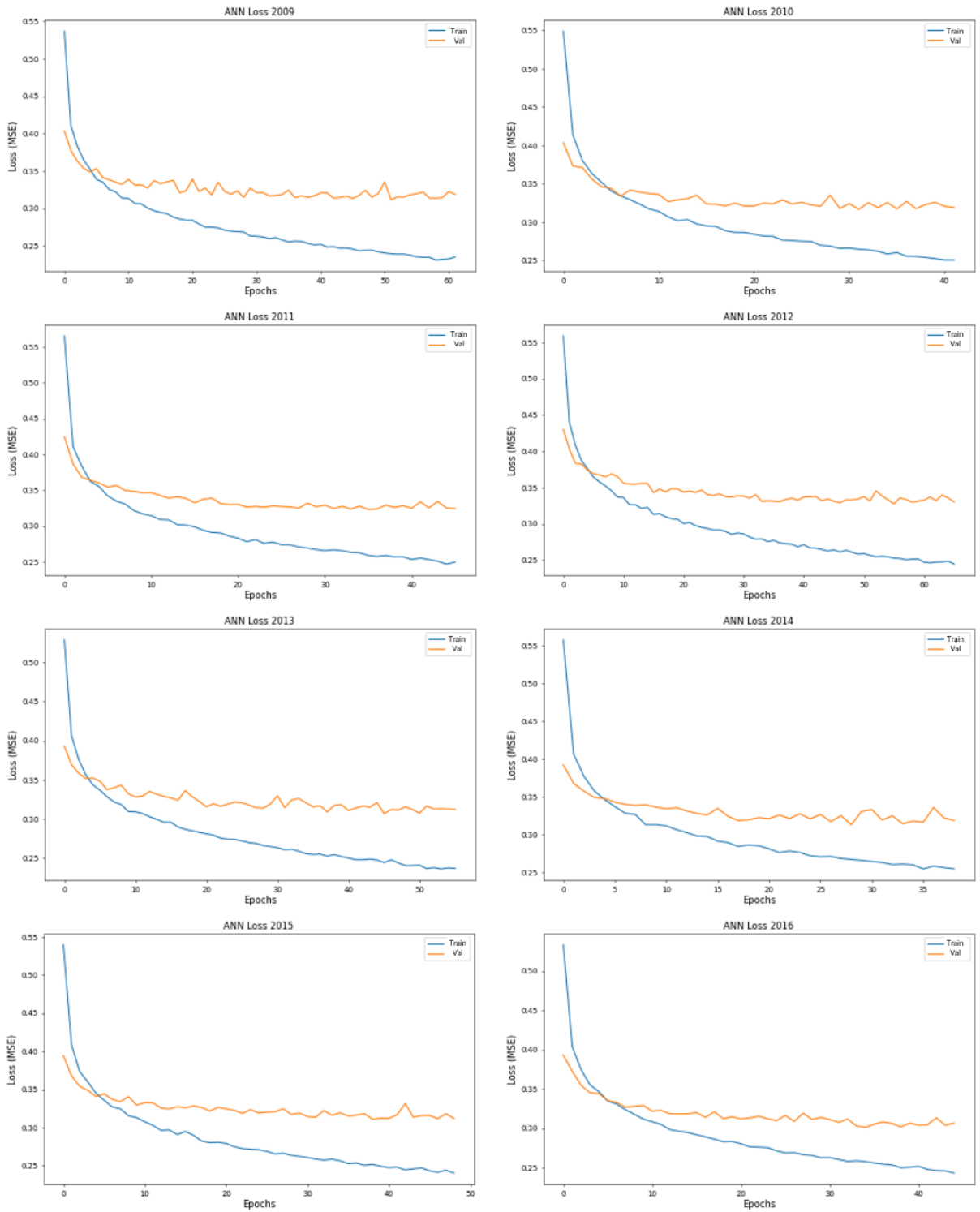
Figure 17: Training and validation loss plots of the Global ANN model with learning rate 10^{-4} for 2008 – 2016.
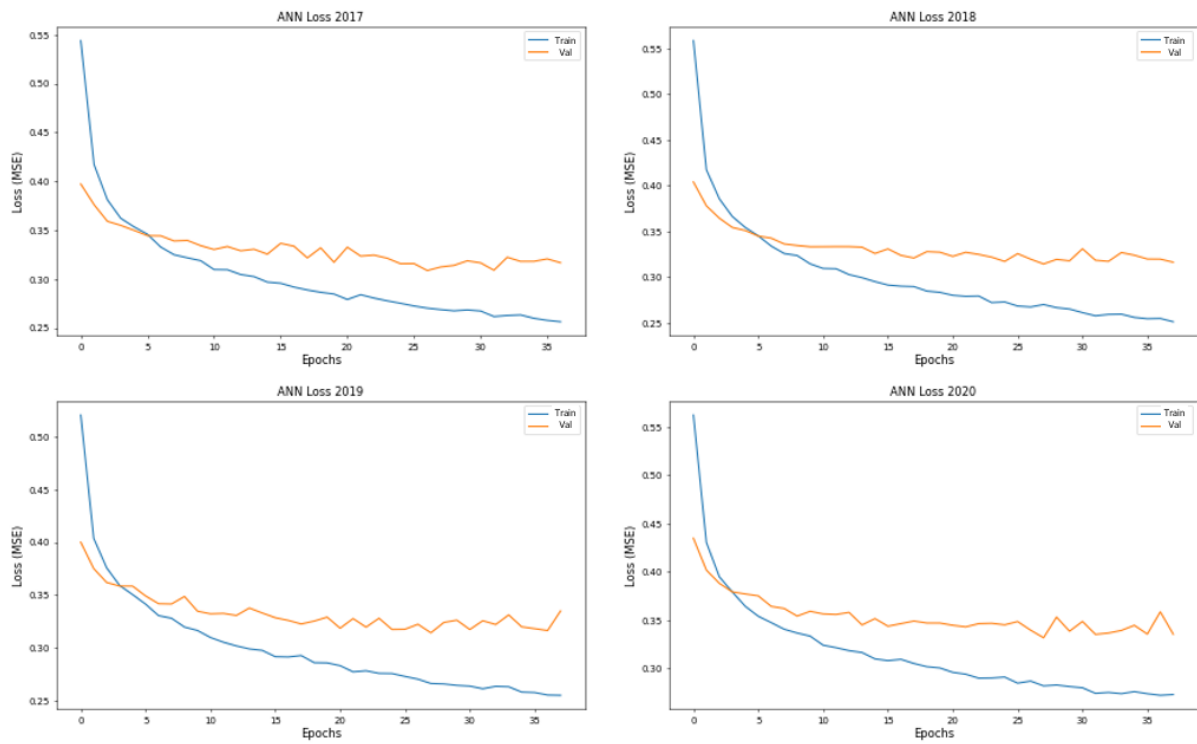
Figure 18: Training and validation loss plots of the Global ANN model with learning rate 10^{-4} for 2017 – 2020.
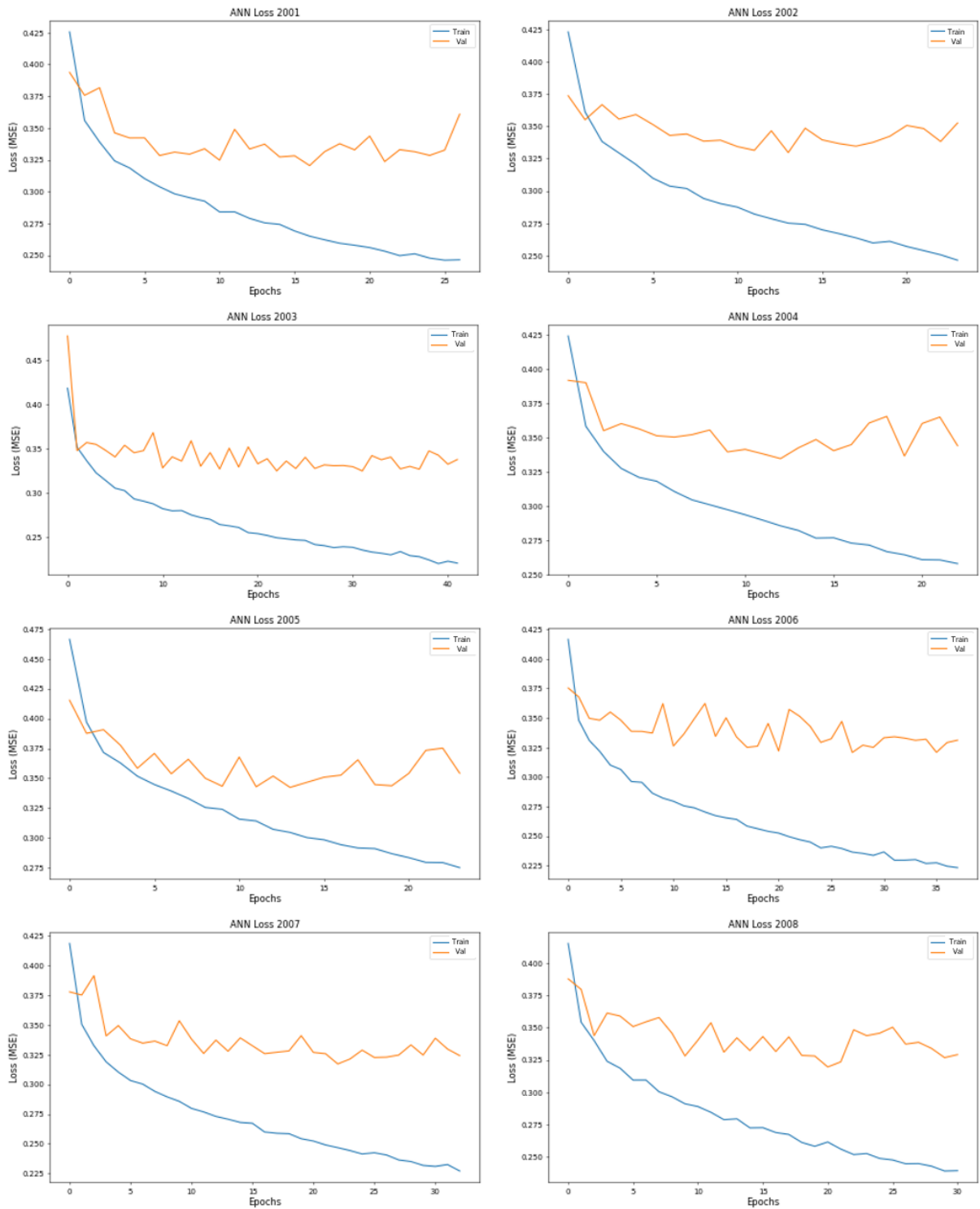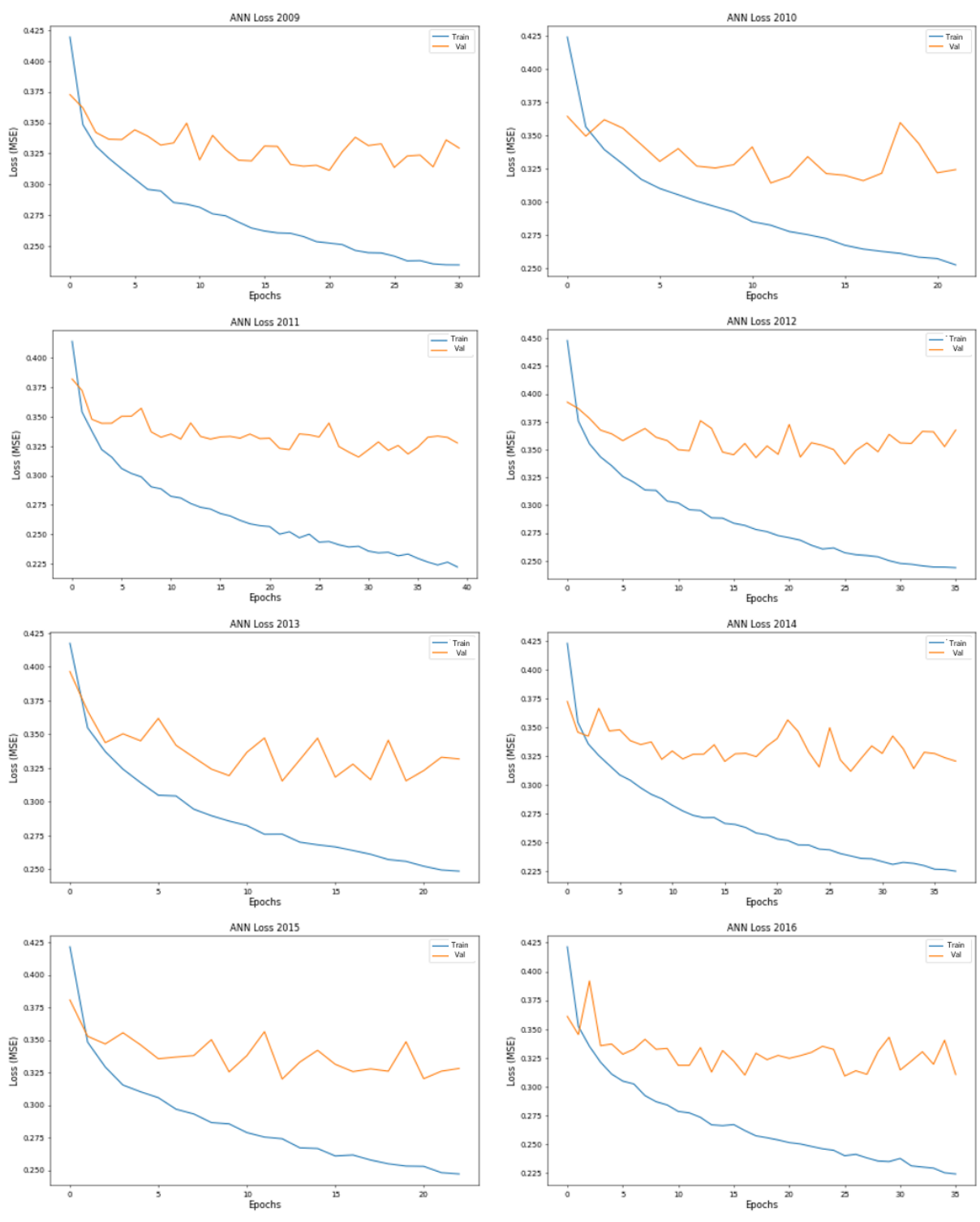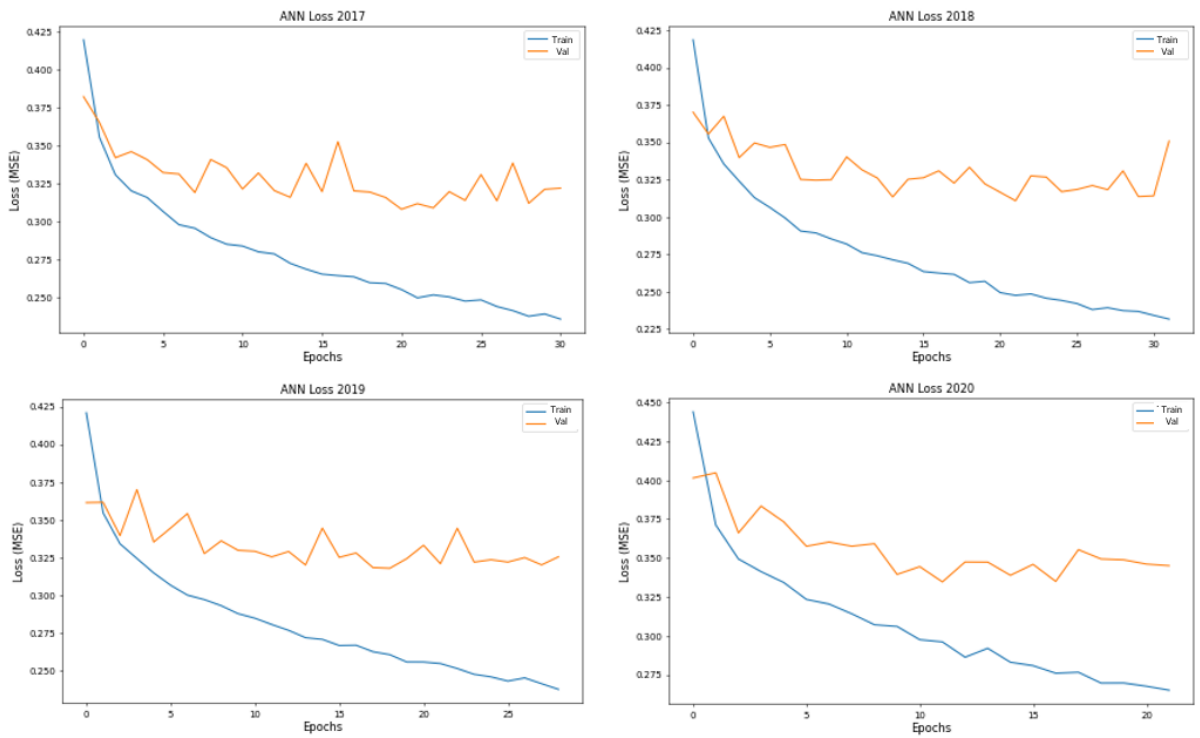
Table S2 – Loss Plots for ANN with learning rate 10^{-3}

Figure 19: Training and validation loss plots of the Global ANN model with learning rate 10^{-3} for 2001 – 2008.

Figure 20: Training and validation loss plots of the Global ANN model with learning rate 10^{-3} for 2009 – 2016.

Figure 21: Training and validation loss plots of the Global ANN model with learning rate 10^{-3} for $2017 - 2020$.
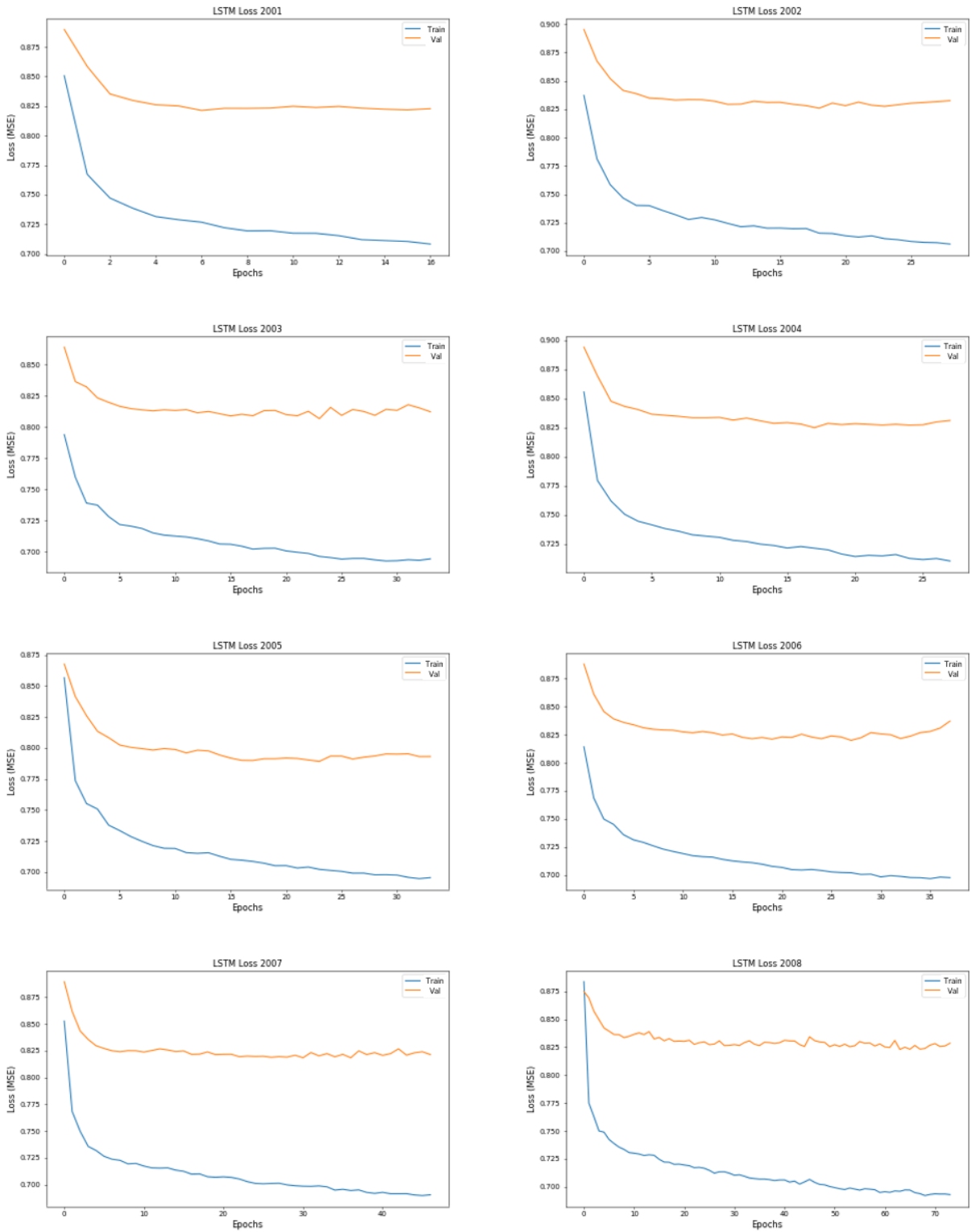
Table S3 – Loss Plots for Bi-LSTM
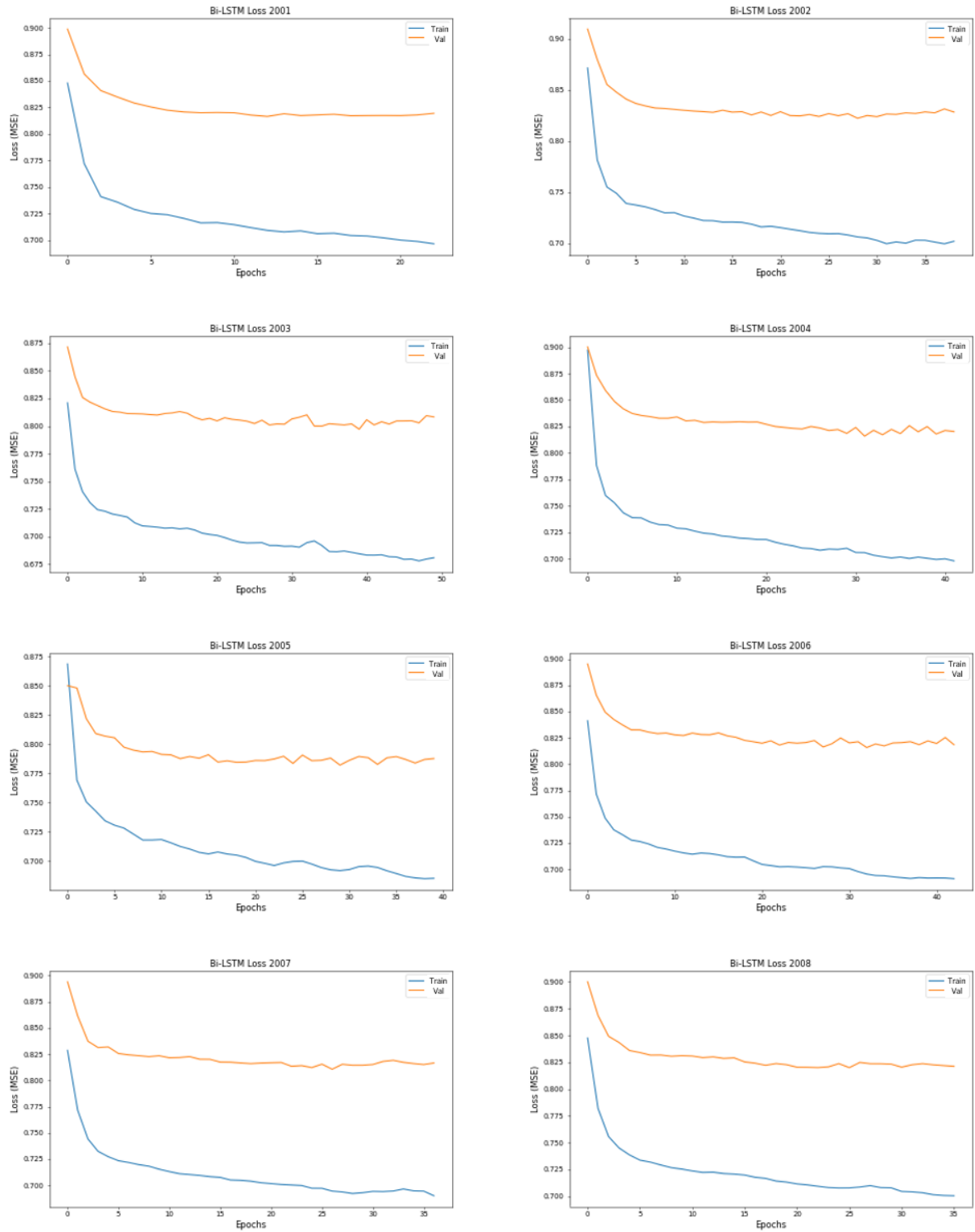
Figure 22: Training and validation loss plots of the Global Bi-LSTM model for 2001 – 2008.
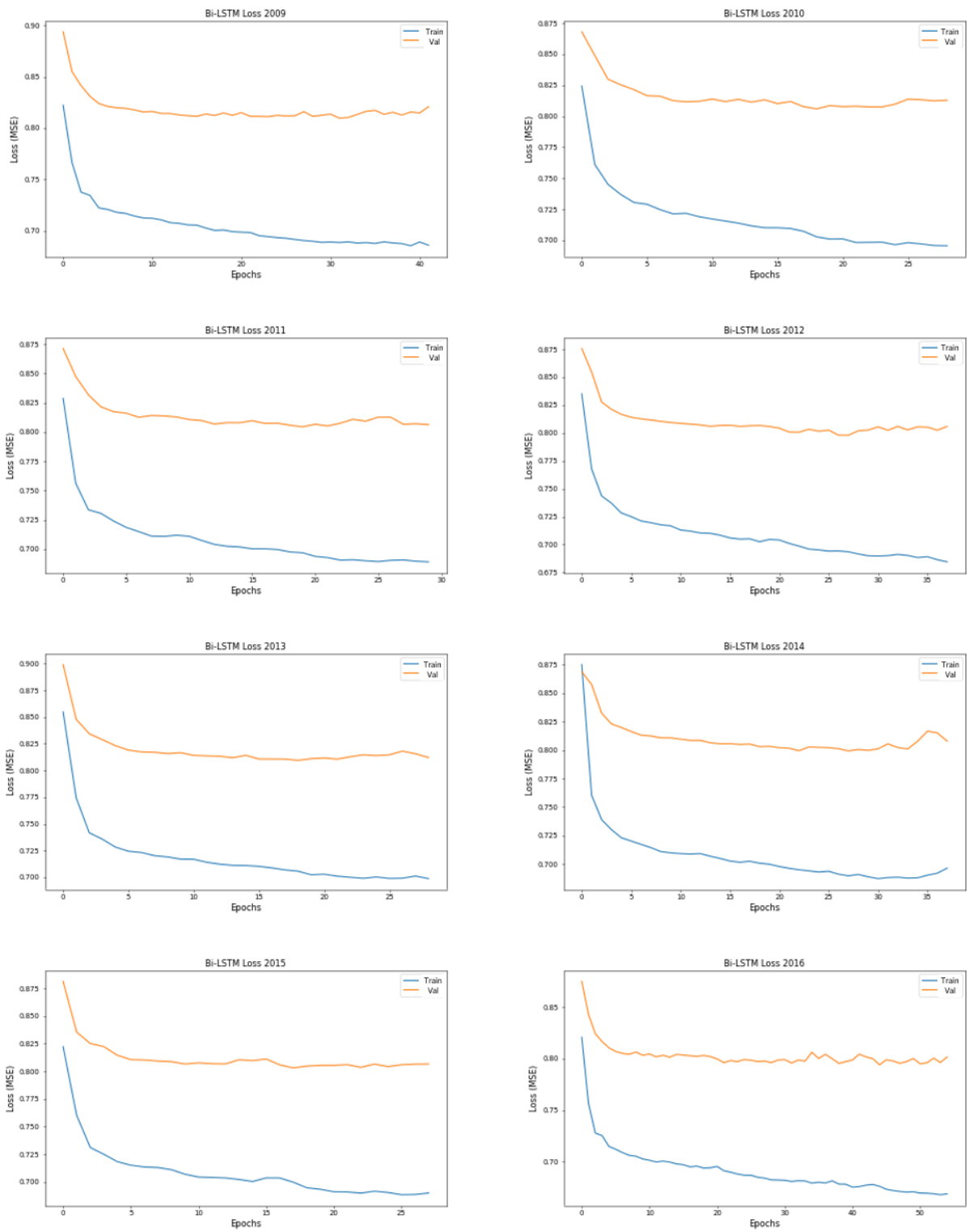
Figure 23: Training and validation loss plots of the Global Bi-LSTM model for 2009 – 2016.
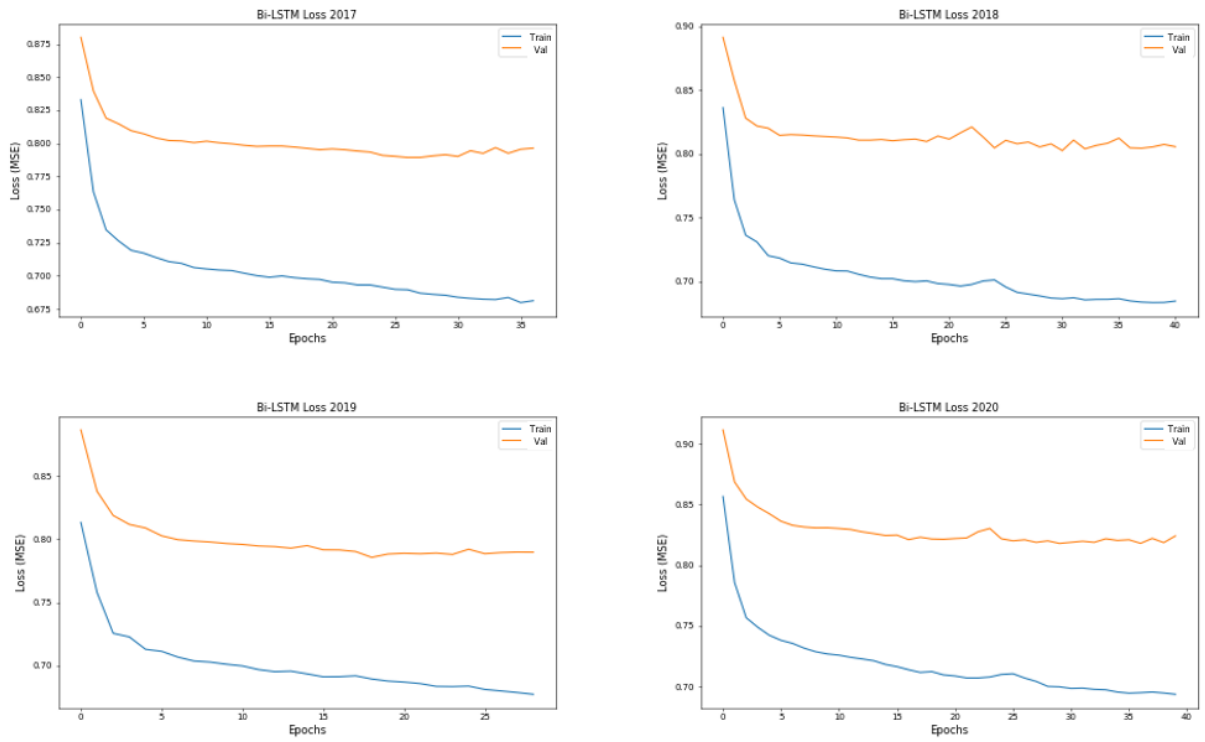
Figure 24: Training and validation loss plots of the Global Bi-LSTM model for 2017 – 2020.

Table S4 – Loss Plots for LSTM

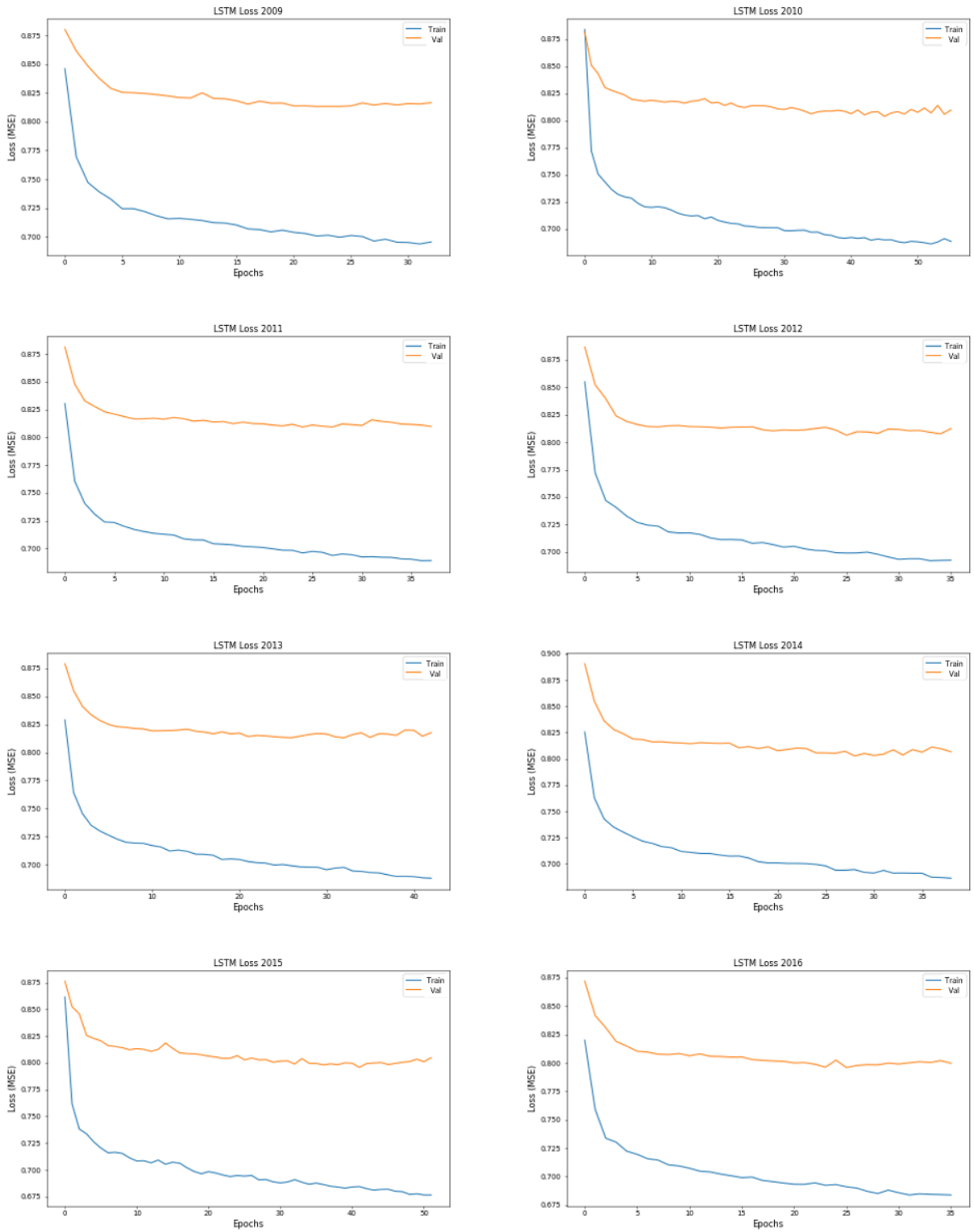Figure 25: Training and validation loss plots of the Global LSTM model for 2001 – 2008.

Figure 26: Training and validation loss plots of the Global LSTM model for 2009 – 2016.
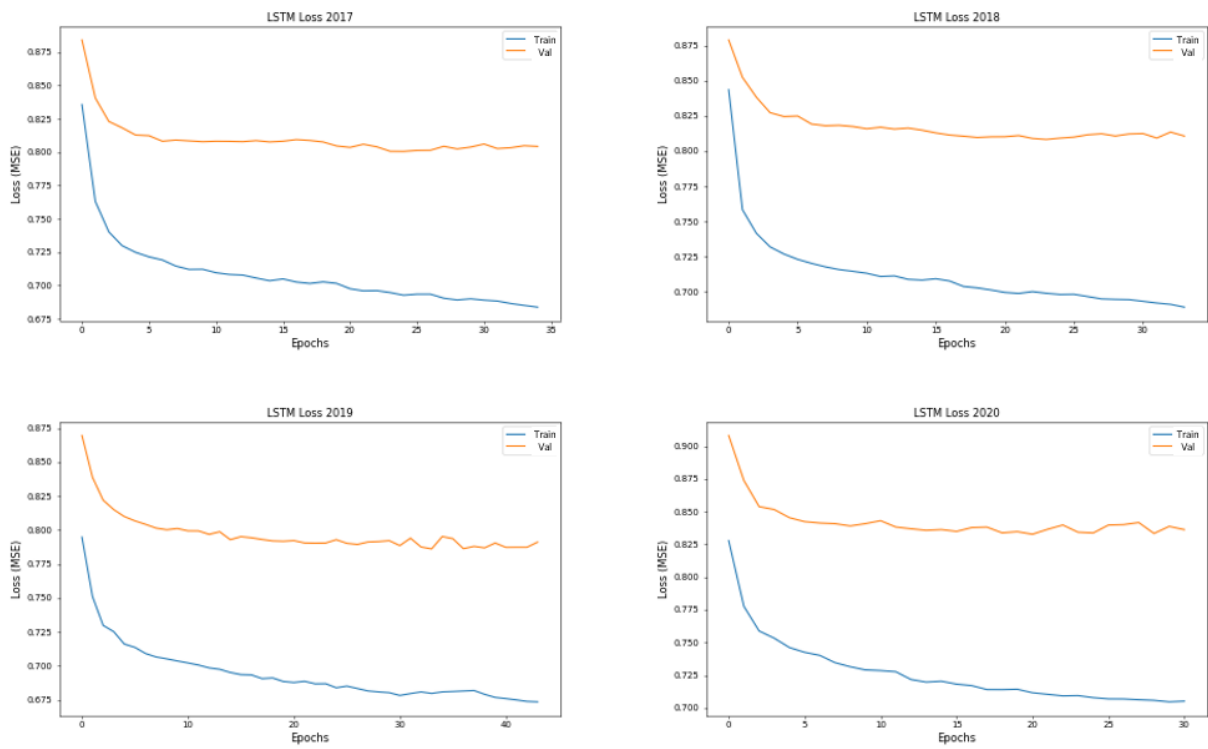
Figure 27: Training and validation loss plots of the Global LSTM model for 2017 – 2020.
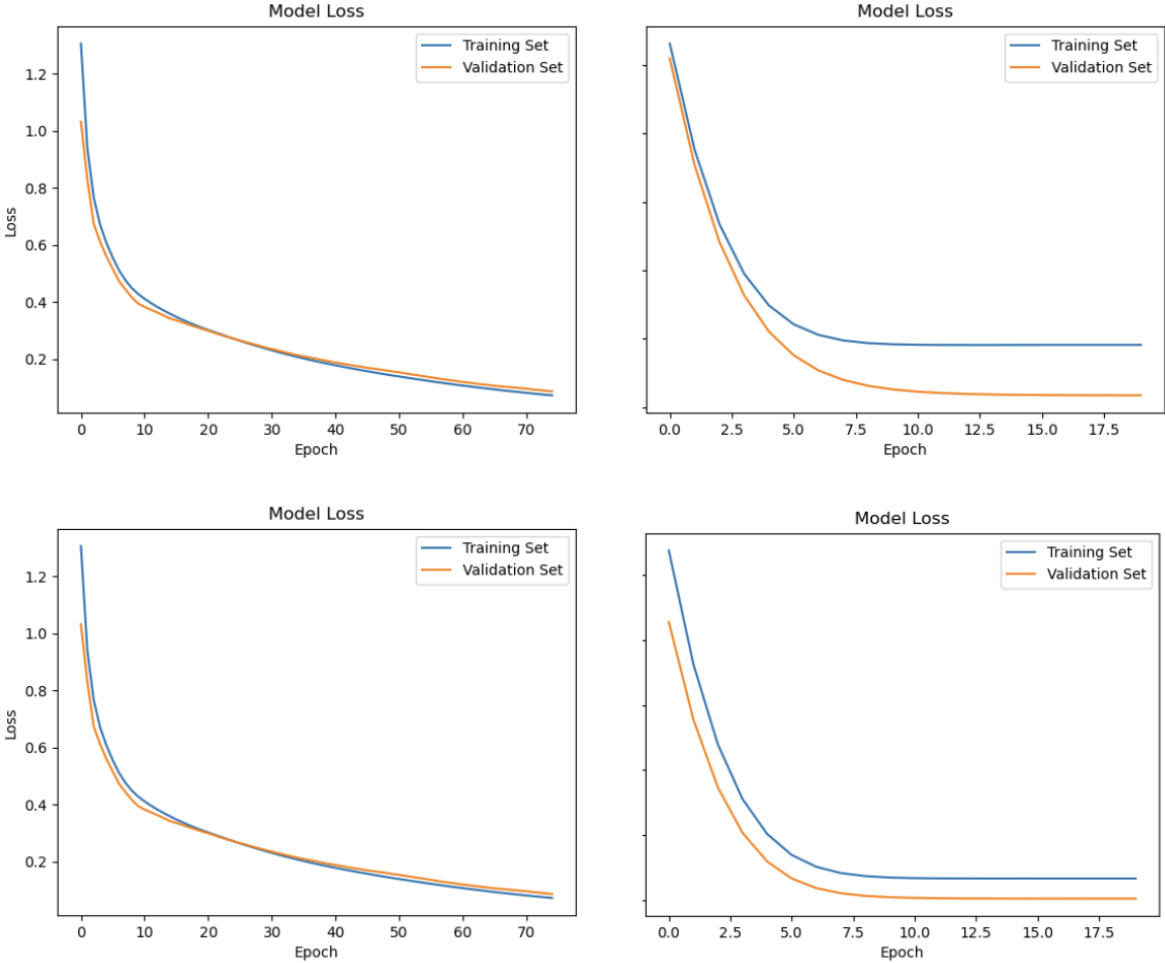
Table S5 – Loss Plots for Transformer based model

Figure 28: Training and validation loss (MSE) plots of the Transformer based model for $2017 - 2020$ using a learning rate 10^{-4} .