REPORT

ON

TIME SERIES ANALYTICS – CROP YIELD PREDICTION

BY

Lakshya Agarwal 2017B5A70904P

Prepared in partial fulfillment of the course

CS F266 – Study Oriented Project

TO

Prof. Poonam Goyal

Associate Professor

Computer Science and Information Systems

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Aug-Dec, 2020

Table of Contents:

Acknowledgement

I would like to thank BITS Pilani and AUGSD for this opportunity to pursue this Study Oriented Project, Time Series Analytics – Crop Yield Prediction.

I would like to extend my sincerest thanks to Prof. Poonam Goyal, Associate Professor, Computer Science and Information Systems, BITS Pilani for the time and effort she has put in guiding me along every step of the way.

Further, I want to express my gratitude towards Research Scholar Ms. Arshveer Kaur for her constant support and unwavering faith in all my endeavors.

Furthermore, I would like to acknowledge with much appreciation the crucial role of my peer Aditya Vishwakarma who has helped me immensely during the course of the project.

Without the help and contribution of the above-mentioned people, this project would not have been a success.

List of Illustrations

1. Introduction

Researchers often use either weather/climate data, satellite remote sensing products, and soil property data or their combination for crop yield prediction at regional scales. The weather data and satellite data are time-series data while the soil property data is commonly considered as constant data that will not change much in the foreseeable future. Because of its advantages such as large scale, continuous, multi-spectral, low cost and long-term archive, remote sensing has become one of the most important tools for yield estimation.

Remote sensing has been used for agricultural applications for over three decades. Several long-running satellites have been widely used in the task. Remotely sensed vegetation indices (VIs) such as the Normalized Difference Vegetation Index (NDVI) have been widely utilized for agricultural mapping and monitoring. Data from the National Oceanic and Atmospheric Administration's (NOAA) Advanced Very High-Resolution Radiometer (AVHRR) have been used since the early 1980's for large scale crop monitoring and yield forecasting. In recent years, the focus of remote sensing-based yield forecasting research has shifted to the National Aeronautics and Space Administration's (NASA) Moderate Resolution Imaging Spectroradiometer (MODIS), which provides improved radiometric calibration and superior spectral and spatial resolution relative to the AVHRR. Both AVHRR and MODIS provide high frequency observations. However, their spatial resolution is quite coarse. MODIS data are available at 250-m, 500-m, and 1000-m spatial resolution depending on the specific product. Higher spatial resolution data from sensors such as the Landsat Thematic Mapper (30 m) have also been used in agricultural applications. However, the repeat period for Landsat is relatively infrequent (16 days), which presents challenges for agricultural applications that rely on high frequency sampling during critical phases of the crop growth cycle. And sentinel dataset is recent started from 2017 and does not have sufficient dataset to train the model.

Recent studies have fused higher spatial resolution map products derived from Landsat with higher frequency MODIS or AVHRR data. However, medium resolution map products derived from Landsat data such as the Cropland Data Layer (CDL), which is produced by the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS), are generally only produced at regional scales and are therefore not available for much of the world. It is therefore important to explore and assess the utility of methods that rely on widely available datasets such as MODIS.

Monitoring growth and yields of crops are exclusively important technologies for food security issues. The methodologies of estimating crop production with remote sensing and crop model have been developed. In fact, crop yields are affected by all the environmental factors (climate, soil etc.) and genetic characteristics including types of cultivar.

To establish more practical model of crop yield estimation, extraction of features which affect crop growth is very important. Generally, crop production per unit area (yield) is on the upward trend due to the improvement of cultivation and its technology. For the time series estimation of yields, influence of such improvements should be considered.

Regardless of the kinds of satellite data being employed, most of the previous researchers prefer to build the relationship between some crop-related indexes and crop yield. The most used indexes extracted from raw satellite data include normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), green chlorophyll vegetation index (GCVI), sun-induced chlorophyll fluorescence (SIF), and the fraction of absorbed photosynthetically active radiation (FPAR), etc. These indexes are essentially multispectral reduction products, which can raise the efficiency but may cause information loss. Besides, weather data(e.g., Land surface temperature (LST), precipitation, and vapor) and soil properties (e.g., clay content, organic carbon content, and PH) have been also included as input for yield prediction. It appears that comprehensive factors can simulate the real growth condition of crops more accurately to improve the performance of yield prediction.

This study used a combination of satellite data, weather data, and soil property data to predict corn yield in the U.S. Corn Belt from 2013 to 2016 at the county-level. A new composite DL model was proposed to take advantage of both CNN and RNN. The multi-level deep learning network (MLDL-Net) framework has two modules. The first module has two steps. The first step is using CNN to extract spatial features from MODIS imagery data and weather data. The second step is using LSTM to extract temporal features from time-series data; The second module is using CNN to extract features from soil property data. To automatically select effective features rather than hand-picking, a histogram-based method was involved to transform natural remote sensing data into fixed-bins histograms and use them to determine which features will be used as inputs. The output from the two modules will be put together as inputs into the final yield prediction step. Validation experiments have been done on the data of the U.S. Corn Belt. The main aims of the work are: (1) evaluate the performance of the proposed method for corn yield prediction in the U.S. Corn Belt; (2) evaluate the influence of different data in the prediction task.

2. Data and methods

2.1 Study region

In this study, the counties in the U.S. Corn Belt were selected as the study area. There are 13 states in the U.S. Corn Belt including North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Wisconsin, Illinois, Michigan, Indiana, Ohio, Kentucky, and Missouri, about 1175 counties in total. Figure 1 shows the selected counties in the GEE. Figure 1 shows the selected counties in the GEE.

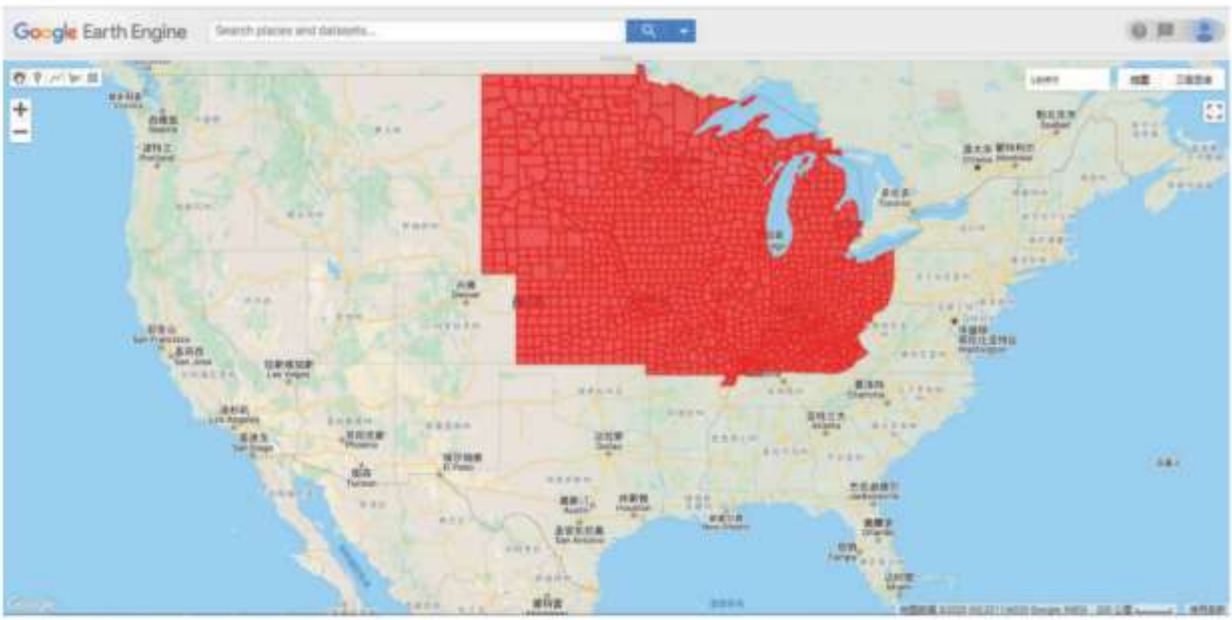


Fig. 1: Study area in the GEE editor(red areas show selected counties in the Corn Belt)

2.2 Data

Considering the related impact factors, i.e., the data availability and generalizability of models, this study uses MODIS surface reflectance (SR) data, MODIS LST data, and Daymet weather data as the time-series data. The openlandmap data was selected as constant soil property data. These data can also fall into two categories, the one is phenological information including MODIS SR data; the other is environmental variables which consist of weather data, MODIS LST, and soil property. A long-term observation can promise mass data for DL, all the data and the corresponding yield data were collected from 2007 to 2016. Moreover, the study time window was set from April 1st to November 30th according to the Usual Planting and Harvesting Dates (UPHD) of U.S. corn [32]. All the data can be freely accessed in the GEE. The details are shown below:

2.2.1 USDA Yield Data

The county-level corn yield data from 2007 to 2016 were collected from the United States Department of Agriculture (USDA) Quick Stats. The data was based on two large panel surveys conducted by the National Agricultural Statistics Service (NASS) and established by the NASS Agricultural Statistics Board (ASB). Figure 2 shows the distribution of the corn yield at the county-level from 2007 to 2016, 11339 records in total. The yield ranges from 1277.77 to 16590.79 kg/ha with a mean of 10016.63 kg/ha. The yield data were used as label data in the DL.

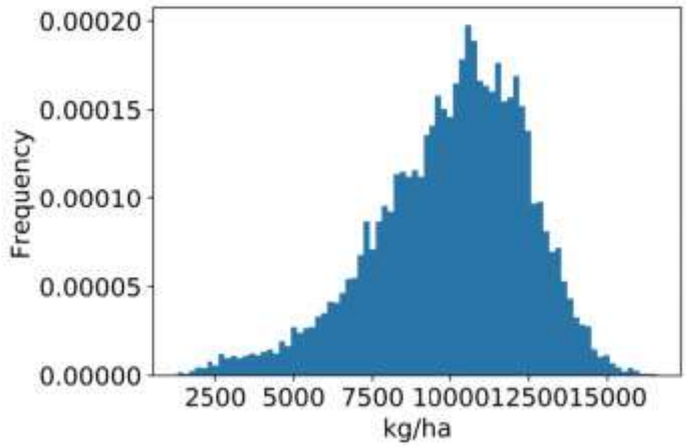


Fig. 2: The distribution of yields from 2007 to 2016

2.2.2 USDA NASS Cropland Data Layers

The Cropland Data Layer (CDL), hosted on CropScape, provides an annual raster, geo-referenced, crop-specific land cover map for the continental United States at 30m spatial resolution. With the help of CDL, we can only focus on the corn data.

2.2.3 MODIS Surface Reflectance

Instead of directly using the handcraft VI, MODIS SR product (MOD09A1) which provides an estimate of the surface spectral reflectance was selected to reflect the state of crop growth. The MOD09A1, with bands 1 to 7, is a 500m gridded, an 8-day composite product derived from the MODIS-Terra top of atmosphere reflectance swaths.

2.2.4 MODIS Land Surface Temperature

The MODIS Land Surface Temperature (LST) products are composed of data from the daily 1 km LST product (MOD11A1) stored on a 1 km grid as the average values of clear sky LSTs during 8 days. MOD11A2 is comprised of daytime and nighttime LSTs, quality assurance assessment, observation times, view angles, bits of clear sky days and nights.

2.2.5 Weather Data

The Daymet dataset provides gridded estimates of daily weather parameters. Seven surface weather parameters are available at a daily time step, 1 km spatial resolution, with a North American spatial extent. It is derived from selected meteorological station data and various supporting data sources. In the study, two important weather parameters, precipitation, and vapor pressure were selected as climatic factors.

2.2.6 Soil Property Data

The GEE provides global predictions for standard numeric soil physical and chemical properties at six standard depths (0, 10, 30, 60, 100, and 200 cm) in GEE. The predictions were based on ca. 150,000 soil profiles used for training and a stack of 158 remote sensing-based soil covariates (primarily derived from MODIS land products, SRTM DEM derivatives, climatic images, and global landform and lithology maps). As shown in Table I, six important properties were selected as soil property data.

TABLE I: Soil properties selected in the model.

Feature	Unit
Clay content mass fraction(CLAY)	%
Sand content mass fraction(SAND)	%
Water content at 33kPa(WATER)	%
pH in H2O(PH)	-
Bulk density(BULK)	kg/m^3
Carbon content(CARBON)	g/kg

2.3 GEE-based Tensor Generation

Tensor generation is the key step for DL. Considering the efficiency of the DL, a dimension reduction was always conducted at first, for instance, most of the previous studies often use the averages of regions to simplify the raw data. However, DL prefers to learn the features from the raw data. Given the assumption that the position information of pixels is unimportant for yield prediction. Then, there is little loss of information in transforming the high-dimensional image into a histogram of pixel count. It is easy to reconstruct the histogram into a tensor for the DL. The histogram-based transformation can not only keep features as much as possible from the raw data but also decrease the dimension sharply. Thanks for the great computing power of GEE, the processing can be performed very efficiently over the U.S. Corn Belt. The key steps are as follows:

2.3.1 Preprocessing

All the data should be collected, and cloud masked in the GEE. As for the time-series data, the daily weather data should be aligned to the 8-day MODIS data by the average values. The time-series data has 30 time-steps during the whole growing season at an 8-day interval. Then, after combining the boundary data, each county will have a 11 bands image composite $M^t_{id;y}$ for time-series data at the t^{th} ($0 < t <= 30$) time step and a 36 bands S_{id} for constant soil property data in each year. The id represents the GEOID of a county and the y denotes a year in the list ranging from the start to the predicted year y0. The 11 bands of $M^t_{id;y}$ consists of 7 bands of the MODIS SR, 2 bands of the MODIS LST, and 2 bands of the weather data. The 36 bands of the S_{id} include 6 types of properties at 6 depths. The CDL data was employed to mask all non-corn pixels for all data.

2.3.2 Histogram-based Tensor Generation

As the inputs of the DL, tensors always need regular shape. Our tensors come from the fixed-bins histogram-based transformation performed on the composites. Each band of composites over a county can be transformed into a histogram with n bins (the n was set to 32 for regulation in the study). Then, the $M^t_{id;y}$ can be transformed into the tensor $TM^t_{id;y} \in R^{1x32x11}$, and the Sid will become the tensor $TS_{id} \in R^{1x32x36}$. Finally, each county will have a yield label from USDA statistics, if any. For example, the tensors $TM^{18}_{20115;2016}$ and TS_{20115} of Kansas Marion county GEOID:20115) in 2016 are visualized in Figure A1 and A2 of Appendix A. The x-axis represents the bin number (32) and the y-axis shows the normalized pixel count. The pixels of each band were distributed into 32 uniform bins according to the DN value.

2.4 DL Architecture

Figure 3 shows the workflow of the prediction. The proposed DL architecture mainly consists of two modules. The first includes two levels for spatial and temporal feature extraction from time-series tensors. The other module is for soil property feature extraction from constant soil tensors. The output is the predicted yield. The detailed descriptions of each level are as follows:

2.4.1 Level I CNN for Time-series Data

The time-series data includes crop phenology and climate information. To explore more spatial features, a CNN method was proposed. As shown in Figure 3, at each time step, the $TM^{t}_{id,y}$ was fed into the CNNs with K (k=2) 2-dimension convolution (Conv2D) layers. The output $TM^{t,K}_{id,y}$ from the Kth convolution layer can be defined in Equation 1:

$$TM^{t,K}_{id,y} = f(TM^{t,K-1}_{id,y} * W^{K} + B^{K}) \qquad (1)$$

where $*$ represents the convolutional operation and $f(\cdot)$ is an activation function. The $TM^{t,K-1}_{id,y}$ is the output from the $(K-1)^{th}$ layer. W^K and B^K are parameters of the k^{th} convolution layer. Each convolution layer is followed by a batch normalization layer, which is used to normalize the input layer by adjusting and scaling the activations for improving the speed, performance, and stability of the networks. After the batch normalization layer, the output is flattened into a long vector, which is sent into a fully connected (FC) layer. The FC offers learns features from all the combinations of the features of the previous layer. Finally, we can obtain a processed feature $VM^{t}_{id,y}$.

2.4.2 Level II LSTM for Time-series Data

The relations of the time-series crop phenologies and climate changes are of significance for yield estimation. The LSTM, a special RNN which can solve the vanishing gradient problem, was employed for time-series feature analysis, as it is capable of learning long-term dependencies. The employed LSTM consists of the cell $c_t \in R$ h (the memory part of the LSTM unit) which contains h LSTM unit's cells at t interval, the input gate $i_t \in R$ h , the output gate $o_t \in R$ h , and the forget gate ft $\in R$ h . Intuitively, the cell can keep track of the dependencies of the input sequence. The input gate controls the extent of the new inputs, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit. The time-series output of level 1 is the input of the LSTM, the whole processing can be defined in Equation 2.

$$
\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \sigma(W_c x_t + U_c h_{t-1} + b_c) \\
h_t &= o_t \circ \sigma(c_t)
\end{aligned}
\qquad (2)
$$

where the operator ◦ denotes the Hadamard product, the σ is the activation function, the xt ∈ R d is the input vector to the LSTM unit, the ht ∈ R h is the hidden state vector, the W ∈ $R^{h\times d}$, U ∈ $R^{h\times h}$ and b ∈ R^{h} are weight matrices and bias vector parameters that need to be learned during training. The superscripts d and h refer to the number of the input features and the hidden units, respectively.

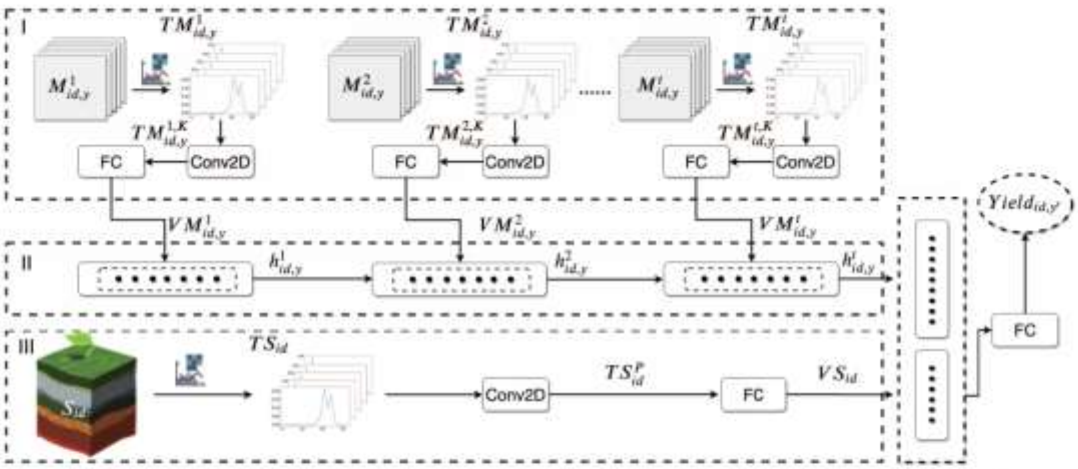


Fig. 3: The architecture of the MLDL-Net. The inputs are the time-series $TM^t_{id,y} \in \mathbb{R}^{1\times 32\times 11}$ tensors and constant $TS_{id} \in \mathbb{R}^{1\times 32\times 36}$ tensors. The output is the predicted yield $yield_{id,y'}$. Level I employs CNN networks for time-series data spatial features exploration at each time step, the CNN network consists of K Conv2D which followed by batch normalization layers and fully connected layers; Level II uses an LSTM network to explore the time-series features; Level III also uses CNN networks which can extract spatial features from soil property data. The output of Level II and Level III was flattened and concatenated into a vector. Finally, the vector was fed into the fully connected layer used for estimation.

2.4.3 Level III CNN for Soil Properties Data

Compared with the time-series data, the soil data of each pixel is constant. To capture the soil property features from the tensor TS_{id}, a CNN based network was designed, which is similar to level 1. The network consists of P Conv2D layers and each layer is followed by a batch normalization layer. Therefore, the output of the P^{th} can be defined in Equation 3. At last, all the features was flattened and fed into the FC layer.

$$TS^P_{id} = f(TS^{P-1}_{id} * W^P + B^P) \qquad (3)$$

where the TS^{P-1}_{id} is the output from the $(P-1)^{th}$ layer. W^P and B^P are parameters of the P^{th} convolution layer.

At the end of the workflow in Figure 3, the output of level 2 and level 3 was concatenated together and fed into a dropout layer; the dropout layer can help prevent the model from overfitting. Finally, an FC with only one unit is used to output the predicted yield $yield_{id,y'}$ of the predicted year y' .

2.5 Model Training and Evaluation

2.5.1 Model Training

More training data can provide a more stable result, to predict the $yield_{id,y'}$ of the target year y' , all the $M^t_{id,y}$ and Sid was collected as the training data, where y range from 2007 to y' , and the time steps t ranges from 0 to 30. Specifically, when t = 30, the model is identified for an after-season prediction, otherwise, it is for in-season prediction. In the study, 14 typical time nodes ranging from May 9th to Nov 25th were selected to explore the potential of the MLDL for in-season and after-season yield prediction. These time nodes mainly concentrate in Jun, Jul, and Aug, from which we hope to find out an in-season time node for a comparable performance of estimation. To ensure unbiasedness, a fixed ratio (0.3) of the training data was randomly separated for validation. During the training, the number of the epochs was set to 200 with a batch size of 16, however, too many epochs may lead to over-fitting of the training dataset, and too few may result in an under-fit model. The mean absolute error (MAE) was selected for a performance measure to monitor early stopping, which was employed as it can stop training once the model performance stops improving after 10 consecutive epochs.

2.5.2 Model Evaluation

To evaluate the performance of the model, the target year y 0 of yield prediction was set to the value from 2013 to 2016 for an average evaluation. The number of training and test data is shown in Table II. Root mean squared error (RMSE) and mean absolute percentage error (MAPE) were selected for overall evaluation measures; Percent Error (PE) was used for the distribution of the error map. Formulas of RMSE, MAPE, and PE are presented in Equations 4, 5, and 6 where xi is the predicted value, xˆi is the observed value, and n is the number of samples. Besides, the R2 was also used to evaluate how well the predicted values can reconstruct the spatial variations of observed yield.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}} \qquad (4)$$

$$MAPE = \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{\hat{x}_i} \right| \cdot \frac{100\%}{n} \qquad (5)$$

$$PE = \frac{|x_i - \hat{x}_i|}{\hat{x}_i} \cdot 100\% \qquad (6)$$

3. Implementation

3.1 Data Export

The model requires the following datasets:

- MODIS SR product: MOD09A1
- MODIS LST product: MOD11A2
- DAYMET product: DAYMET_V3
- Mask Product: MCD12Q1
- Clay Product: SOL_CLAY-WFRACTION_USDA-3A1A1A_M/v02
- Soil pH product: SOL_PH-H2O_USDA-4C1A2A_M/v02
- Water Product: SOL_WATERCONTENT-33KPA_USDA-4B1C_M/v01
- Soil carbon content product: SOL_ORGANIC-CARBON_USDA-6A1C_M/v02
- Soil bulk density product: SOL_BULKDENS-FINEEARTH_USDA-4A1H_M/v02
- Sand content product: SOL_SAND-WFRACTION_USDA-3A1A1A_M/v02

Using Earth Engine API in python above mentioned products were downloaded in the form of Image Collections for the required period which is 2007 to 2016. TIGER: US Census Counties 2018 data is used to divide images as per the counties required. The United States Census Bureau TIGER dataset contains the 2018 boundaries for primary legal divisions of US states. In most states, these entities are termed "counties". In Louisiana, these divisions are known as "parishes".

The images finally stored in .tif format, for each product a single image represents a county (hereafter referred as image) named as “state_county.tif”. Each image stores the data for the whole-time range and for each day all bands are stored together. Below mentioned example will be able to explain this in a better way. Say we have data for 3 days D1, D2 and D3 and each day have 4 bands say B1, B2, B3 and B4. So, data is stored in the sequentially as D1B1 D1B2 D1B3 D1B4 D2B1 D2B2 D2B3 D2B4 D3B1 D3B2 D3B3 D3B4.

Dimensions of each image will be l X b X (#days*#bands) where l represents length of county and b represents breadth of the county.

3.2 Data Cleaning

First, all required bands (image, temperature, mask and daymet/weather) are loaded for each county in numpy array using gdal python library as we process them. All the pixel values of bands are filtered and scaled from 0-5000 as mentioned in the table below:

Table 1. The actual data range of features.

Feature	Original Min	Original Max	New Min	New Max
MOD09A1	−100	16,000	1	5000
MOD11A2	7500	65,535	12,400	15,600
PRECIPITATION	0	200	0	35
PRESSURE	0	10,000	0	3200

where new min and mew max denotes the filtering range after which specified scaling is done.

Mask band having value 12 denotes Croplands: at least 60% of area is cultivated cropland. So, all pixels having values not equals to 12 are replaced to 0 otherwise to 1.

Now each image for each product is further divided into years and further all the products (image, temperature, weather) for each year are combined (hereafter referred as composite image), followed by masking by multiplying the mask layer to composite image.

The structure of composite image looks like

l X b X (#timeStepsInAYear *(#bandsImage+ #bandsTemperature + #bandsWeather))

where # timeStepsInAYear denotes number of timesteps in a year as per temporal resolution

#bandsImage, #bandsTemperature, #bandsWeather denotes number of bands used from image surface reflectance, LST and daymet dataset respectively.

Composite image obtained after cleaning is saved in format "year_state_county.npy".

3.3 Histogram Generation

Each of the composite image is loaded as they are processed. Each image is filtered for required timespan only which is from day 91 to day 335 in case of corn which is basically time of crop cycle. After this histogram generation is done with 32 bins equally spaced for values from 0 to 5000. For histogram generation each band of the composite image is individually converted into a histogram. Histogram of each day (11bands) is appended in a list & further these are combined year-wise (as there in composite image) and appended in a master list. Simultaneously respective county crop yield for year is also appended in another list. Finally, these are stored in .npz format from which we can easily access during model training and evaluation.

3.4 Model

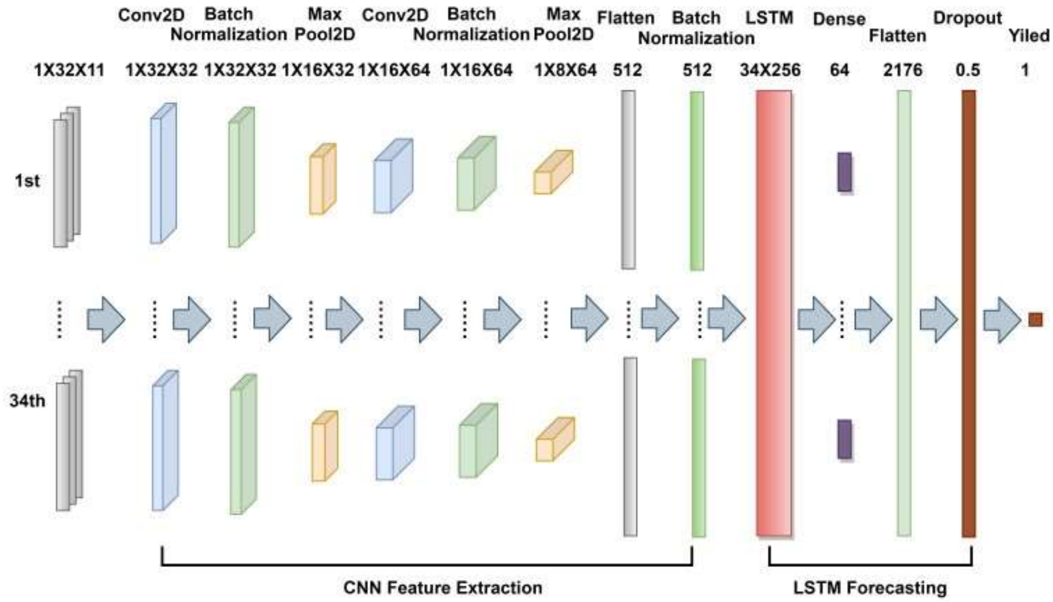


Figure 3. The architecture of the proposed CNN-LSTM model.

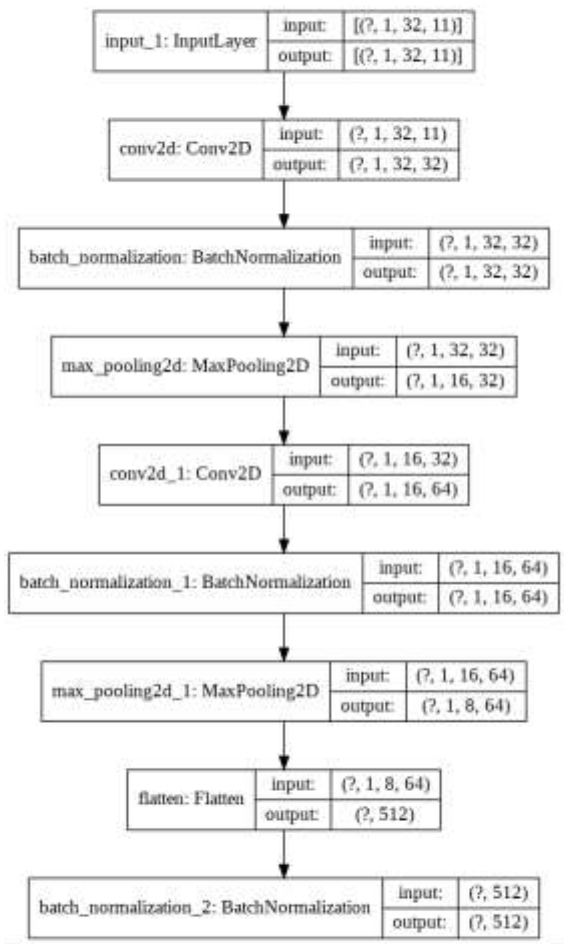


Fig. 5 CNN Model for time-series data

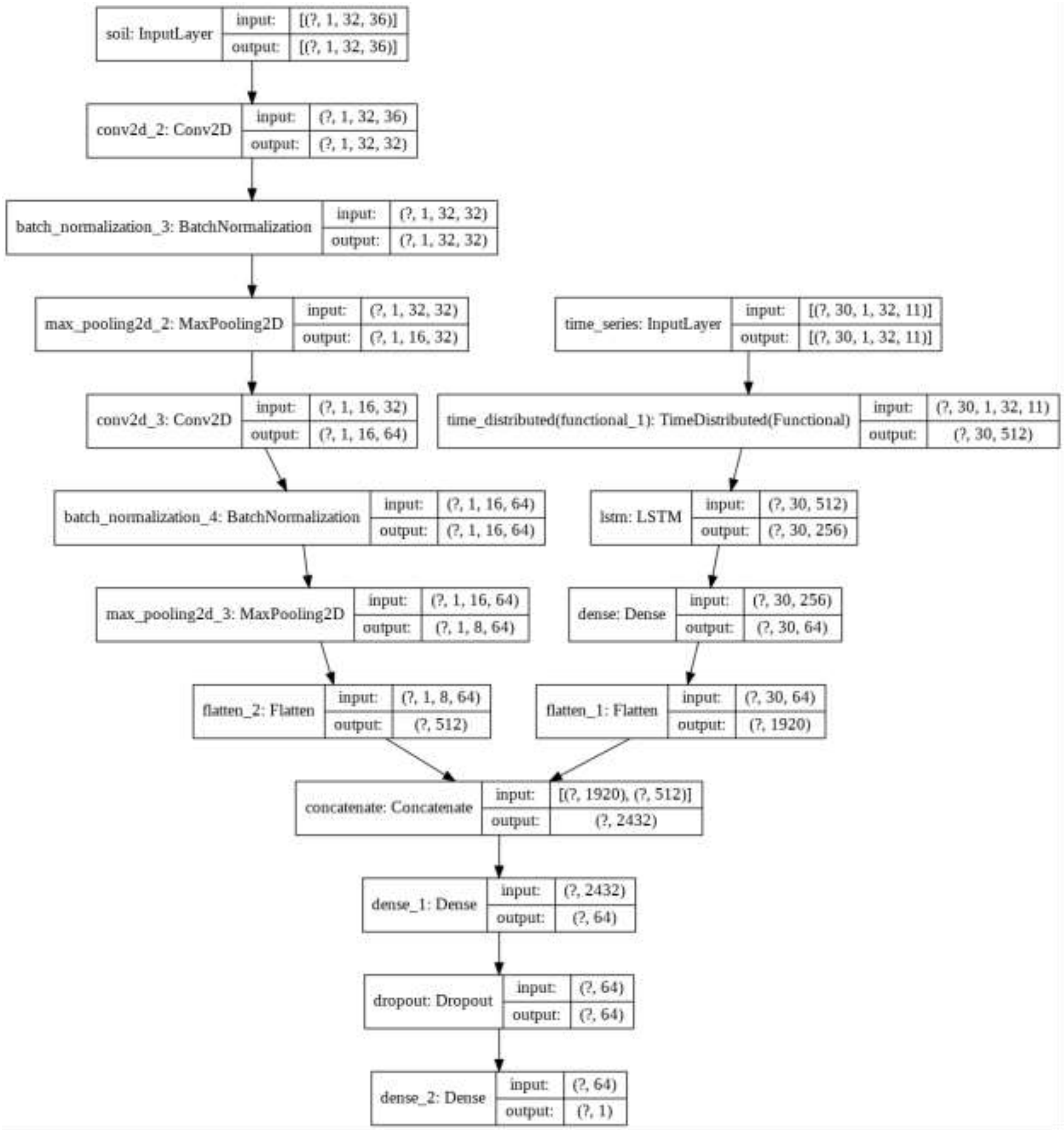


Fig. 6 Complete LSTM-CNN Model including CNN for static soil data

4. Results and Discussion

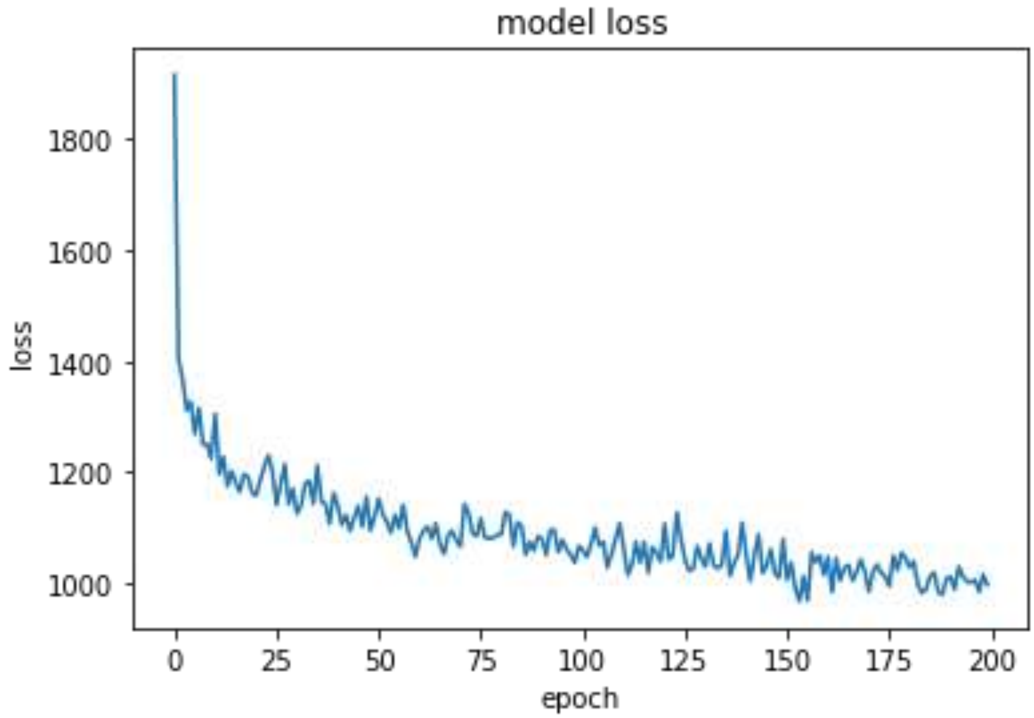
Below mentioned are model name with some other changes (if any, explained with file name) in the following format:

Datasize_epochs_batchSize_learningRate*10^{-3}

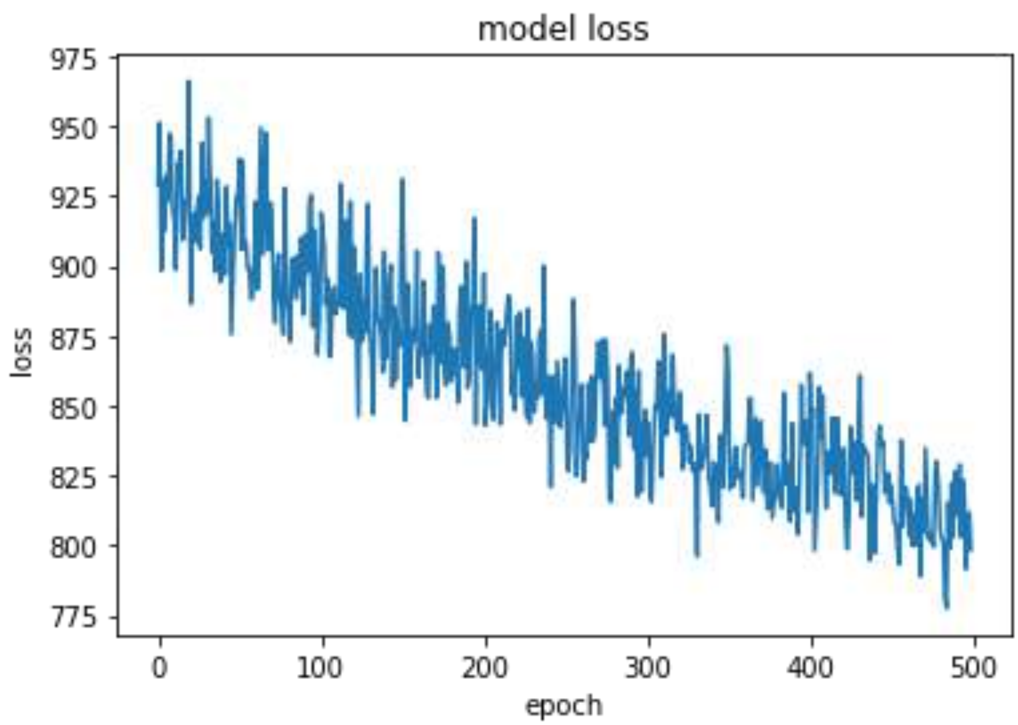
MODEL NAME **MAPE**

1. 4602_200_16_2_cnn_32_64.h5: **19.367494878310467**
   Normal Model
2. 4602_200-16_2_cnn_32_64_noDropout.h5: 19.3051744722527
   Dropout layer removed
3. 4602_200_16_2_cnn_48_96.h5: 19.553417575843206
4. 4602_200-16_2_cnn_32_64_poolSoil_1_3: 20.088272127683805
   Soil pooling changed to (1,3) instead of (1,2)
5. 4602_200-16_2_cnn_32_64_lstm_128_64: 19.35577232332208
   LSTM FC changed to 128 from 256.
6. 4602_200-16_2_lstm_128_64_cnnSoil_32_48: 18.9132849220148
   LSTM FC changed to 128 from 256 and soil CNN layers changed to 32 then 48 instead of 32 then 64.
7. 4672_200-16_2_lstm_128_64_cnnSoil_24_36: **16.91209778642009**
   LSTM FC changed to 128 from 256 and soil CNN layers changed to 24 then 36 instead of 32 then 64.
8. 4672_200-16_2_lstm_128_64_cnnSoil_12_24: 20.519011973045956
   LSTM FC changed to 128 from 256 and soil CNN layers changed to 12 then 24 instead of 32 then 64.
9. 4672_200-16_2_lstm_128_64_cnnSoil_24_36_soilmaxpool_1_3: 18.51287954215208
   LSTM FC changed to 128 from 256, soil CNN layers changed to 24 then 36 instead of 32 then 64, soil max pooling changed to (1,3) instead of (1,2).
10. 4672_200-16_2_lstm_128_64_soilsinglecnn_64: 19.546269380915504
    LSTM FC changed to 128 from 256 and single soil CNN layer of 64.
11. 4672_200-16_2_lstm_128_64_cnnSoil_48_64: 18.871213871488997
    LSTM FC changed to 128 from 256 and soil CNN layers changed to 48 then 64 instead of 32 then 64.

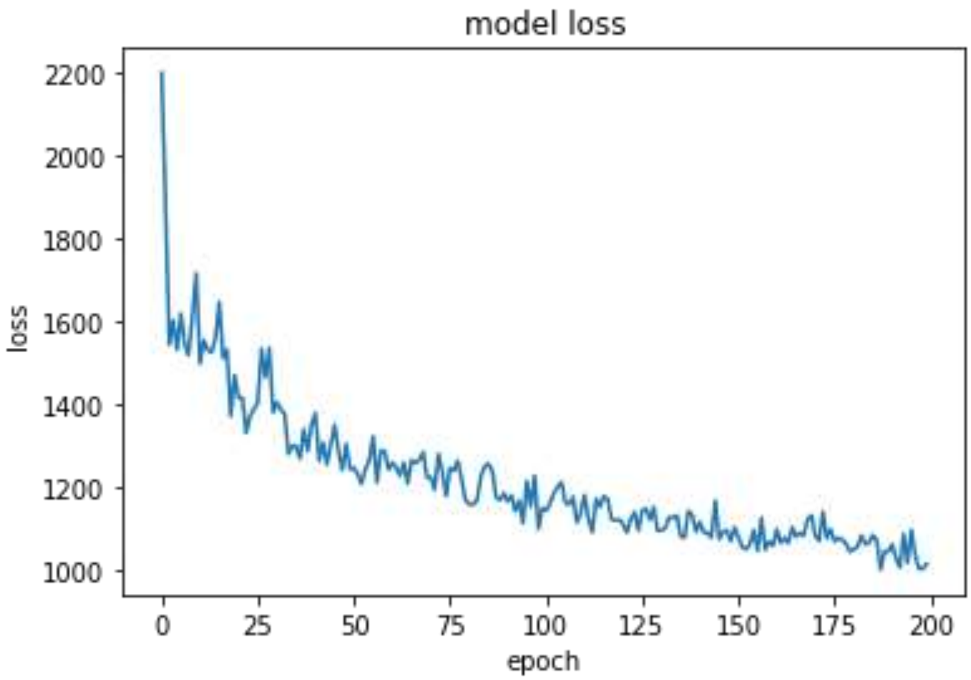
12. 4510_opt_200_16_2_cnn_48_64: 20.734207636552348
CNN time series data changed to 48 and 64 from 32 and 64.



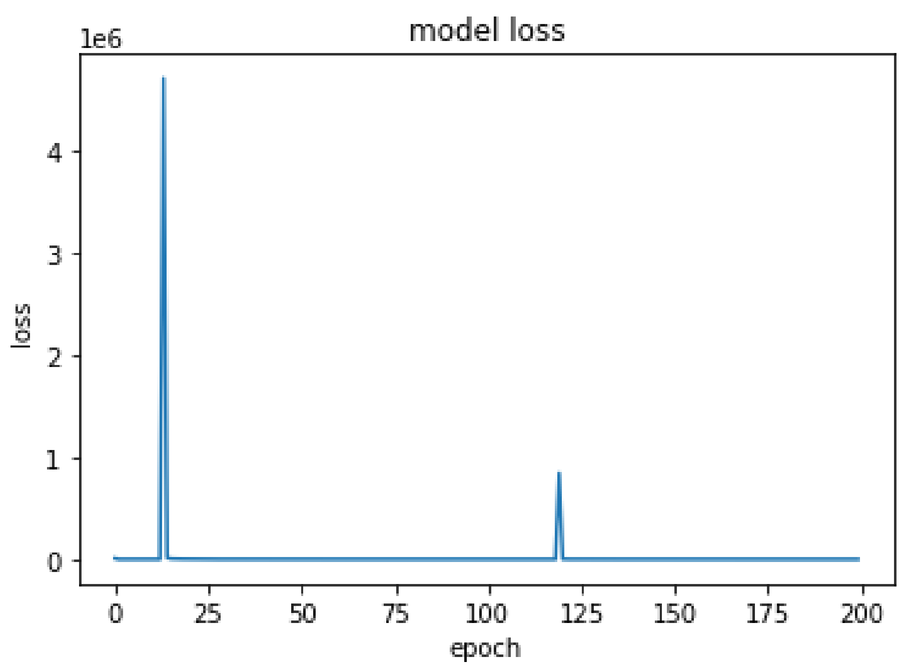
13. 4510_opt_500_16_2_cnnSoil_24_36: 18.779222869374525
Epochs changed to 500 and soil CNN layers changed to 24 then 36 instead of 32 then 64.



14. 4510_opt_200_16_10_cnnsoil_48_64: 18.71670954551692
    soil CNN layers changed to 24 then 36 instead of 32 then 64 and learning rate to 0.01 from 0.002

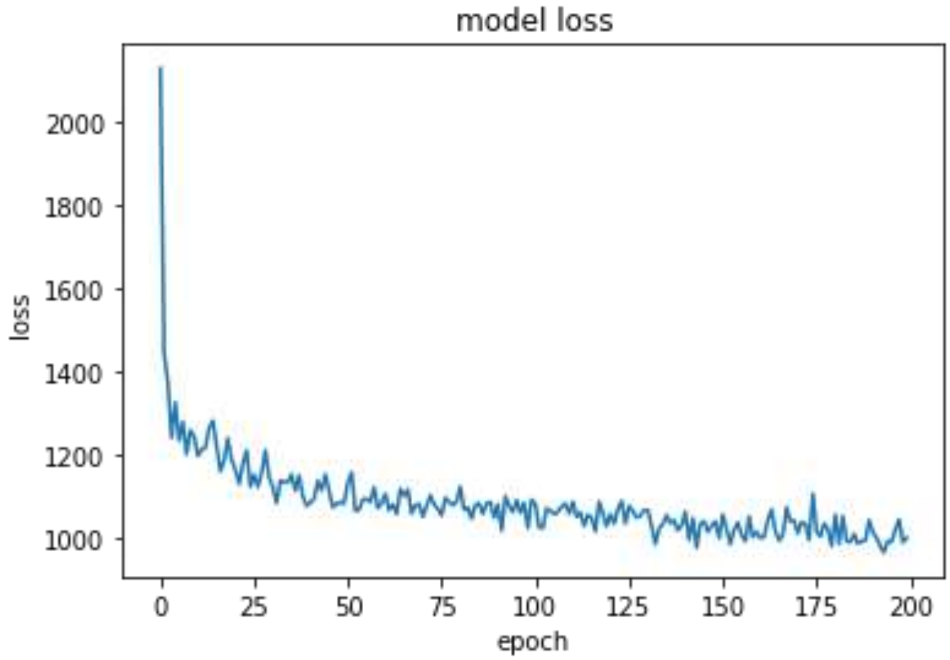


15. 4510_opt_200_16_30_cnnsoil_48_64: **16.675054973269063**
    soil CNN layers changed to 24 then 36 instead of 32 then 64 and learning rate to 0.03 from 0.002

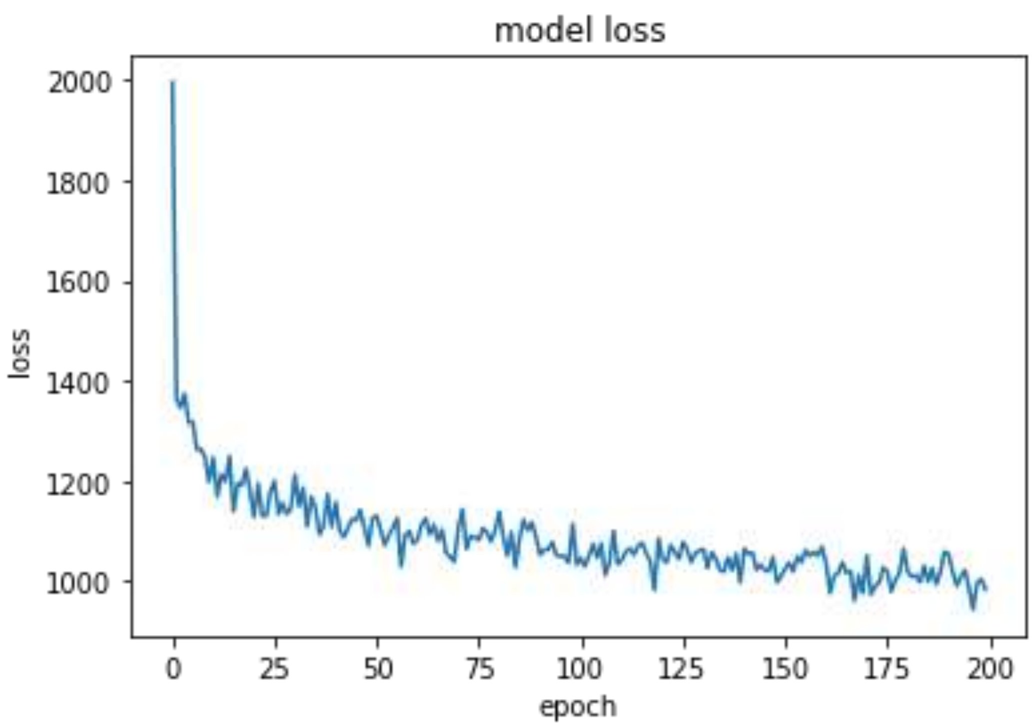


16. 4510_opt_200_16_2_lstm_128_64_cnnsoil_48_64: 19.6375690629349

    LSTM FC changed to 128 from 256 and soil CNN layers changed to 48 then 64 instead of 32 then 64.



17. 4510_opt_200_16_2_lstm_192_64_cnnsoil_48_64: 18.261329940802597
    LSTM FC changed to 192 from 256 and soil CNN layers changed to 48 then 64 instead of 32 then 64.



Conclusive remarks after trying out above variations of model:

- It is observed that learning rate of 0.002(with MAPE 19.37%) performs best out of tried learning rate of 0.001, 0.002, 0.01 and 0.03.
- For soil data CNN model specification were not given and two CNN layers with depth 48 and 64 one after other respectively worked best **(with MAPE 16.67%)** worked best out of tried combinations of (48,64), (32,64), (32, 48), (24,36), (24,48), (12,24) and single CNN of 64 depth.
- After flattening LSTM output of time series data two fully connected layers are applied. Two variation of number of neurons are tried out, (256,64) and (128,64). (128,64) neuron FC layers worked slightly better than (256,64), probably due to a smaller number of training set available.
- Removing dropout layer doesn't improve or degrade the model though purpose of dropout layer is to prevent overfitting.
- Increasing max-pooling size to (1,3) from (1,2) does not improve the results.

Best result obtained till now is MAPE of **16.67%** (obtained in 15th variant) as compared to authors' end season prediction MAPE of 8.1%.

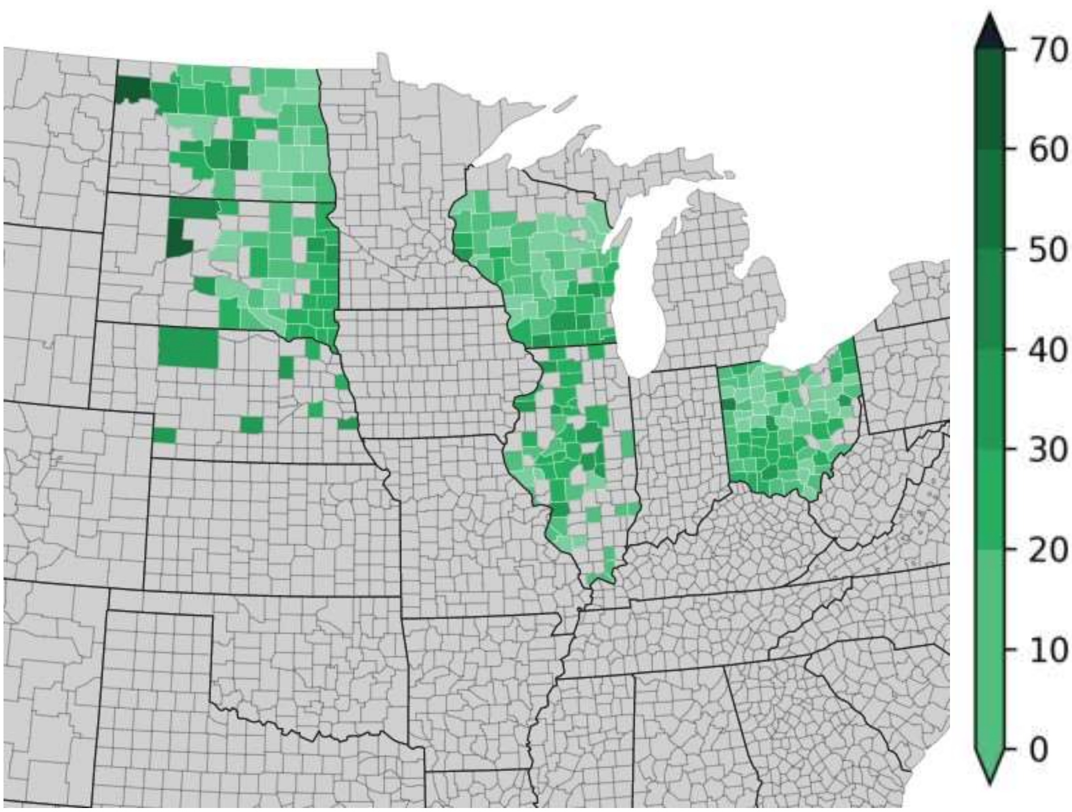


Fig 8. Map depicting MAPE a/c to color bar given on left in % MAPE.

Some important points to be noted which may be the responsible for the results obtained:

- Model is currently trained on maximum of 4672 as compared to 11750 as done by authors of the paper.
- All the details of the model are not provided by authors and many details of model are picked from “County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model” which is one of the recent publications of the authors and have similar implementation except the inclusion of static soil factor.

5. References

1. Jie Sun, Zulong Lai, Liping Di, Member, IEEE, Ziheng Sun, Jianbin Tao, and Yonglin Shen. Multi-level deep learning network for county-level corn yield estimation in the U.S. Corn Belt, 2020
2. Jie Sun, Liping Di, Ziheng Sun, Yonglin Shen and Zulong Lai. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model, 2019
3. USDA, "The USDA Economics, Statistics and Market Information System," https://usda.library.cornell.edu/?locale=en
4. NASS, "Usda national agricultural statistics service," https://www.nass.usda.gov/Quick Stats/index.php/, 2019
5. USDA-NASS, "USDA National Agricultural Statistics Service Cropland Data Layer," https://nassgeodata.gmu.edu/CropScape
6. E. Vermote, "MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 ," NASA EOSDIS Land Processes DAAC, doi:10.5067/MODIS/MOD09A1.006, 2015
7. Z. Wan, S. Hook, and G. Hulley, "MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006 ," NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MOD11A2.006, 2015.
8. P. Thornton, M. Thornton, B. Mayer, Y. Wei, R. Devarakonda, R.S.Vose, and R. Cook., "Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version3," ORNL DAAC. doi:10.5067/MODIS/MOD11A2.006, 2017
9. GEE, "Google Earth Engine," https://developers.google.com/earth-engine