Scaling up observations on plant phenology using remote sensing and machine learning

Kailun(Lucas) Jin[a,b], Duncan MacIntyre[a,c] Jingqi(Irena) Zhang[a,d]

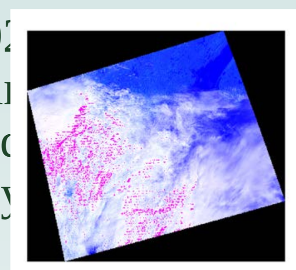
[a]Data Science Institute, University of British Columbia; [b]Department of Computer Science, University of British Columbia;
[c]Department of Physics & Astronomy, University of British Columbia; [d]Department of Electrical and Computer Engineering, University of British Columbia;

Abstract

Our project is to predict the date of plant blooming onset. The changes in the timing of flowering are essential signals of environmental changes. For example, the blooming time could influence the synchronization between plants and pollinators. It could also affect the sensitivity of crop yield to climate. Also, the timing of plant flowering is strongly related to pollen allergies. We are motivated to build a large-scale and more flexible tool to predict when and where the blooming will start based on the MODIS remote sensing satellite data, and we try to realize this goal with machine learning. We focus on predicting the flowering of lilacs and honeysuckle. We hope that our model will do a better job than the first bloom index, which is a short-term yearly forecast of the onset date of flowering published by the National Phenology Network. With remote sensing technology, we hope to directly grab relevant geographical features of the study regions and the blooming conditions of the plants for the machine learning model, so that predicting the onset of flowering of other plant species or even crops becomes possible.

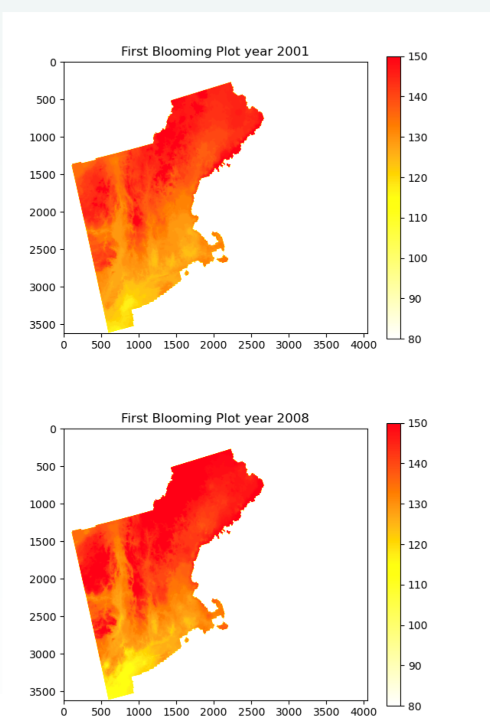
Introduction

1. We used satellite images in 2000-202 MODIS as the primary data for the i features to our machine learning mod
   - We train our model with both 8-day and daily 250m resolution data
   - 8-day data: with 13 bands; daily data with 8 bands



MODIS_daily250_small_2022-05-19 via QGIS

1. Complimentary data --- the land cover data/ properties data/ PRISM climate data
   - The land cover data: the study regions' land classification region (2001-2020)
   - Properties data: the longtitude and latitude information for each pixel within the study regions
   - PRISM climate data: temperature/ precipitation information

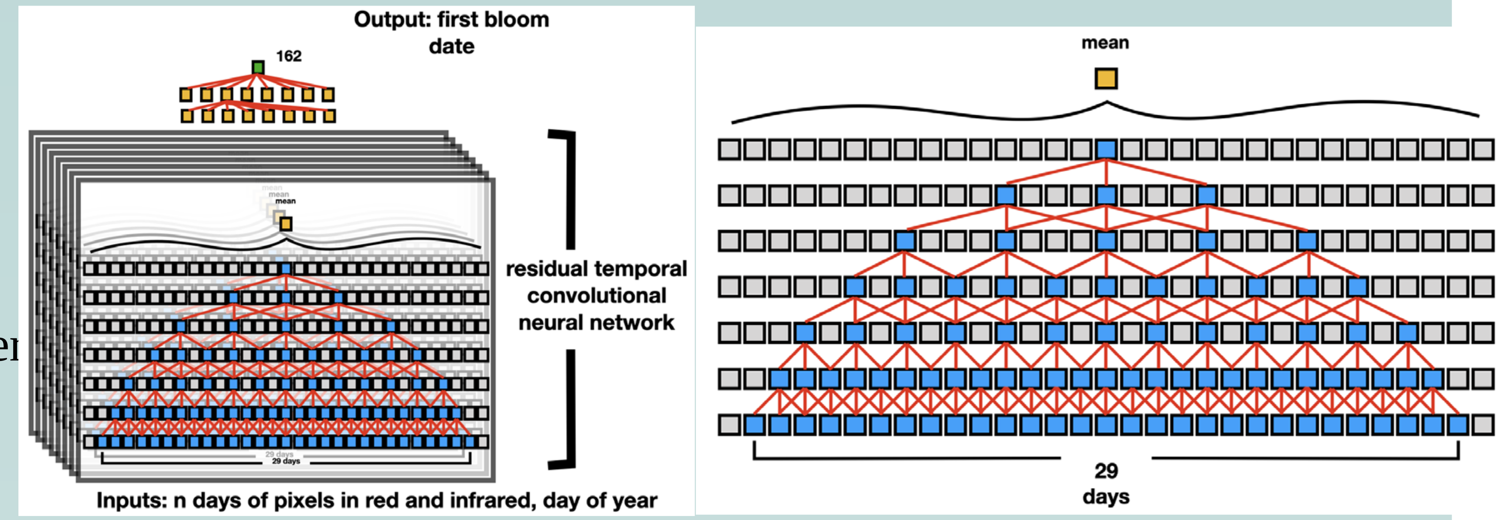


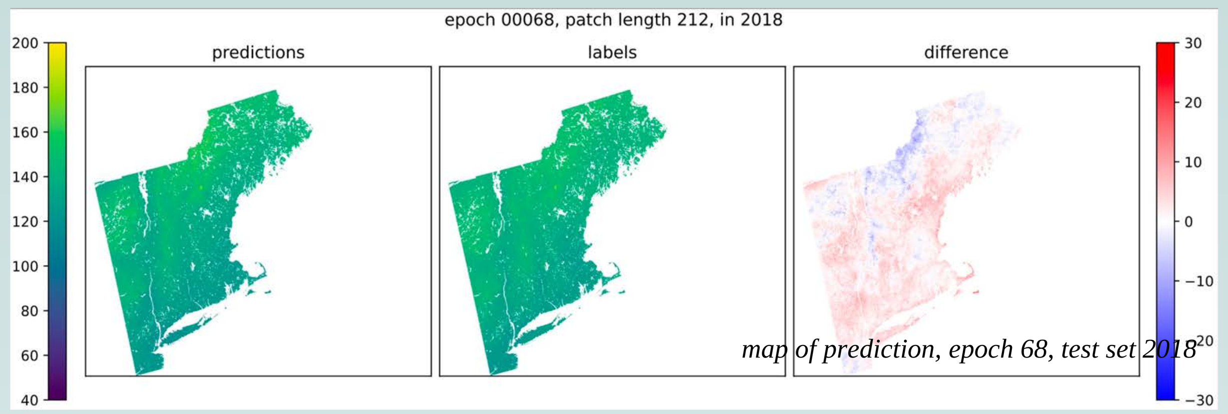
3. First Bloom Index
- A temperature-based metric provided by the National Phenology Network
- The metric we trained our model to predict
- A rough proxy to the flowering onset date based on historical observation and temperature

Methodology

Satellite images from MODIS

- Input: MODIS satellite data
1. One pixel over n days (red and infrared)
2. day of year
3. cosine calendar
- Data from different years are used in differ
1. 2001-2014: training set
2. 2015-2017: Validation set
3. 2018-2020: test set
- Output: First Bloom Date





map of prediction, epoch 68, test set 2018

Maos of Prediction

- 3 types of maps
1. Maps of prediction
2. Maps of labels
3. Maps of differences (labels minus prediction)
- Comparing the maps plotted with different models from some selected epochs along one single training process

Transfer learning

- We use either of these methods to pad other places:
1. Padding those places with fake (but reasonable) data. We use median in our project.
2. Pad those places with nothing. In other words we have shorter input (ignore other places completely, since we have variable input length)

Selection of study Region

- 2 main direction:
1. the effect of the size of study regions or the dataset
2. the effect of randomly selected pixels

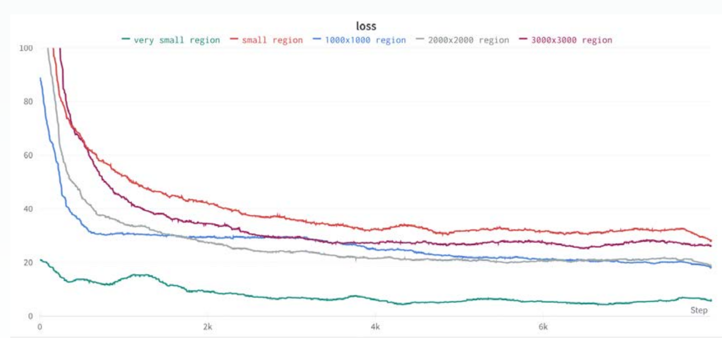
Results

Training loss and validation loss

- Training loss would decrease when there is a reduction in the size of the study region
- Validation loss would decrease when there is a reduction in the size of the study region
-The plot of the validation loss shares a similar pattern as training loss
- Over-fitting problem persists even with the smallest study region as the training set.

Model Improvements

- Models improvement (map)
1. differences between prediction and labels (no matter is the prediction date earlier/later than the First Bloom Index) reduces
2. Colors darkening.



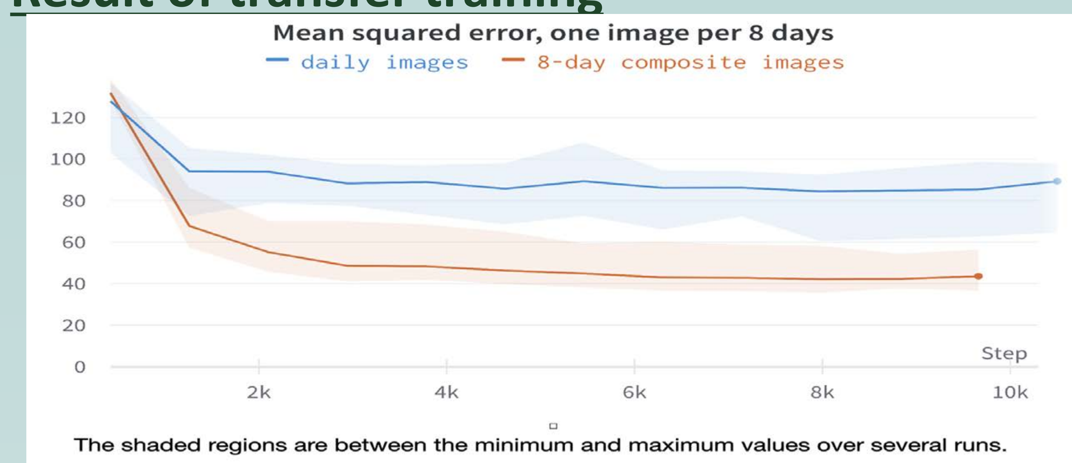
Comparison Group 1 Training Loss

Comparison Group 1 Validation Loss

map of prediction, epoch 68, test set 2020

Results (cont'd)

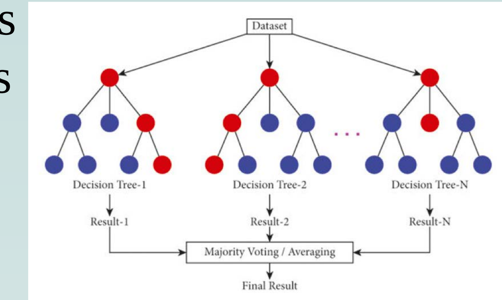
Result of transfer training



Conclusions

- Future Work
- - Train on ground truth labels
- - Compare with other models
- - Use spatial information
1. 1. Blur in preprocessing
2. 2. Spatial convolution
- - Use an ensemble model

- Compare with Other Models



- A Better Model for Social Goods
- Ideally, a perfect model could predict fairly accurate, so that:
- - Guiding artificial pollinator placement
- - Improving honeysuckle quality and drive the development of related - medicine industries
- - Maintain a better ecological balance by improved ability to intervene

Acknowledgements

- We would like to express our deepest appreciation to all people at UBC Data Science Institute, especially Dr. Raymond Ng;
- We would also like to extend our gratitude to Dr. Joséphine who gave us this topic and datasets, and always helping us as a sponsor.
- We would also like to recognize the advices from Dr. Mathias Lecuyer, who helped us multiple times during lunch time!

References

[1]Ault, T. R., et al. (2015). Trends and Natural Variability of Spring Onset in the Coterminous United States as Evaluated by a New Gridded Dataset of Spring Indices, Journal of Climate, 28(21), 8363-8378.
[2]Qinchuan Xin, et al. Evaluations and comparisons of rule-based and machine-learning-based methods to retrieve satellite-based vegetation phenology using MODIS and USA National Phenology Network data, International Journal of Applied Earth Observation and Geoinformation, Volume 93, 2020, 102189, ISSN 1569-8432.
[3] Toby R. Ault, et al.A Matlab© toolbox for calculating spring indices from daily meteorological data,Computers & Geosciences,Volume 83,2015,Pages 46-53,ISSN 0098-3004