Predicting Soil Salinity from Satellite Data: Applications in Estimating Agricultural Damages in California’s Central Valley

Amanda Quay  
*Engineering and Public Policy, Civil and Environmental Engineering*  
*Carnegie Mellon University*  
*Pittsburgh, PA, United States*

1. Abstract

Soil salinity affects agricultural production in arid regions worldwide. Soil salinity reduced agricultural revenues in California’s critical agricultural regions by an estimated 3.7 billion dollars in 2014 alone (1). High-density salinity sampling is costly, yet crucial for assessing the impact of salinization on agriculture. Remote sensing by satellite imagery analysis can offer a cheap and synoptic approach to salinity monitoring. Free satellite repositories contain cleaned and calibrated satellite imagery, as well as additional spatially-resolved variables that may improve soil salinity predictions (e.g., precipitation, elevation).

In this project, machine learning techniques were applied to a dataset of Landsat images for the western San Joaquin valley and ground-truthed, high-fidelity salinity data for 22 agricultural fields comprising 542 hectares, broken into over 5000 individual pixel features. This dataset has been used to create salinity maps using the Canopy Response Salinity Index (CRSI) and simple regressions. (2). The goal of the project was to develop an improved regional-level salinity predictor capable of estimating salinity in fields at a high resolution (30m resolution) using neural network and random forest models. Multi-year Landsat spectral bands have been shown to provide a reliable signal for indirect detection of soil salinity through remote sensing (3) By observing multi-year Landsat spectral bands into a vegetation index (i.e., CRSI), reliable spectral signals can be used to estimate soil salinity by observing vegetation and using it as an indicator for salinity (4).

Policy applications of a reliable remotely-sensed salinity estimation include impact assessment of soil salinity on regional agriculture, and improvements in yield predictors by incorporating improved salinity models into existing synthetic datasets (5).
2. Proposal Details

Univariate and multivariate models developed by Scudiero et al. provide baseline salinity predictions (2). After replicating these efforts, random forest and neural network algorithms were applied and compared to the traditional predictive models. For all predictive models, a leave-one-field-out (LOFO) strategy will be employed to ensure that the features passed into the algorithm don’t represent the outcome data. Predictions improved with these machine learning techniques when one anomalous field was excluded from the dataset, offering a highly resolved measure of soil salinity without costly ground-truthing.

With the excluded field, we explored the minimum number of 'sample' pixels required to generate a reasonable accurate salinity map of a new field. This analysis allows discussion of the limitations and extent of the ML algorithms developed for this application and how translatable this technique is to a regional or global analysis.

References


Appendix I: Dataset Variables

Fields are split into 30m-by-30m pixels to match the Landsat imagery, increasing the dataset from 22 fields to over 5000 pixel points. Some features are field-resolved, and some are pixel-resolved.

- ’Field_ID’: the unique field identifier (1-22), field-resolution
- ’aspect’: the compass direction that the pixel’s slope faces, pixel-resolution
- ’average_temperature’: the average temperature in the year of the outcome data (2013), field-resolution
- ’band_x_mean_2007’ through ’band_x_mean_2013’: each of the 7 Landsat spectral bands (x = [1, 7]) has been averaged annually for the test year, as well as the six years previous for better multi-year stability of salinity estimated from the spectral data, pixel-resolution
- ’elevation’: elevation at pixel
- ’maxCRSI_year’: year that maximum CRSI value for a pixel was observed
- ’max_CRSI’: maximum CRSI value measured at pixel over all years considered
- ’salinity’: (Y) salinity at pixel
- ’slope’: field slope at pixel
- ’total_precipitation’: annual total precipitation collected from Google Earth Engine for 2013
- ’ucr_precipitation’: provided by UC Riverside with the salinity dataset
- ’ucr_temperature’: provided by UC Riverside with the salinity dataset Keras was used for the neural network models tested and sklearn was used for the random forest models tested in an ipython environment.