World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:11, No:07, 2017

Model-Driven and Data-Driven Approaches for Crop Yield Prediction: Analysis and Comparison

Authors : Xiangtuo Chen, Paul-Henry Cournéde

Abstract : Crop yield prediction is a paramount issue in agriculture. The main idea of this paper is to find out efficient way to predict the yield of corn based meteorological records. The prediction models used in this paper can be classified into model-driven approaches and data-driven approaches, according to the different modeling methodologies. The model-driven approaches are based on crop mechanistic modeling. They describe crop growth in interaction with their environment as dynamical systems. But the calibration process of the dynamic system comes up with much difficulty, because it turns out to be a multidimensional non-convex optimization problem. An original contribution of this paper is to propose a statistical methodology, Multi-Scenarios Parameters Estimation (MSPE), for the parametrization of potentially complex mechanistic models from a new type of datasets (climatic data, final yield in many situations). It is tested with CORNFLO, a crop model for maize growth. On the other hand, the data-driven approach for yield prediction is free of the complex biophysical process. But it has some strict requirements about the dataset. A second contribution of the paper is the comparison of these model-driven methods with classical data-driven methods. For this purpose, we consider two classes of regression methods, methods derived from linear regression (Ridge and Lasso Regression, Principal Components Regression or Partial Least Squares Regression) and machine learning methods (Random Forest, k-Nearest Neighbor, Artificial Neural Network and SVM regression). The dataset consists of 720 records of corn yield at county scale provided by the United States Department of Agriculture (USDA) and the associated climatic data. A 5-folds cross-validation process and two accuracy metrics: root mean square error of prediction(RMSEP), mean absolute error of prediction(MAEP) were used to evaluate the crop prediction capacity. The results show that among the data-driven approaches, Random Forest is the most robust and generally achieves the best prediction error (MAEP 4.27%). It also outperforms our model-driven approach (MAEP 6.11%). However, the method to calibrate the mechanistic model from dataset easy to access offers several side-perspectives. The mechanistic model can potentially help to underline the stresses suffered by the crop or to identify the biological parameters of interest for breeding purposes. For this reason, an interesting perspective is to combine these two types of approaches.

Keywords : crop yield prediction, crop model, sensitivity analysis, paramater estimation, particle swarm optimization, random forest
Conference Title : ICCSDA 2017 : International Conference on Computational Statistics and Data Analysis
Conference Location : London, United Kingdom
Conference Dates : July 24-25, 2017