**INTRODUCTION**

As with many bio-systems, plant growth is a highly complex and dynamic environmentally linked system. Therefore, growth and yield modeling is a significant scientific challenge . Modeling approaches vary in a number of aspects (including, scale of interest, level of description, integration of environmental stress, etc.). According to (Todorovski and Dzeroski, 2006; Atanasova et al., 2008) two basic modeling approaches are possible, namely, "knowledge-driven" or "data-driven" modeling. The knowledge driven approach relies mainly on existing domain knowledge. In contrast, a data-driven modeling approach is capable of formulating a model solely from gathered data without necessarily using domain knowledge.

Data driven models (DDM) include classical Machine Learning techniques, artifical neural networks (Daniel et al., 2008), support vector machines (Pouteau et al., 2012), and generalized linear models. Those methods have many desirable characteristics, such as imposing fewer restrictions, or assumptions, the ability to approximate nonlinear functions, strong predictive abilities, and the flexibility to adapt to inputs of a multivariate system (Buhmann, 2003). According to Singh et al., 2016 and reviewed by Liakos et al., 2018 Machine Learning (ML), linear polarizations, wavelet-based filtering, vegetation indices (NDVI) and regression analysis are the most popular techniques used for analyzing agricultural data. However and besides the aforementioned techniques, a new methodology which is recently gaining momentum is deep learning (DL)(Goodfellow et al., 2016). DL belongs to the machine

learning computational field and is similar to ANN. However, DL is about “deeper” neural networks that provide a hierarchical representation of the data by means of various operations. This allows larger learning capabilities, and thus higher performance and precision. A strong advantage of DL is feature learning, i.e., automatic feature extraction from raw data, with features from higher levels of the hierarchy being formed by composition of lower level features (Goodfellow et al., 2016). DL can solve more complex problems particularly well, because of the more complex related models (Pan and Yang, 2010). These complex models employed in DL can increase classification accuracy and reduce error in regression problems, provided there are adequately large data-sets available describing the problem.

Gonzalez-Sanchez et al.( 2019) presented a comparative study of ANN, SVR, M5-prime, KNN ML techniques and Multiple Linear Regression for crop yield prediction in ten crop datasets. In their study, Root Mean Square Error (RMS), Root Relative Square Error (RRSE), Normalized Mean Absolute Error (MAE) and Correlation Factor (R) were used as accuracy metrics to validate the models. Results showed that M5-Prime achieved the lowest errors across the produced crop yield models. The results of that study ranked the techniques from the best to the worst, according to RMSE, RRSE, R, and MAE resulting, in the following order: M5-Prime, kNN, SVR, ANN and MLR. Another study by (Nair and Yang-Won, 2016) applied four ML techniques, SVM, Random Forest (RF), Extremely Randomized Trees (ERT) and Deep Learning (DL) to estimate corn yield in Iowa State. Comparisons of the validation statistics showed that DL provided more stable results, overcoming the overfitting problem.

Stem diameter is considered as one of the important parameters describing the growth of plants during vegetative growth stage. Also, the variation of stem diameter has widely been used to derive proxies for plant water status and, is therefore applied in optimisation strategies for plant-based irrigation scheduling in a wide range of species. Plant stem diameter variation (SDV) refers to plant stem periodic shrinkage and recovery movement during the day and night, and this periodic variation is related to plant water content and can be used as an indicator of the plant water content change. During active vegetative growth and development, crop plants rely on the carbohydrate gained from photosynthesis and the translocation of photo-assimilates from the site of synthesis to sink organs (Yu et al., 2015).

The fundamentals of stem diameter variations have been well documented in a substantial amount of literature (Vandegehuchet et al., 2014). It has been documented that SDV is sensitive to water and nutrient conditions and is closely related to the responses of crop plants to the changes of environmental conditions (Kanai et al., 2008). The stem diameter is an important parameter describing the growth of crop plants under abiotic stress during vegetative growth stage. Therefore, it is important to generate stem diameter growth models able to predict the response of SDV to environmental changes and plant growth under different conditions. Many studies emphasize the need to critically review and improve SDV models for assessment of environmental impact on crop growth (Hinckley and Bruckerhoff, 2011). SDV daily models have been developed to accurately predict inter-annual variation in annual growth in balsam fir (Abies balsamea L) (Duchesene and Houle, 2011). Inclusion of daily data in growth-climate models can improve predictions of the potential growth response to climate by identifying particular climatic events that escape to a classical dendroclimatic approach (Duchesene and Houle, 2011). However, models for predicting SDV and plant growth using environmental variables have so far remained limited.

Tomato crop growing in greenhouse environment is considered as a dynamic and complex system, with few models having been studied for it up to now. In the literature TOMGRO and TOMSIM (Jones et al., 1999), (Heuvelink, 1996) are considered as the main applicable dynamic growth models. Those models are dependent on physiological processes, and they represent biomass partitioning, crop growth, and yield as a function of several climate and physiological parameters. However, due to their limited application to practical

settings, their complexity, the difficulty in estimating initial parameter values and the need for calibration and validation in every new environment, growers uptake has been limited.

The Tompousse model was developed by (Abreu et al., 2000) to predict tomato yield in terms of the weight of harvested fruits. The model was developed by examining the relationship between environmental parameters in a heated greenhouses in the Southern part of France. A linear relationship between flowering rate and fruit growth was the basic assumption used in this model. However, the model performance was poor when tested in unheated plastic greenhouses in Portugal. Another tomato yield model was proposed by Adams (Adams, 2002), based on a form of graphical simulation tool. The main objective of the model was to represent weekly fluctuations of greenhouse tomato yield in terms of fruit size and harvest rate. Hourly climate data were used to estimate the rate of growth of leaf truss and the flower production. Yield seasonal fluctuations were generally infuenced by periodic variations of solar radiation and air temperature. According to (Qaddoum et al., 2013), there is a large number of tools that can help farmers in making decisions. These can provide yield rate prediction, suggest climate control strategies, synchronise crop production with market demands.