

EGU2020-13035

<https://doi.org/10.5194/egusphere-egu2020-13035>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



A test development of a data driven model to simulate chlorophyll data at Tongyeong bay in Korea

Sung Dae Kim and Sang Hwa Choi

Korea Institute of Ocean Science and Technology, Busan, Korea, Republic of (sdkim@kiost.ac.kr)

A pilot machine learning(ML) program was developed to test ML technique for simulation of biochemical parameters at the coastal area in Korea. Temperature, chlorophyll, solar radiation, daylight time, humidity, nutrient data were collected as training dataset from the public domain and in-house projects of KIOST(Korea Institute of Ocean Science & Technology). Daily satellite chlorophyll data of MODIS(Moderate Resolution Imaging Spectroradiometer) and GOCI(Geostationary Ocean Color Imager) were retrieved from the public services. Daily SST(Sea Surface Temperature) data and ECMWF solar radiation data were retrieved from GHRSSST service and Copernicus service. Meteorological observation data and marine observation data were collected from KMA (Korea Meteorological Agency) and KIOST. The output of marine biochemical numerical model of KIOST were also prepared to validate ML model. ML program was configured using LSTM network and TensorFlow. During the data processing process, some chlorophyll data were interpolated because there were many missing data exist in satellite dataset. ML training were conducted repeatedly under varying combinations of sequence length, learning rate, number of hidden layer and iterations. The 75% of training dataset were used for training and 25% were used for prediction. The maximum correlation between training data and predicted data was 0.995 in case that model output data were used as training dataset. When satellite data and observation data were used, correlations were around 0.55. Though the latter correlation is relatively low, the model simulated periodic variation well and some differences were found at peak values. It is thought that ML model can be applied for simulation of chlorophyll data if preparation of sufficient reliable observation data were possible.