

Development Of Soil, Water And Carbon Datasets For Land Surface Modeling

Wei Shangquan¹(shgwei@mail.sysu.edu.cn), Qingliang Li², Feini Huang¹, Gaosong Shi¹, Cheng Zhang², Zili Xiong¹, Wenyue Sun¹, Yongjiu Dai¹
 1 School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, China; 2 Changchun Normal University, Changchun, China

A 1-km daily soil moisture dataset over China using in-situ measurement and machine learning

This dataset is derived with in-situ measurements of 1,648 stations, named as SMC11.0 (Soil moisture over China based on In-situ data). SMC11.0 provides 10-layer soil moisture with 10 cm intervals up to 100 cm deep at daily resolution over the period 2000-2020. Random Forest is used to predict soil moisture using ERA5-land time series, leaf area index, land cover type, topography and soil properties as covariates. Using in-situ soil moisture as the benchmark, two independent experiments are conducted to investigate the estimation accuracy of the SMC11.0: year-to-year experiment (ubRMSE ranges from 0.041-0.052 and R ranges from 0.883-0.919) and station-to-station experiment (ubRMSE ranges from 0.045-0.051 and R ranges from 0.866-0.893). SMC11.0 can be useful for various hydrological, meteorological, and ecological analyses and modeling, especially for those applications requiring high resolution SM maps.

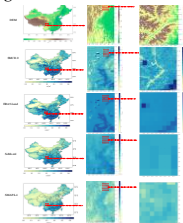


Fig.1 Soil moisture maps from different products on 1st January 2016

Downloading: <https://doi.org/10.11888/Terre.tpdc.272415>

Citation: Li, Q., Shi, G., Shangquan, W., Nourani, V., Li, J., Li, L., Huang, F., Zhang, Y., Wang, C., Wang, D., Qiu, J., Lu, X., & Dai, Y. (2022). A 1 km daily soil moisture dataset over China using in situ measurement and machine learning. *Earth Syst. Sci. Data*, 14, 5267-5286. <https://doi.org/10.5194/essd-14-5267-2022>

A 1-km 10-day global carbon fluxes dataset using in-situ measurement

This dataset is derived with in-situ measurements of 280 stations, named as GCFD (Global carbon fluxes dataset). GCFD provides three carbon flux variables including gross primary production (GPP), terrestrial ecosystem respiration (RECO) and net ecosystem exchange (NEE) at 10-day resolution over the period from January 1999 to June 2020. Convolutional Neural Network is used to predict carbon fluxes using fraction of absorbed photosynthetically active radiation, leaf area index, 2 m temperature, solar shortwave radiation downward, latent heat, sensible heat, soil temperature and soil moisture as covariates. Using in-situ carbon fluxes as the benchmark, two independent experiments are conducted to investigate the estimation accuracy of the GCFD: station-to-station experiment (RMSE ranges from 1.580-1.772 $gC\ m^{-2}\ d^{-1}$ and R^2 ranges from 0.54-0.78) and year-to-year experiment (RMSE ranges from 2.174-2.270 $gC\ m^{-2}\ d^{-1}$ and R^2 ranges from 0.20-0.66). GCFD can be useful for applications requiring high resolution carbon flux maps.

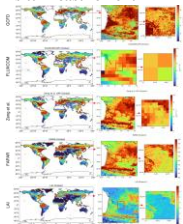


Fig.3 GPP maps of the global and an oceanic climate area from different products on 10 July 2010

Downloading: <https://doi.org/10.11888/Terre.tpdc.300009>

Citation: Shangquan, W., Xiong, Z., Nourani, V., Li, Q., Lu, X., Li, L., Huang, F., Zhang, Y., Sun, W. & Dai, Y. (2023). A 1 km Global Carbon Flux Dataset Using In Situ Measurements and Deep Learning. *Forests*, 14(5): 913. <https://doi.org/10.3390/f14050913>.

A China dataset of soil properties for land surface modeling (version 2)

This dataset maps 23 soil physical, chemical, and fertility properties at six standard depth layers (0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm) across China with a spatial resolution of 90 meters. The source data includes 8,979 soil profiles from the Second National Soil Survey of China, 1,641 soil profiles from the World Soil Information Service, 76 soil profiles from the First National Soil Survey of China, and 614 soil profiles from regional databases. This dataset includes pH, sand, silt, clay, bulk density, organic carbon content, gravel, alkali-hydrolysable nitrogen, total nitrogen, cation exchange capacity, porosity, total potassium, total phosphorus, available potassium, available phosphorus, and soil color (in both Munsell and RGB formats). The dataset is provided in raster format, available in both Tiff and netCDF formats. The units for each soil property are detailed in the readme documentation. This dataset offers superior data quality compared to the first edition and is suitable for a wide range of geoscience research, including land surface process modeling.

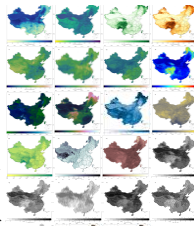


Fig.2 Soil property maps of China at 0.5 cm depth interval

Downloading: <https://doi.org/10.11888/Terre.tpdc.301235>

Citation: Shi G, Shangquan, W. (2024). A China dataset of soil properties for land surface modeling (version 2), submitted to *Earth System Science Data*.

LandBench 1.0: A Benchmark Dataset and Evaluation Metrics for Data-Driven Land Surface Variables Prediction

We provide a set of benchmark data LandBench1.0, used for data driven land surface variables prediction. The data set is based on ERA5 - land, ERA5, SoilGrid, the SMSC and MODIS data sets generated and provided 0.5, 1, 2, and 4 degrees resolution daily global data. Data preprocessing, convenient to use in a data-driven model. We also provide a benchmark data set kit, available at <https://github.com/2023ATAI/LandBench>. This toolkit promoted the existing method to achieve, a new prediction model of development and utilization of the unified evaluation index, the kit also included used in high resolution land surface variables predicted global address mapping technology. Benchmark data set for the purpose of computer and the earth system is to encourage cooperation between scientists, in order to develop effective land surface variables prediction based on machine learning data driven model.

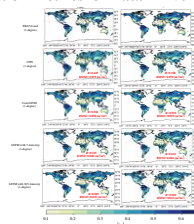


Fig.4 Global soil moisture (m³/m³) of ERA5-land and various deep learning predictive models at 1-degree resolution

Downloading: <https://doi.org/10.11888/Atmos.tpdc.300294>

Citation: Li, Q., Zhang, C., Shangquan, W., Wei, Z., Yuan, H., Zhu, J., Li, X., Li, L., Li, Gan, Liu, P., & Dai, Y. (2024). LandBench 1.0: A benchmark dataset and evaluation metrics for data-driven land surface variables prediction. *Expert Systems with Applications*, 243, 122917. <https://doi.org/10.1016/j.eswa.2023.122917>.