Project: **1318**

Project title: **CLINT - climate intelligence**

Principal investigator: **Christopher Kadow**

Report period: **2022-05-01 to 2023-04-30**

Project overview

The main objective of CLINT (https://climateintelligence.eu/) is the development of an Artificial Intelligence framework composed of Machine Learning algorithms to process big climate datasets for improving Climate Science in the detection, causation, and attribution of Extreme Events, including tropical cyclones, heatwaves and warm nights, and extreme droughts, along with compound events and concurrent extremes. For this purpose, partners of the CLINT project need to have access to computational resources and storage that allow them to train the developed Machine Learning algorithms. The CLINT project has 15 partners, including DKRZ, POLIMI (Politecnico di Milano), JLU (University Justus-Liebig - Giessen) and ECMWF (European Centre for Medium-Range Weather Forecasts). The project is part of the H2020 Programme supported by the European Union.

Partner contributions

DKRZ is contributing to the CLINT project by developing a machine learning application called CRAI (Climate Reconstruction AI): https://github.com/FREVA-CLINT/climatereconstructionAI. CRAI is a deep learning based inpainting technology to reconstruct missing values in climate datasets. In the context of the CLINT project, CRAI has been used to train models to reconstruct observational datasets of extreme events such as HadEX3 (https://www.metoffice.gov.uk/hadobs/hadex3/) in order to support the detection, causation and attribution of extreme events. The training of the models has been carried out on the GPU nodes. Most of the resources have been devoted to the hyperparameter tuning, i.e., the search of the configuration of parameters which gives the best performance on the validation set with the selected evaluation metrics.

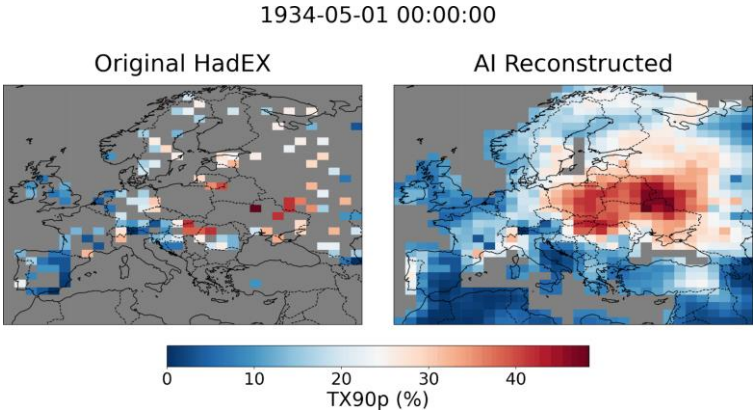

Figure 1: Example output of the AI infilled observational HadEX dataset.

CRAI has been selected to be the first machine learning algorithm to be deployed as a Web Processing Service (WPS) for AI-enhanced Climate Science. This task has been achieved by creating a WPS called “Duck” (https://github.com/climateintelligence/duck) by using the PyWPS library (https://pywps.org/) and the cookiecutter from the Birdhouse ecosystem (https://github.com/bird-house/cookiecutter-birdhouse). The service has been deployed on a virtual machine at DKRZ and automated using Ansible playbook scripts. https://clint.dkrz.de/

Polimi’s application is a contribution to Climate Intelligence as a deep learning application for synthetic generation of Extra-Tropical Cyclones (ETCs) atmospheric fields in the North Atlantic (NA). These events are a major driver of flooding, extreme rainfall and wind speeds, and storm surges in coastal regions. In the first step the ProGAN was trained using the Python deep learning API Keras based on TensorFlow platform. For the second step of the application, we

started from the ERA5 reanalysis data of mean sea level pressure, meridional and zonal wind component, and the rainfall rate in the domain (Lat. 20°N – 70°N Long. 100°W – 20°E) for the period going from January 1959 to December. Now we are training the network on a resized training set with examples resized to 128x128 to have a first glamps of result on the new domain. The training of the model is performed in PyTorch framework and the NVIDIA CUDA Deep Neural Network library (cuDNN).

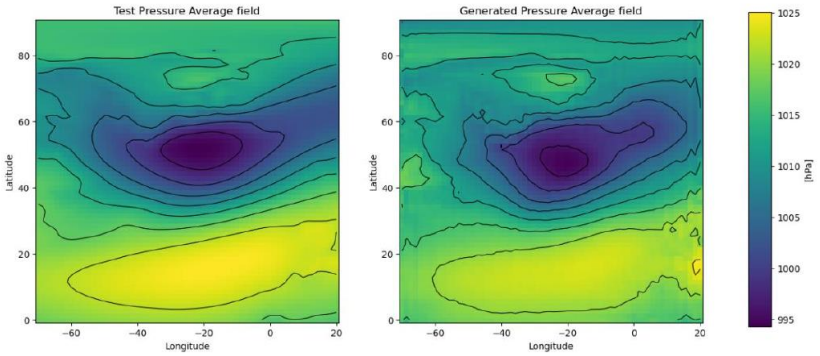

Figure 2: Example output of generated pressure field by the GAN.

JLU's is aiming at identifying connectivities of large-scale heatwaves and droughts of global scale. Artificial intelligence is used to enhance the method and solve identified weaknesses. The downscaling and bias adjustment pipelines are major AI pipelines also AI-driven utilizing Convolutional Neural Networks built with the principles of UNETs and incorporating multi-modal input and ensemble output processes.

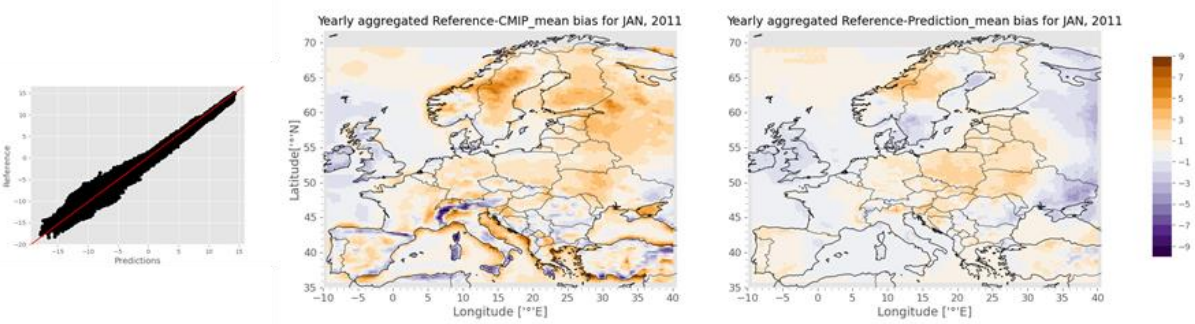

Figure 3: Example output metrics from the downscaling and bias adjustment pipelines.

Conclusion of the compute period

The CLINT project was in its first phase. It produced promising scientific results and important services. The upcoming phase will work upon these results and accelerate scientific output. All marked conference contributions are in preparation for journal publications. The resources allocated are well used, and probably will use all requested resources.

Resource	Granted	Utilization	Remaining
Levante CPU nodes (Node hours)	2022-07-01: 2700 2023-01-01: 6300 Total: 9000	5766	3234
Levante GPU nodes (Node hours)	2022-07-01: 14000 2023-01-01: 16000 Total: 30000	19662	10338
Levante storage (TiB)	2022-07-01: 60 2023-01-01: 205 Total: 265	31	234
Swift Object Storage (GiB)	10240		10240