Imperial College London
Department of Earth Science and Engineering
MSc in Applying Computational Science and Engineering

Independent Research Project

Project Plan

Deep learning surrogate model for wildfire probability prediction on a global scale

by

Caili Zhong

Email: caili.zhong21@imperial.ac.uk

GitHub username: acse-cz421

Repository: https://github.com/ese-msc-2021/irp-cz421

Supervisors:

Dr. Sibo Cheng

Dr. Matthew Kasoar

July/September 2022

1. Introduction

1.1 Motivation, Aims and Objective

Fire is one of the most destructive disasters, and every year fires have a significant economic, social, and environmental impacts on a global scale. Especially in the case of forest fires, which are often sudden and extremely destructive and difficult to deal with immediately [1]. Globally, an average of more than 200,000 forest fires occurs each year, burning more than 1% of the world's forested area [2]. In addition, wildfires not only destroy trees, but also cause serious damage to the forest environment and ecological structure. The physicochemical properties of the soil will also change with the occurrence of wildfire, which can result in serious ecosystem imbalances [3]. Therefore, it is important to study real-time monitoring and early warning of fire danger in key forest protection areas. Early warning of wildfire can help to extinguish fires at an early stage and significantly reduce the cost and damage of fighting them. The main direction of this project is to predict the probability of wildfires.

Furthermore, global fire frequency is related to land use, vegetation type and meteorological factors. When fire risk weather meets combustible vegetation, it may cause wildfires, and fires can damage vegetation and soil health when they occur. Data from validated natural environment models can then be used to make predictions about forest fires. Numerous valid and relevant models have been constructed, such as Earth System Models [4] and Dynamic Global Vegetation Models (DGVM) [5]. The Joint UK Land Environment Simulator-INteractive Fire and Emissions algorithm for Natural envirOnments (JULES-INFERNO) is an example of the DGVM, which can effectively simulate the probability of fire occurrence based on geographic features (vegetation, climate, etc.) [6]. However, implementing such a surface simulation is extremely challenging and computationally expensive due to the complexity of the physical model and geographical features. For example, predicting the probability of a 100-year wildfire would take several weeks [7]. It is necessary to build a surrogate model to address the computational cost issue.

The aim of this project is to propose a deep learning surrogate model based on JULES-INFERNO that predicts the probability of wildfire occurrence in future by feeding in a set of time-series wildfire data, which could significantly improve forecasting efficiency. The built model will be trained using the data generated by JULES-INFERNO and the trained model will be tested and evaluated using unseen scenarios. To address the problem of oversized features, the project uses Convolutional Autoencoder (CAE) and Principal Component Analysis (PCA) to reduce dimensionality. In order to construct the most suitable surrogate model, this project plans to build three deep learning prediction models (PCA + Long Short Term Memory (LSTM), CAE+LSTM, Convolutional LSTM (ConvLSTM)). This paper presents some related research, project methodology, preliminary implementation, as well as a plan for future work.

1.2 Literature Review

JULES is a model that can simulate land surface vegetation, primarily using physical models to simulate the processes of land-use, fire, and climate interactions with vegetation dynamics. However, the model formulation of disturbances, particularly fire, drought, and tree mortality, is not sufficiently constrained [8]. INFERNO was constructed based on a simplified parameterisation of fire counts proposed by Pechony and Shindell, which was experimentally shown to provide an efficient prediction of large-scale fires based on climatology variables [9]. In 2016, JULES invoked a new fire disturbance term from the INFERNO model, and the new JULES-INFERNO model can effectively represent the complexity and interaction between land-use change and fire [6]. However, such a model would require plenty of simulations, which would also increase computational cost. A method to improve this situation is to use a surrogate model substitutes the original high precision simulation model. The surrogate model aims to output a result that approximates the original model but is less computationally intensive to solve [10]. The next section describes the approach chosen for this project to implement this surrogate model.

There are numerous models exist for wildfire prediction, with a mass of papers focusing on the use of Machine Learning (ML) methods to directly predict the final area burned by wildfires. Cortez and Morais [11] compared four different ML methods (Decision Tree, Random Forest, Artificial Neural Network and Support Vector Machines (SVM)) to predict burned area using fire and climate data from Montesano Natural Park in north-eastern Portugal and showed that SVM had the best prediction performance. Zhang et al. [12] compared Convolutional Neural Networks (CNN) and SVM with wildfire data from Yunnan Province, China, and found that for fire monitoring (classification), that the accuracy of CNN is 13.7% higher than that of SVM. Liang et al. [13] compared Back Propagation Neural Networks, Recurrent Neural Networks and LSTM using data on the scale of wildfires in Alberta, Canada, and showed that LSTM had the highest accuracy (90.9%). For the time series prediction problem, Cao et al. [14] found that using a combination of LSTM and CNN had better fire prediction performance. Therefore, this project plans to implement an LSTM neural network combined with a CNN for wildfire probability prediction.

2. Methodology and Preliminary Results

2.1 Methodology

The main objective of this project is to implement a deep learning surrogate model for global wildfire probability prediction based on the JULES-INFERNO model. Firstly, collecting the data generated by the JULES-INFERNO model, the variables to be used in this project were extracted from the original data and pre-processed them according to the network model requirements, then divided the dataset into training and test sets. Secondly, due to the large feature size of the variables, the data needs to be dimensional reduction. The project will then use the reduced dimensional data for subsequent predictions. The current relatively mature LSTM is chosen as the core for wildfire prediction model. After building the initial prediction model, the network will be trained using the training set, and then the trained network will be evaluated using the test set to determine whether the model is satisfactory. Finally, the prediction model will be updated by adjusting the network parameters or changing the network structure to improve the accuracy and flexibility. The basic flowchart of the study is shown in Figure 1.

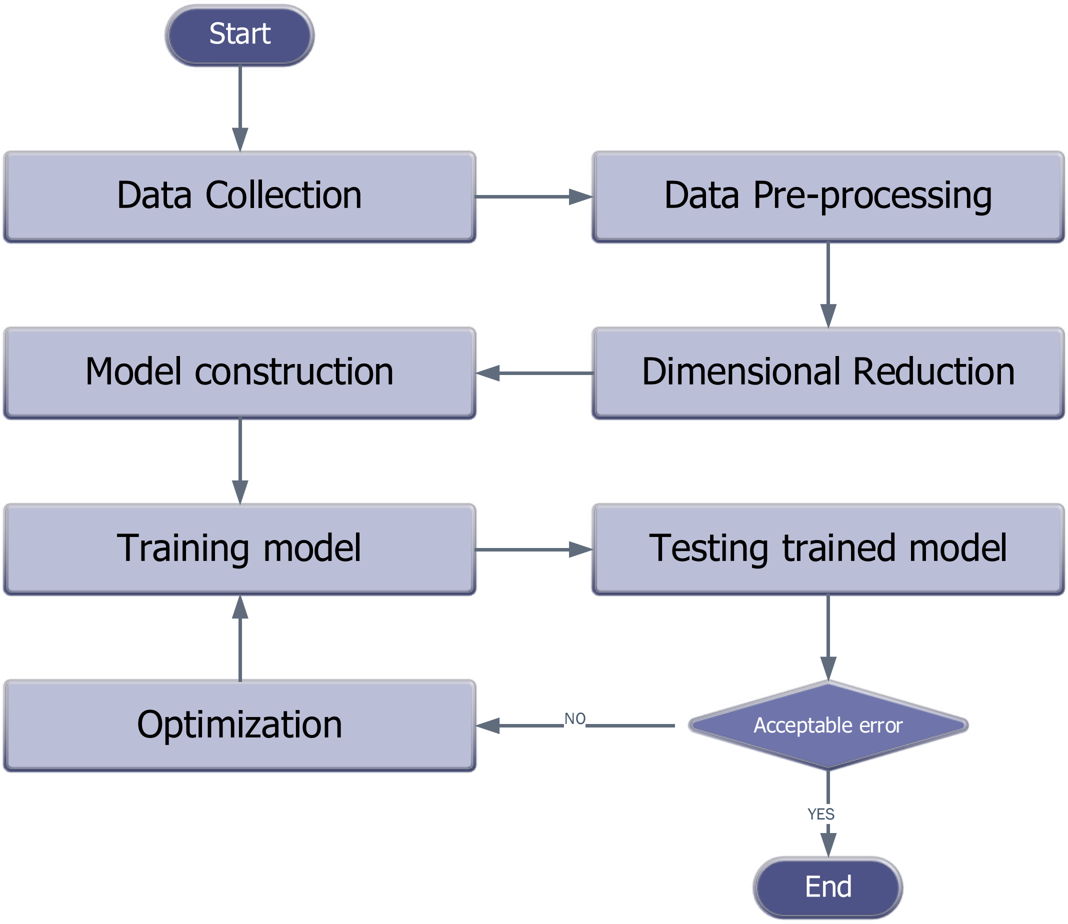

Fig. 1 Flow Chart

Specifically, for data dimensionality reduction, this project has chosen two current mainstream methods: CAE and PCA. In this project, the original data reduced from 112*192 dimensions to 7*6 dimensions. In terms of the construction of the prediction model, the project will mainly refer to the LSTM. In addition, the ConvLSTM mentioned in the literature review has also performed well for wildfire prediction, so this project will also attempt to implement it in subsequent research. For the inputs and outputs, the project attempt to implement a Multi-parallel input and Multi-step output (e.g., taking 5 sets of data as input can predict the next 5 sets).

Overall, the project aims to implement three machine learning models for wildfire probability prediction (CAE+LSTM, PCA+LSTM and ConvLSTM) and compare them to evaluate the most suitable surrogate model.

2.2 Preliminary Implementation and Results

2.2.1 Data pre-processing

The project collected JULES-INFERNO data from 1961 to 1990, with each year being recorded monthly, that a total of 360 samples were included. The original model uses the natural environment to simulate the vegetation dynamics, that each simple provides data on land, weather, fire, and vegetation [6]. At present, the 'burnt_area_gb’ (Grid box mean burnt area fraction) feature was extracted from the data and then converted into 2-dimensional data based on latitude and longitude information. Figure 2 plots the 'burnt_area_gb' for 12 months in 1961, with the horizontal and vertical coordinates indicating latitude and longitude respectively. In order to enable training and testing, the dataset was split in a ratio of 8:2. Apart from that, the data were reshaped according to the requirements of different network models. Finally, to make the training more effective, the dataset was enhanced by setting the offset to obtain more predicted input sequences.

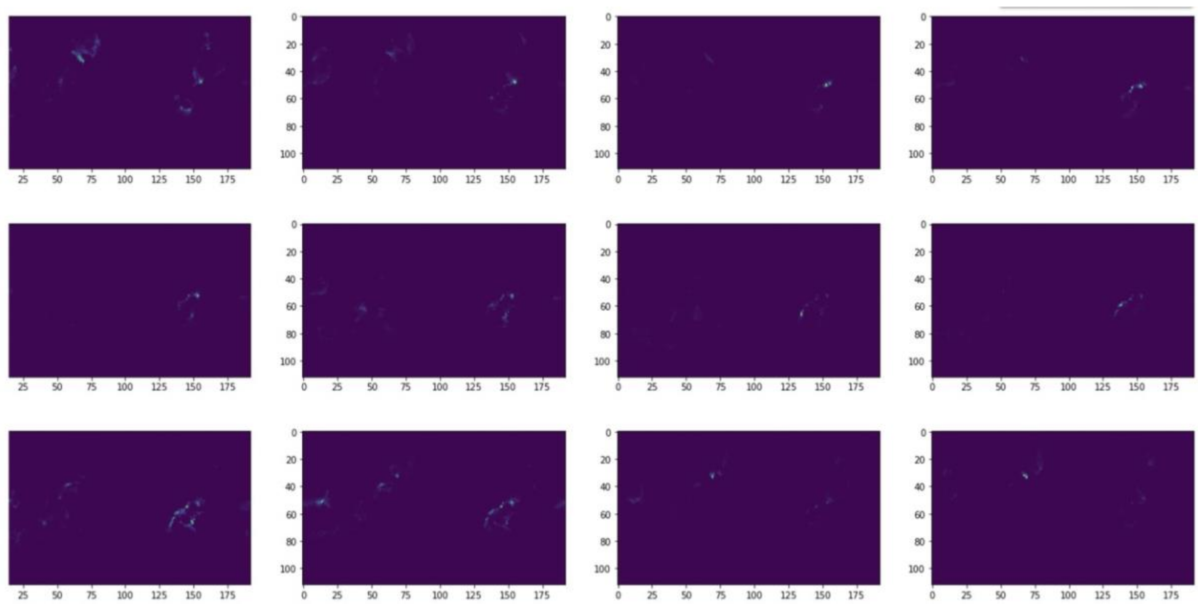

Fig. 2   Data Previewing

2.2.2 PCA + LSTM

The project currently implements a simple PCA combined with LSTM model, which can use 5 set data to predict the next. The Mean Squared Error (MSE) loss obtained by evaluating the trained model using the test set is 6.4071135238304645e-06. A simple comparison visualisation is shown in Figure 3.

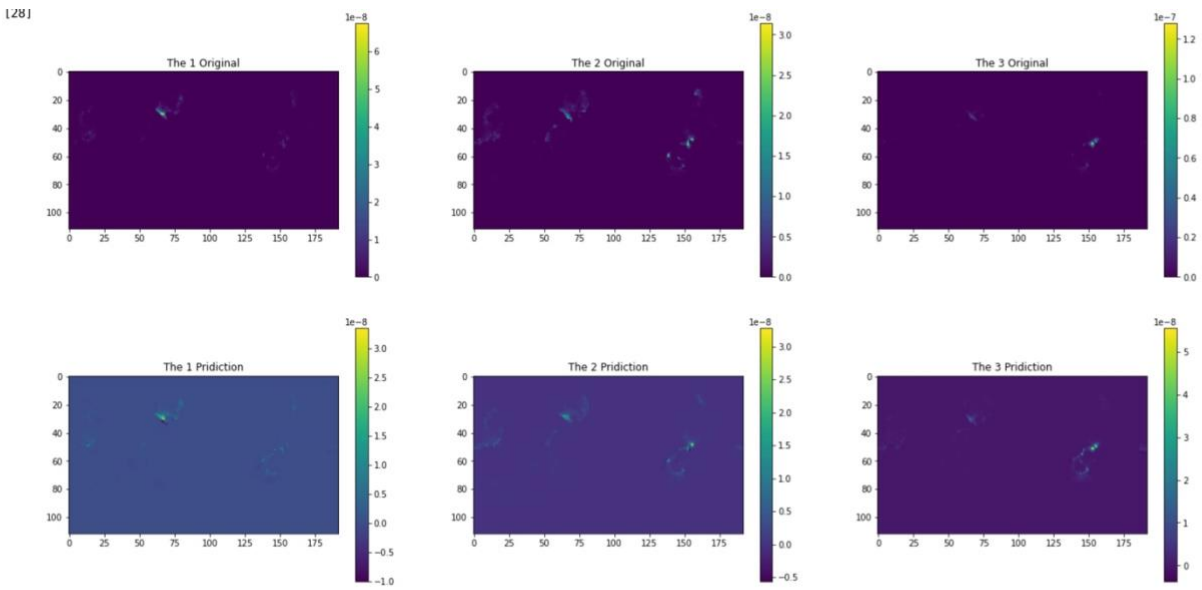

Fig. 3   PCA+LSTM comparison visualization

2.2.3 CAE + LSTM

A simple CAE combined with LSTM model was also constructed. However, the MSE is higher than the last method which is equal to 1.1008991860608025e-05. The forecast performance is not satisfied, and comparison visualisation is shown in Figure 4.

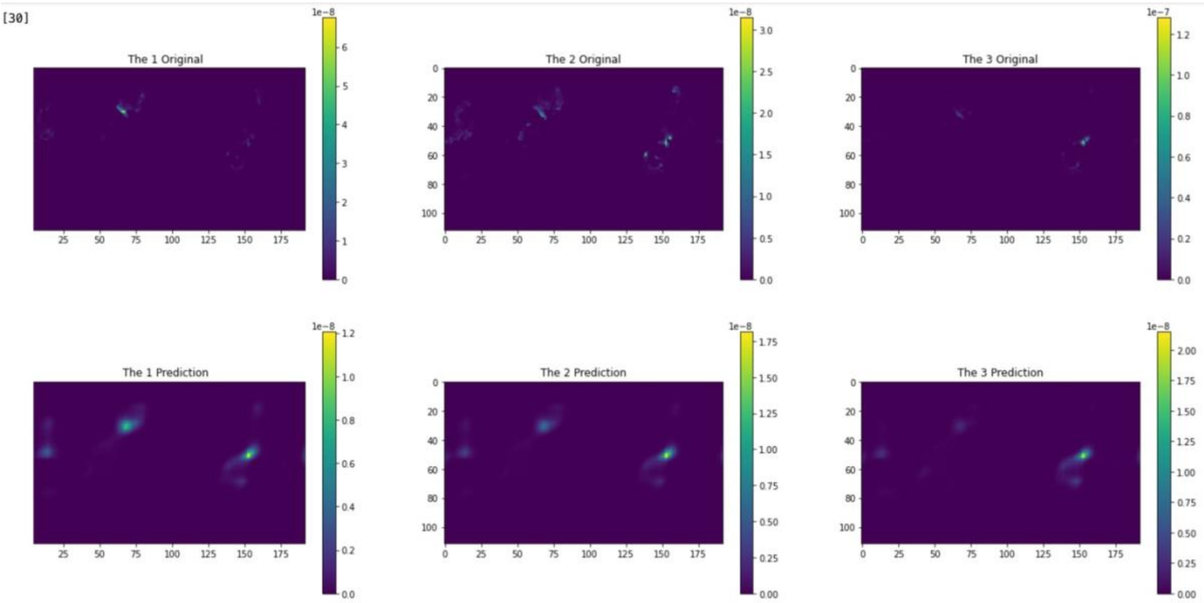

Fig. 4 CAE+LSTM comparison visualizations

3. Future work

The next main objective is to construct the ConvLSTM wildfire prediction model and then optimise these three models to ultimately compare the most suitable surrogate model for wildfire prediction. In addition, both the PCA+LSTM and CAE+LSTM models implemented so far only achieve one-step prediction for multiple inputs, after which multi-step prediction will need to be implemented. The Gantt chart (Figure 5) below shows a preliminary plan.

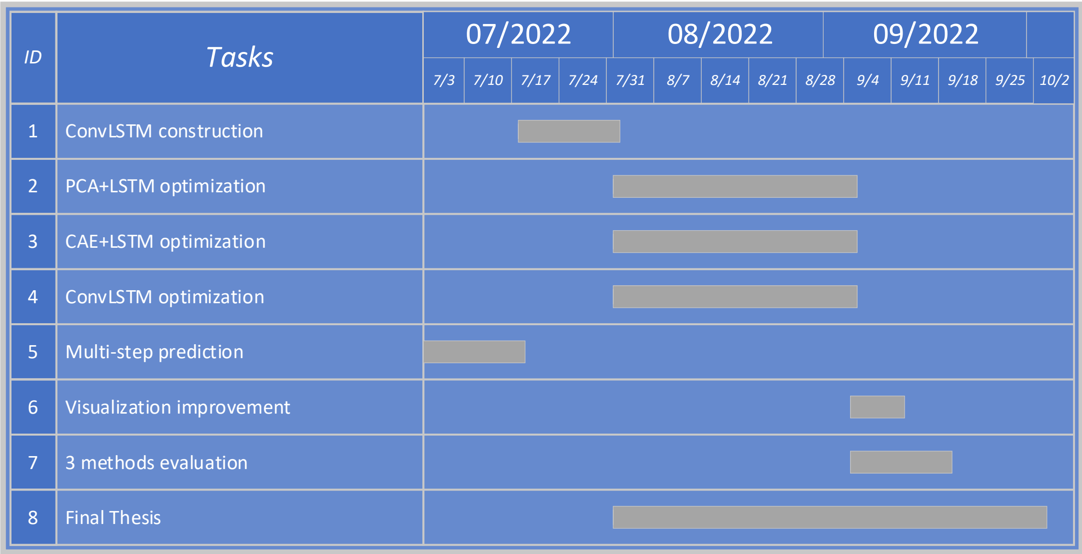

Fig. 5 Gantt Chart

4. Reference

[1] Aldersley, A., Murray, J. S and Cornell, E. S. (2011). 'Global and regional analysis of climate and human drivers of wildfire', *Science of the Total Environment*, 409 (18), pp. 3472-3481. doi: 10.1016/j.scitotenv.2011.05.032.

[2] Pais, C., Miranda, A., Carrasco, J and Shen, Z. M. (2021). 'Deep fire topology: Understanding the role of landscape spatial patterns in wildfire occurrence using artificial intelligence', *Environ Model Softw*, 143, pp. 105-122. Available at: www.sciencedirect.com/science/article/pii/S1364815221001651.

[3] Marques, S., Marto, M., Bushenkov, V and McDill. (2017). 'Addressing wildfire risk in forest management planning with multiple criteria decision making methods', *Sustainability*, 9, pp. 298. Available at: https://www.mdpi.com/181470.

[4] Claussen, M. et al. (2002). 'Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models', *Climate Dynamics*, 18, pp. 579–586. doi: 10.1007/s00382-001-0200-1.

[5] Prentice, I. C. and Cowling, S. A. (2013). 'Dynamic global vegetation models', *In S. A. Levin (Ed.), Encyclopedia of biodiversity*, 2, pp. 670-689. doi: 10.1016/B978-0-12-384719-5.00412-3.

[6] Burton, C. et al. (2019). 'Representation of fire, land-use change and vegetation dynamics in the Joint UK Land Environment Simulator vn4.9 (JULES)', *Geosci. Model Dev.*, 12, pp. 179–193. doi: 10.5194/gmd-12-179-2019.

[7] Mangeon, S., Voulgarakis, A., Gilham, R., Harper, A., Sitch, S., and Folberth, G. (2016). ‘INFERNO: a fire and emissions scheme for the UK Met Office's Unified Model’, *Geosci. Model Dev.*, 9, pp. 2685–2700. doi: 10.5194/gmd-9-2685-2016.

[8] Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V. and House, J. I. (2014). ‘Carbon and Other Biogeochemical Cycles’, *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 465-570. Available at: https://www.ipcc.ch/report/ar5/wg1.

[9] Pechony, O. and Shindell, D.T. (2009). ‘Fire parameterization on a global scale’, *J. Geophys. Res.*, 114. doi: 10.1029/2009JD011927.

[10] Khowaja, K. and Shcherbatyy, M. (2021). ‘Wolfgang Karl Härdle Surrogate Models for Optimization of Dynamical Systems’, *arXiv*. doi: 10.48550/arXiv.2101.10189.

[11] Cortez, P. and Morais, R. A. (2007). ‘A data mining approach to predict forest fires using meteorological data’. Available at: https://repositorium.sdum.uminho.pt/handle/1822/8039.

[12] Zhang, G., Wang, M. and Liu, K. (2019) ‘Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China’, *International Journal of Disaster Risk Science*, 10(3), pp. 1-18. doi: 10.1007/s13753-019-00233-1.

[13] Liang, H., Zhang, M. and Wang, H. (2019). ‘A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors’, *IEEE*. doi: 10.1109/ACCESS.

[14] Cao, Y., Yang, F., Tang, Q. and Lu X. (2019) ‘An Attention Enhanced Bidirectional LSTM for Early Forest Fire Smoke Recognition’, *IEEE*, p. 1. doi: 10.1109.