Climate Change Prediction and its Consequences on Major Crops with Economic Loss

Loc Nguyen

Degree of Data Science
SP Jain School of Global Management, Lidcombe, NSW 2141, Australia

Supervisor: JP Aggarwal
Submitted Date: April 23, 2024

Abstract

The Climate Change Impact Prediction project is an innovative initiative that leverages the power of machine learning to forecast the impact of climate change in specific locations. By deploying a random forest model, the project aims to predict temperature changes and their associated impacts on four major crops—maize, rice, soybean, and wheat—along with the resulting economic consequences. This user-friendly website-based tool serves as a bridge to raise awareness about climate change, offering valuable insights and recommendations to help local communities adapt to the ongoing environmental shifts.

Keywords: Climate Change, Machine Learning, Crop Production, Economic.

Introduction

Crop productivity is often the initial sector to suffer from the direct and indirect effects of climate change on various food system processes (Nguyen et al., 2024). Studies on crop production projections under future climate change date back to the early 1980s. Researchers have been able to project the effects of climate change on crop yields under different scenarios since the 1990s by using crop simulation models and data on future climate. Since then, hundreds of studies have simulated crop yields under various growing conditions and climate scenarios using crop simulation models. National and international organisations, particularly the Intergovernmental Panel on Climate Change (IPCC) Working Group II, which offers scientific evidence pertinent to policy regarding the effects of and adaptation to climate change, have reviewed and evaluated the results on a regular basis (Aggarwal et al., 2019). While the overall effects are negative, there are significant regional variations, according to review studies spanning the last five IPCC assessment cycles.

Prior to 2010, the majority of simulation research was carried out by lone research groups with various climate models, target years, spatial resolution with local management, and cultivar circumstances. However, since 2010, a great deal of work has gone into coordinating modelling studies through the Agricultural Model Intercomparison and Improvement Project (AgMIP). AgMIP compares outputs from various crop models using inputs that are standardised. The significance of machine learning based on multiple crop models has been bolstered by early AgMIP activities that have separated sources of uncertainty in crop yield projections and shown that yield projections vary among crop models and that model ensemble mean or median often works better than a single model. Thus, in this project, we aimed to:

- Predict Temperature Changes: Utilizing the random forest model, the project predicts temperature fluctuations in various regions based on user inputs, such as local temperature and location.
- Assess Impact on Crops: The project examines how these temperature changes affect major crops like maize, rice, soybean, and wheat, offering detailed analysis of potential yield reductions or other crop-specific impacts.
- Calculate Economic Loss: Beyond crop impacts, the project estimates the economic loss that may result from the changing climate, providing an essential context for understanding broader societal impacts.
- Provide Adaptation Recommendations: The project provides tailored recommendations to help local communities adapt to climate change, such as altering farming practices, implementing water management strategies, or diversifying crops.
- Raise Awareness and Influence Policy: By making this tool available to the public, the project aims to raise awareness about climate change, encouraging local government agencies to create policies that address climate-related issues affecting their communities.

Methodology

Data

In this study, we used a global database that was acquired using two different methods, potentially useful for the IPCC Working Group II assessment (Hasegawa et al., 2022). The dataset can be used as a reliable foundation for analyses starting with the sixth IPCC assessment since it includes data from all six cycles of the assessment.

The dataset includes the most pertinent variables to assess how climate change will affect 21st-century maize, rice, soybean, and rice yields (Figure 1). These include the following: geographic coordinates, crop species, CO2 emission scenarios, CO2 concentrations, current temperature and precipitation levels, degrees of local and global warming, anticipated changes in precipitation, and the relative changes in yield as a percentage of the baseline period obtained with or without CO2 effects.

```python
df.head()
```

	ID	Ref No	Future_Mid-point	Baseline_Mid-point	Time slice	Climate scenario	Scenario source	Method	Scale	Crop	...	CO2 ppm	Impact relative to 2005_NAD_%	Impact relative to 2005_NAD_yield	Impact relative to 2005_AD_%	Impact relative to 2005_AD_yield
0	2	1	2054	1995	MC	RCP8.5	CMIP5	SCOPUS-NewSearch	Regional	Maize	...	564.31311	-16.277966	5.603880	11.627119	7.945800
1	4	2	2055	1998	MC	RCP8.5	CMIP5	SCOPUS-NewSearch	Regional	Maize	...	570.51669	-19.649123	8.464073	21.929825	13.634139
2	14	5	2070	2002	EC	RCP4.5	CMIP5	SCOPUS-NewSearch	Regional	Maize	...	524.30217	-25.911390	4.067413	-15.957978	4.648447
3	15	5	2070	2002	EC	RCP4.5	CMIP5	SCOPUS-NewSearch	Regional	Maize	...	524.30217	-25.911390	4.067413	32.388735	7.470705
4	16	5	2070	2002	EC	RCP4.5	CMIP5	SCOPUS-NewSearch	Regional	Maize	...	524.30217	-25.911390	4.067413	-21.225272	4.340967

Figure 1 Dataset

The dataset is described in Figure 2.

```python
df.describe()
```

	ID	Future_Mid-point	Baseline_Mid-point	latitude	longitude	Current Average Temperature (dC)_area_weighted	Current Average Temperature_point_coordinate (dC)	Current Annual Precipitation (mm) _area_weighted	Current Annual Precipitation (mm) _point_coordinate	Local delt from 20
count	2005.000000	2005.000000	2005.000000	2004.000000	2004.000000	2004.000000	2004.000000	2004.000000	2004.000000	2004.000
mean	4801.131671	2056.022444	1994.990025	26.506349	25.973802	18.220760	17.819436	927.068110	897.250841	1.820
std	2722.004729	20.313069	12.093692	20.971099	72.243844	6.708541	7.753867	545.996800	644.093368	1.050
min	2.000000	2020.000000	1975.000000	-38.416100	-107.800000	1.489925	-8.371055	0.967242	0.967242	0.164
25%	1110.000000	2050.000000	1990.000000	19.856300	-3.749220	12.548148	12.548148	570.196723	421.181280	1.002
50%	5401.000000	2050.000000	2000.000000	32.760000	35.243300	16.520588	16.520588	798.284578	762.041285	1.550
75%	7009.000000	2080.000000	2005.000000	38.963700	87.200000	24.854590	24.980152	1247.933810	1247.933810	2.31
max	8685.000000	2100.000000	2012.000000	61.524000	138.253000	29.727396	29.727396	3076.049282	3891.503190	5.550

8 rows × 21 columns

Figure 2 Dataset description

We then chose a subset of studies for climate-scenario-based simulations that contained the following terms related to climate scenarios: "RCP," "RCP2.6," "RCP6.0," "RCP4.5," "RCP8.5," "CMIP5," and "CMIP6." The term "Representative Concentration Pathways" (RCP) refers to a series of greenhouse gas concentration trajectories that describe various future levels of greenhouse gas emissions (van Vuuren et al., 2011). The value that follows RCP represents the radiative forcing level (Wm^{-2}) attained by the end of the twenty-first century. This number rises in proportion to the amount of greenhouse gases released into the atmosphere. Coupled Model Intercomparison Project (CMIP) Phases 5 and 6 involve groups of various earth system models (ESMs) providing global-scale climate projections based on various RCPs.

The yields of the baseline period, projected yields with and without adaptations, geographic coordinates, crop species, greenhouse gas emission scenarios, and options for adaptation were all extracted. Additionally, we made every effort to measure variations in CO2 concentration and local and global temperatures. Apart from gathering information from published works, data also got in touch with a few grid simulation study authors to get combined findings for specific nations or areas. In response, the authors of the three grid simulation studies gave data on annual temperature, precipitation, and baseline and projected yields that were combined for various nations or regions (Figure 3). Next, an average of the outcomes from each ESM was calculated.

The effect of climate change on yield (YI) is calculated using simulated grain mass per unit land area. Yield is defined as:

$$YI(\%) = (Y_f / Y_b - 1) * 100$$

where the baseline yield is denoted by Y_b and the future yield by Y_f. In one study20, yields were separately simulated from the pre-industrial era to the end of the 21st century under both climate change and counterfactual non-climate change scenarios. Yield increases resulting from non-climatic technological factors were also taken into account. Because it incorporates the effects of both technological factors and climate change, the YI found using the above equation under the climate change scenario was not entirely relevant in this instance. As a result, YI for this study was calculated using the average yield under climate change for the years 2001–2010.

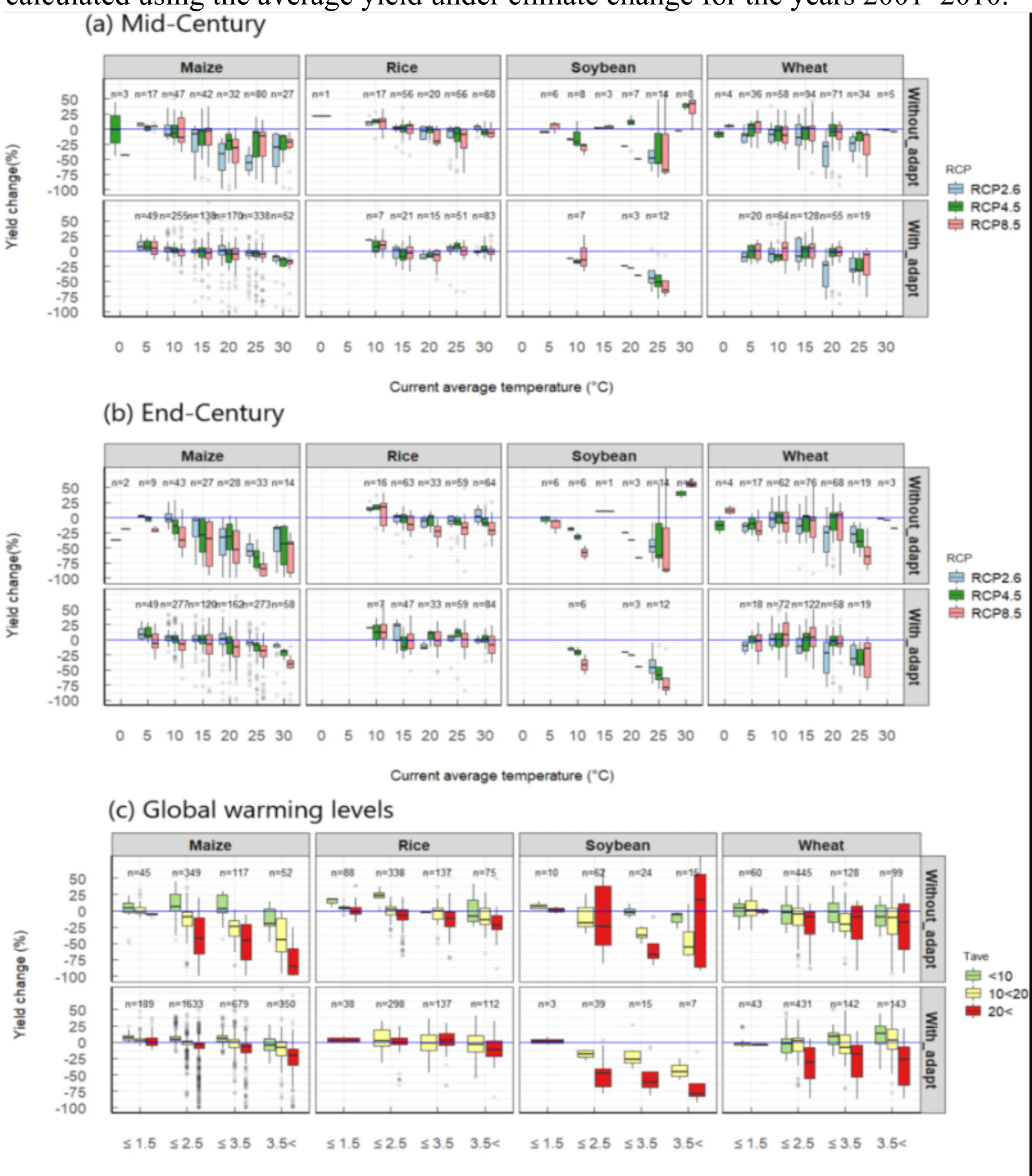

Figure 3 Changes in expected yield with respect to the baseline period (2001–2010). (a) End-century (EC) projections under three RCP scenarios by current annual temperature (Tave), (b)

Mid-century (MC) projections under RCP8.5 scenario without adaptation, with upper panels indicating positive impacts and lower panels negative impacts, and (c) Yield change as a function of global temperature rise from the pre-industrial period by three Tave levels. The median is shown by the middle line in the box, which represents the interquartile range (IQR).

Adaptation to Climate Changes

Various management or cultivar options are included in the dataset (Figure 5). The options for adaptation are divided into categories such as cultivar, irrigation, tillage, planting time, soil organic matter management, and fertiliser. More specifically, we consider it an adaptation in the fertiliser option if the timing and quantity of fertiliser application differ from the current conventional method. When it comes to the irrigation option, we classify it as adaptation since the management is modified to account for future climatic conditions if the simulation programme decides how to schedule irrigation based on crop growth, climate, and soil moisture conditions. We do not view irrigation as an adaptation if rainfed and irrigated conditions are simulated independently. The use of cultivars with a higher heat tolerance or from different maturity groups than traditional cultivars is referred to as cultivar option. The planting time option relates to a change in planting time from the standard schedule. When testing different planting times, we choose the one that produces the highest yield. The option for managing soil organic matter involves applying crop residue and/or compost. In contrast to no conventional tillage, the tillage option relates to reduced- or no-till cultivation. It calculates YI from the ratio of yield with adaptation under climate change to baseline yield without adaptation when studies take adaptation options into account (Figure 4).

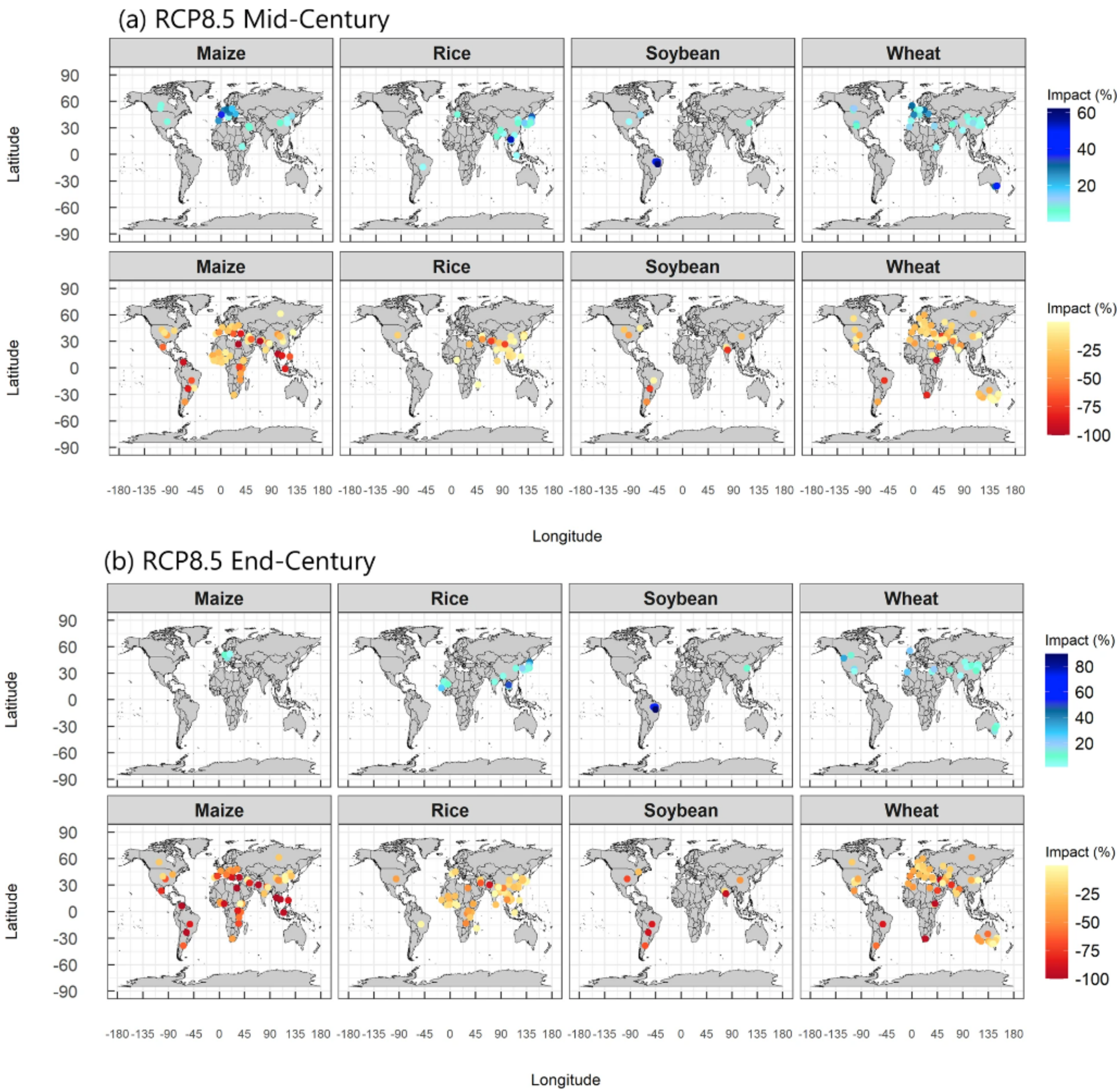

Figure 4 Impacts of climate change (percentage of yield change from baseline period) on four crops under RCP8.5 without adaptation. A mid-century; a. End-Century; b. Maps with bluish symbols indicate gains in yield, while those with reddish symbols indicate losses in yield.

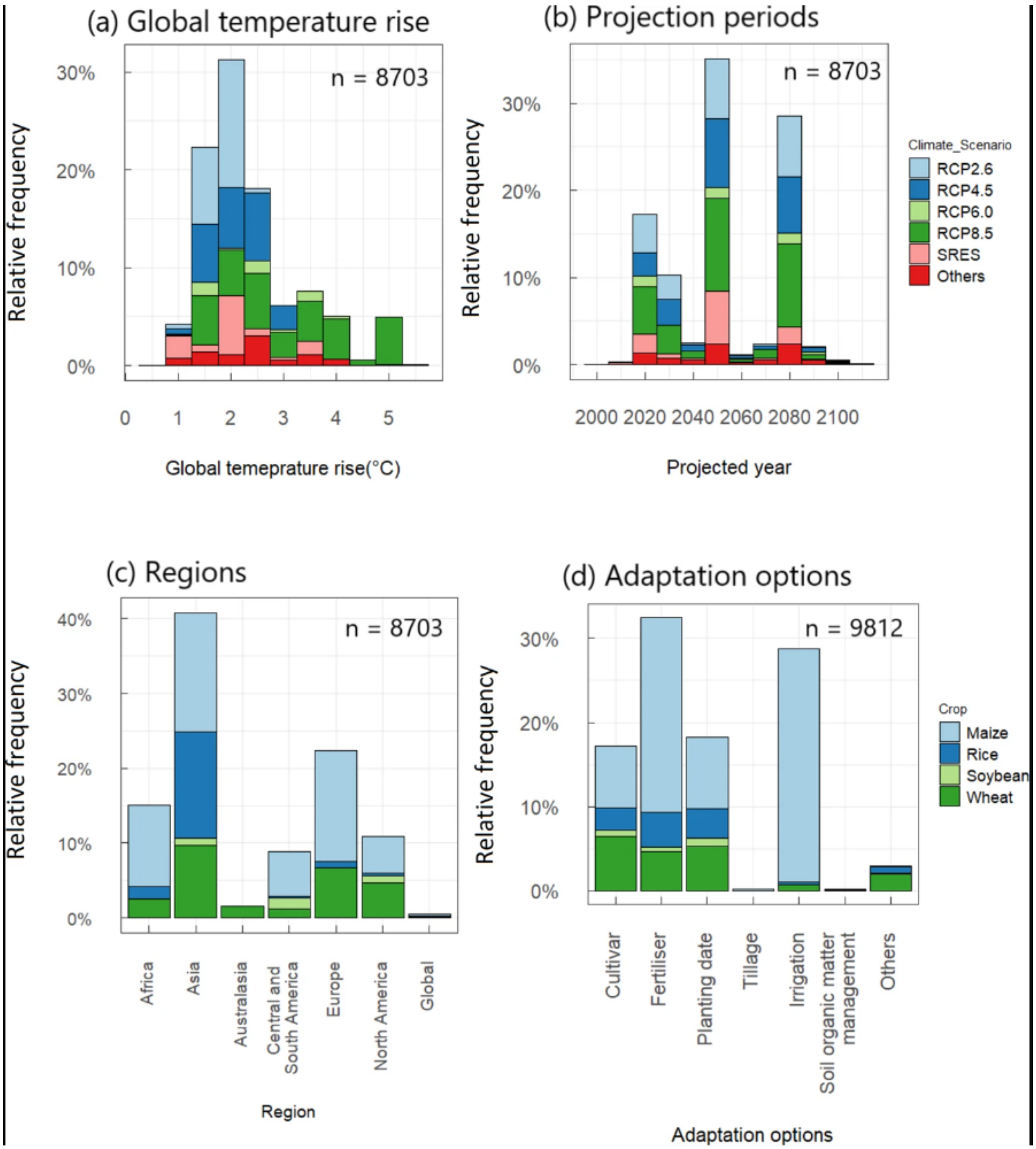

Figure 5 Crop yield simulation data availability and breakdown. The following factors influence the global temperature rise: (a) the rise from the preindustrial era and climate scenarios; (b) projected time periods (midpoint years) and climate scenarios; (c) crop species and IPCC regions29; and (d) crop species and adaptation options. Because we add up all of the options used in the simulations, even the ones that use multiple options, you will notice that n = 9812 in adaptation options (d) exceeds the total number of simulations (8703). The number of simulation outcomes is n.

Temperature and Precipitation Changes

There are significant ramifications for both the rise in local temperature (ΔTl) and the rise in global mean temperature (ΔTg) from the baseline period. The former has a direct impact on crop yield and growth, while the latter is a global goal related to mitigation efforts. We made every effort to extract both ΔTl and ΔTg from the literature; however, not many studies have ΔTg available. In these situations, we used the Warming Attribution Calculator to estimate ΔTg (http://wlcalc.climateanalytics.org/choices).

Baseline

Since the baseline periods for the studies varied, we used a linear interpolation method, as described by (Aggarwal et al., 2019), to correct YI, ΔTl, ΔTg, and ΔPr to the 2001–2010 baseline period. To be more precise, the impacts YI were first divided by the difference in years between the baseline period midpoint year of the initial study and the midpoint year of the future period. The year difference from our reference baseline period midpoint year (2005) was then multiplied by the impact per year. For 2001–2010, the same technique was used to express ΔTl and ΔPr.

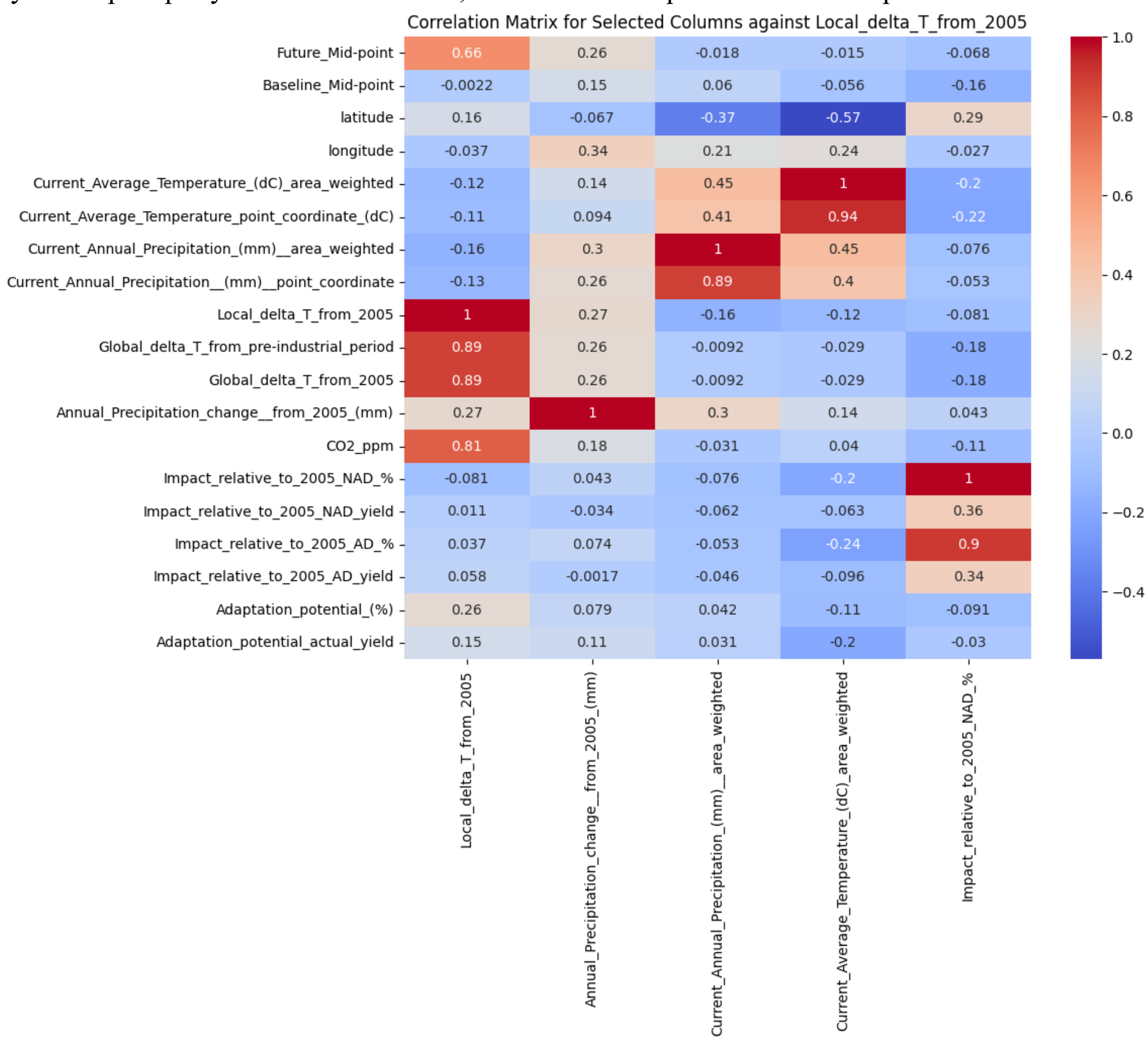

Figure 6 Correlation Matrix

Model Construction

In this study, we employed a Random Forest Regression to predict local temperature rise (Figure 7). Random forest regression is a machine learning technique used to predict continuous values. It

is an extension of the random forest algorithm, which is widely known for its robustness and accuracy.
Random forest regression builds upon the idea of decision trees, but it uses an ensemble of multiple trees instead of a single one. Each tree makes a prediction, and the final prediction is the average of the individual predictions. Like a regular decision tree, each tree in the forest is constructed by splitting the data based on certain features to minimize an error metric (like mean squared error). Random forests use "bootstrap sampling," where each tree is trained on a randomly selected subset of the training data with replacement. To encourage diversity among the trees, random forests use "feature bagging." At each node split, the algorithm considers a random subset of features instead of all features. This helps reduce overfitting. The final prediction is obtained by averaging the predictions from all trees in the forest. This aggregation helps smooth out individual tree biases and reduces overfitting, resulting in more robust and accurate predictions.
Due to ensemble averaging, random forests are generally accurate. They're less prone to overfitting compared to single decision trees. Random forests can capture complex relationships without assuming a linear relationship between features and the target variable. They provide insights into the importance of different features in predicting the outcome.
However, it also has some limitations. Due to the ensemble nature, random forest regression models are less interpretable than single decision trees. Building multiple trees requires more computational resources than a single tree.

```
# Random Forest Modeling
X_train = DATA_train.drop('Local_delta_T_from_2005', axis=1)
y_train = DATA_train['Local_delta_T_from_2005']

X_train2 = DATA_train2.drop('Local_delta_T_from_2005', axis=1)
y_train2 = DATA_train2['Local_delta_T_from_2005']

X_train3 = DATA_train3.drop('Local_delta_T_from_2005', axis=1)
y_train3 = DATA_train3['Local_delta_T_from_2005']

X_train4 = DATA_train4.drop('Local_delta_T_from_2005', axis=1)
y_train4 = DATA_train4['Local_delta_T_from_2005']
```

```
"""
Models with mode columns
"""
# Create Random Forest Models
mod_rf = RandomForestRegressor(n_estimators=500, random_state=42)
mod_rf.fit(X_train, y_train)

mod_rf2 = RandomForestRegressor(n_estimators=500, random_state=42)
mod_rf2.fit(X_train2, y_train2)

mod_rf3 = RandomForestRegressor(n_estimators=500, random_state=42)
mod_rf3.fit(X_train3, y_train3)

mod_rf4 = RandomForestRegressor(n_estimators=500, random_state=42)
mod_rf4.fit(X_train4, y_train4)
```

```
RandomForestRegressor
RandomForestRegressor(n_estimators=500, random_state=42)
```

Figure 7 Fitting Random Forest Regressor to Data

The features inputs in the model are 'Local_delta_T_from_2005', 'Global_delta_T_from_2005', 'latitude', 'longitude', 'Climate_scenario', 'Future_Mid-point', 'Current_Average_Temperature_(dC)_area_weighted', 'Current_Average_Temperature_(dC)_area_weighted', 'Country', 'Time_slice', 'Fertiliser', 'Irrigation', 'Cultivar', 'Adaptation_type' (Figure 8).

```
print("Features in DATA_train:", DATA_train.columns)

# Print the names of the features used to fit the model
# (This code is not provided in the original user input)
# print("Features used for model fitting:", mod_rf.feature_names_in_)
```

```
Features in DATA_train: Index(['Local_delta_T_from_2005', 'Global_delta_T_from_2005', 'latitude',
       'longitude', 'Climate_scenario', 'Future_Mid-point',
       'Current_Average_Temperature_(dC)_area_weighted',
       'Current_Average_Temperature_(dC)_area_weighted', 'Country',
       'Time_slice', 'Fertiliser', 'Irrigation', 'Cultivar',
       'Adaptation_type'],
      dtype='object')
```

Figure 8 Feature Input

Output:
Cross-validation Mean Square Errors (MSE) are reported in Figure 9.

```
Cross-validation MSE for mod_rf: 0.004607168318592218
Cross-validation MSE for mod_rf2: 0.004607168318592218
Cross-validation MSE for mod_rf3: 0.004607168318592218
Cross-validation MSE for mod_rf4: 0.004607168318592218
Feature Importance for mod_rf: [8.19789061e-01 3.81059637e-02 3.06389982e-02 2.26831862e-02
 9.71946684e-03 1.60229071e-02 1.52548080e-02 8.82099991e-03
 2.08809678e-02 1.19413582e-03 2.97924082e-04 1.35416821e-03
 1.52374131e-02]
Feature Importance for mod_rf2: [8.19789061e-01 3.81059637e-02 3.06389982e-02 2.26831862e-02
 9.71946684e-03 1.60229071e-02 1.52548080e-02 8.82099991e-03
 2.08809678e-02 1.19413582e-03 2.97924082e-04 1.35416821e-03
 1.52374131e-02]
Feature Importance for mod_rf3: [8.19789061e-01 3.81059637e-02 3.06389982e-02 2.26831862e-02
 9.71946684e-03 1.60229071e-02 1.52548080e-02 8.82099991e-03
 2.08809678e-02 1.19413582e-03 2.97924082e-04 1.35416821e-03
 1.52374131e-02]
Feature Importance for mod_rf4: [8.19789061e-01 3.81059637e-02 3.06389982e-02 2.26831862e-02
 9.71946684e-03 1.60229071e-02 1.52548080e-02 8.82099991e-03
 2.08809678e-02 1.19413582e-03 2.97924082e-04 1.35416821e-03
 1.52374131e-02]
```

Figure 9 Mean Square Error and Feature Importance

As showed in the result section, the MSEs are relatively small, implying the model is robust and well-validated.

Website Development

The project is designed to be user-friendly and accessible to a wide audience (Figure 10).

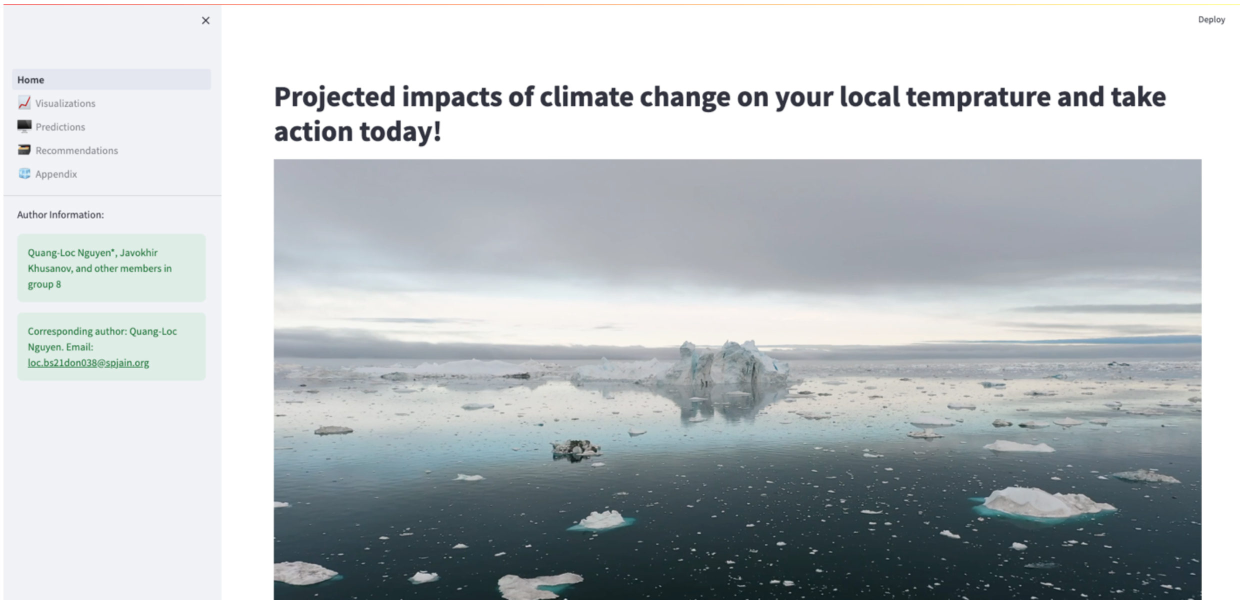

Figure 10 User-friendly web page in the home screen

The website also includes visualizations to help users understand the potential impacts on their community (Figure 11). Graphs and charts illustrate projected temperature increases over time, changes in crop yields, and estimated economic losses. This visual approach facilitates comprehension, allowing users to grasp the significance of climate change in their local context.

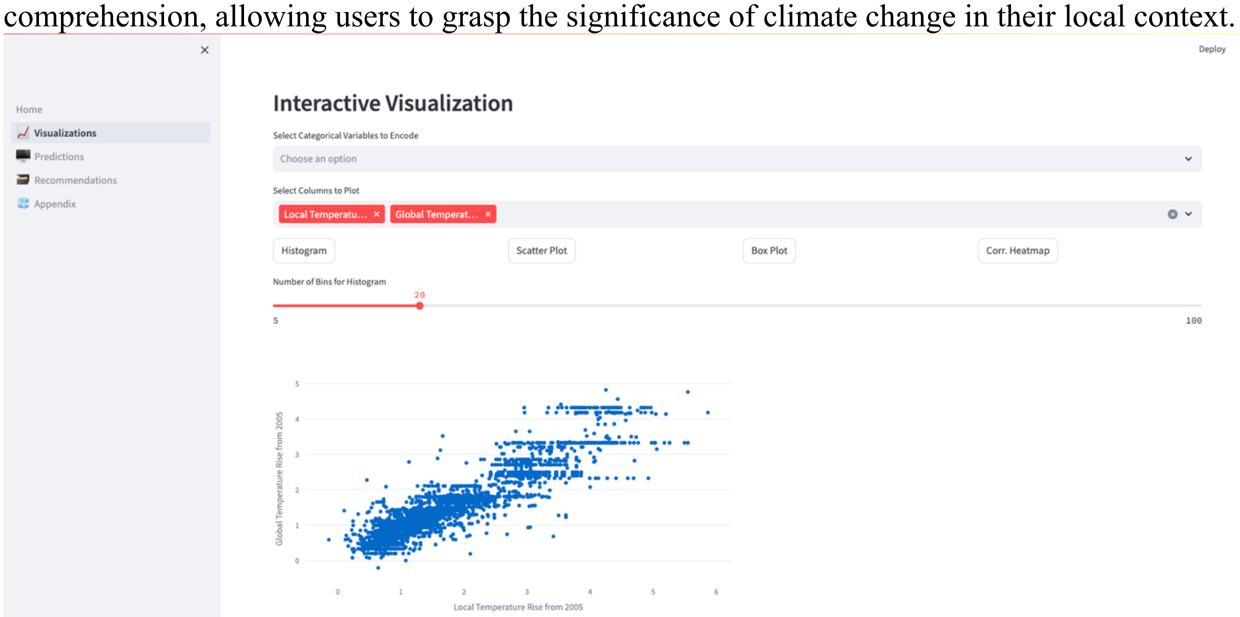

Figure 11 Interactive Visualization

Prediction

Users can input their local temperature and location data into the website, and the random forest model generates predictions for climate change in their area Figure 12. The model considers various factors such as historical temperature trends, geographical characteristics, and agricultural data to provide accurate predictions.

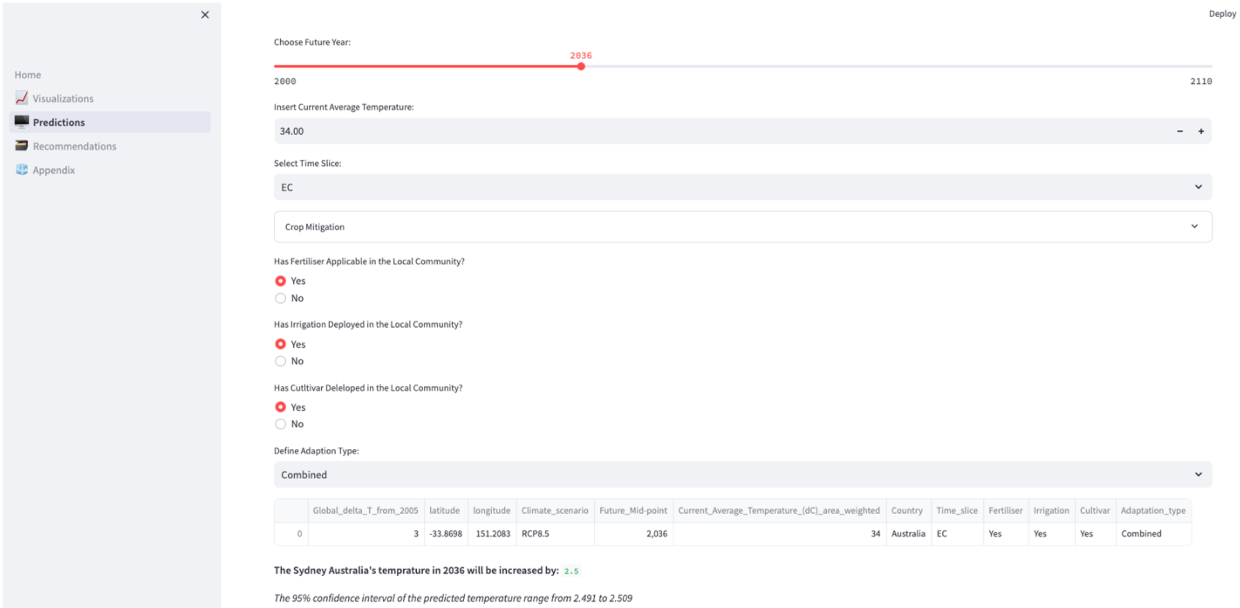

Figure 12 An example of model prediction

Furthermore, the website will help measure how the four crops production change in the expected year (Figure 13).

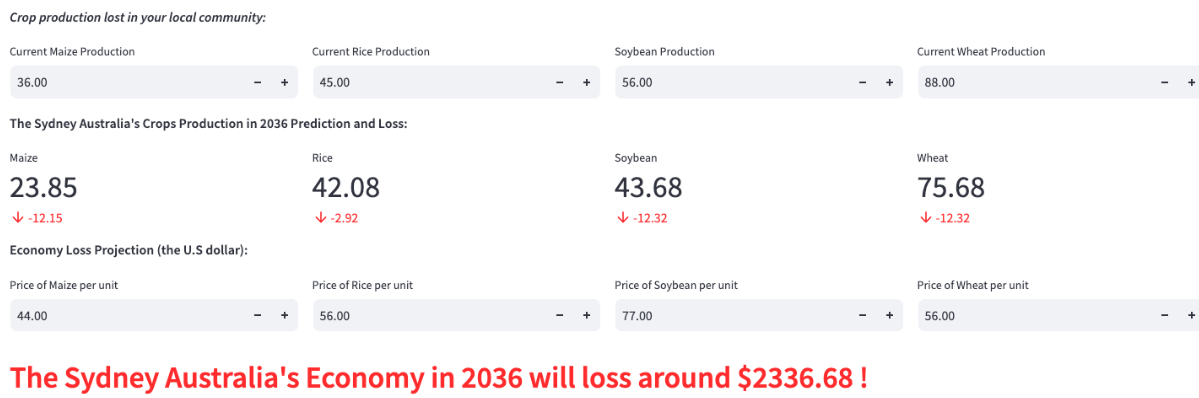

Figure 13 Example of crops production projection

Recommendations and Adaptation Strategies

The project goes beyond mere prediction, offering practical recommendations for adapting to climate change (Figure 14). These recommendations are tailored to the user's specific location and crop data, providing personalized guidance on topics such as:

- Climate-Resilient Crops: Suggestions for crops that are more resistant to temperature changes and adverse weather conditions.
- Water Management: Strategies for efficient water usage, including irrigation techniques and rainwater harvesting.
- Crop Rotation and Diversification: Recommendations for diversifying crops to reduce risk and promote sustainable farming practices.

Deploy

Recommendations for Adaptation and Mitigation

Adapting Agriculture in the Changing World

Strategies to Adapt and Mitigate

Adapting agriculture to ensure crop production in the face of climate change requires a combination of strategies that enhance resilience, conserve resources, and optimize productivity. Here are some key strategies:

Crop Diversification: Planting a variety of crops helps spread risk. Different crops have different tolerances to temperature, precipitation, and pests, so diversification can help buffer against the impacts of climate variability.

Breeding Resilient Varieties: Developing and using crop varieties that are more resilient to climate stressors such as drought, heat, floods, and pests. This involves traditional breeding techniques as well as emerging technologies like genetic engineering.

Water Management: Implementing efficient irrigation systems, water conservation practices, and water harvesting techniques to ensure crops have adequate water during changing climate conditions, including both droughts and floods.

Soil Health Management: Practices such as conservation tillage, cover cropping, and adding organic matter to soil help improve soil structure, fertility, and water retention, making it more resilient to extreme weather events.

Agroforestry and Agroecology: Integrating trees and shrubs into agricultural landscapes (agroforestry) and adopting agroecological practices help increase biodiversity, improve soil health, enhance water retention, and provide natural pest control.

Figure 14 Recommendation Page

Conclusion

The project's ultimate goal is to drive positive change at the local community level. By providing accurate predictions and actionable insights, the project seeks to empower individuals and local governments to make informed decisions about climate change (M.-H. Nguyen et al., 2023; Vuong et al., 2023). For example, local government agencies can use this information to develop policies that support sustainable agriculture, mitigate economic losses, and promote climate resilience.

Furthermore, the project fosters awareness and encourages community engagement. It provides a platform for individuals to learn about the effects of climate change in their area, fostering a sense of responsibility and encouraging proactive measures. This engagement has the potential to create a ripple effect, where informed communities advocate for broader climate action and contribute to global efforts to combat climate change (M.-H. Nguyen et al., 2022; Q. L. Nguyen et al., 2023).

Overall, the Climate Change Impact Prediction project represents a significant step towards understanding and addressing the impacts of climate change. By combining machine learning with a user-centric approach, the project provides a valuable resource for local communities, promoting awareness, adaptation, and informed policy-making.

Reference:

Aggarwal, P., Vyas, S., Thornton, P., & Campbell, B. M. (2019). How much does climate change add to the challenge of feeding the planet this century? *Environmental Research Letters*, *14*(4), 043001. https://doi.org/10.1088/1748-9326/aafa3e

Hasegawa, T., Wakatsuki, H., Ju, H., Vyas, S., Nelson, G. C., Farrell, A., Deryng, D., Meza, F., & Makowski, D. (2022). A global dataset for the projected impacts of climate change on four major crops. *Scientific Data*, *9*(1), 58. https://doi.org/10.1038/s41597-022-01150-7

Nguyen, M.-H., Duong, M.-P. T., Nguyen, Q.-L., La, V.-P., & Hoang, V.-Q. (2024). In search of value: The intricate impacts of benefit perception, knowledge, and emotion about climate change on marine protection support. *Journal of Environmental Studies and Sciences*. https://doi.org/10.1007/s13412-024-00902-8

Nguyen, M.-H., Khuc, Q. V., La, V.-P., Le, T.-T., Nguyen, Q.-L., Jin, R., Nguyen, P.-T., & Vuong, Q.-H. (2022). Mindsponge-Based Reasoning of Households' Financial Resilience during the COVID-19 Crisis. *Journal of Risk and Financial Management*, *15*(11), Article 11. https://doi.org/10.3390/jrfm15110542

Nguyen, M.-H., Nguyen, M.-H. T., Jin, R., Nguyen, Q.-L., La, V.-P., Le, T.-T., & Vuong, Q.-H. (2023). Preventing the Separation of Urban Humans from Nature: The Impact of Pet and Plant Diversity on Biodiversity Loss Belief. *Urban Science*, *7*(2), Article 2. https://doi.org/10.3390/urbansci7020046

Nguyen, Q. L., Nguyen, M.-H., La, V.-P., Bhatti, I., & Vuong, Q. H. (2023). *Enterprise's Strategies to Improve Financial Capital under Climate Change Scenario – Evidence of the Leading Country* (SSRN Scholarly Paper 4351342). https://doi.org/10.2139/ssrn.4351342

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1), 5. https://doi.org/10.1007/s10584-011-0148-z

Vuong, Q.-H., Nguyen, Q.-L., Jin, R., Nguyen, M.-H. T., Nguyen, T.-P., La, V.-P., & Nguyen, M.-H. (2023). Increasing Supply for Woody-Biomass-Based Energy through Wasted Resources: Insights from US Private Landowners. *Sustainability*, *15*(11), Article 11. https://doi.org/10.3390/su15118667