

8-6-2021

## **Robust Data Submission Pipeline For AgMIP Global Economics Modelers**

Raziq Ramli  
*Purdue University*, mramli@purdue.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/duri>



Part of the [Computational Engineering Commons](#)

---

### **Recommended Citation**

Raziq Ramli (2021), "AgMIP GlobalEcon Data Submission Tool," <https://mygeohub.org/resources/agmipsub>.

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.  
Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

---

# Robust Data Submission Pipeline for AgMIP Global Economics Modelers

Raziq R. Ramli  
Purdue College of Science  
[mramli@purdue.edu](mailto:mramli@purdue.edu)

---

# What is AgMIP?

**Agricultural Model Inter-comparison and Improvement Project**

**International effort to improve our capabilities in assessing the impacts of climate change and other driving factors on:**

- **Agriculture**
- **Food security**
- **Poverty**



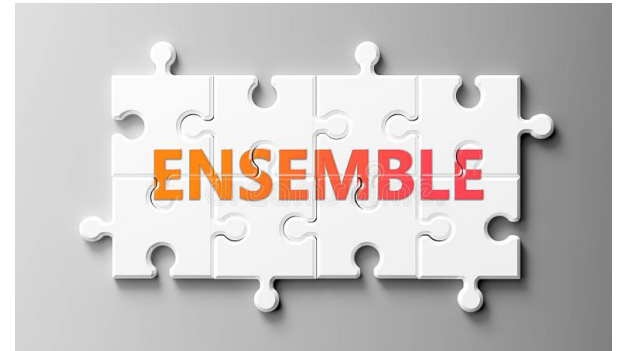
# AgMIP Global Economics Team

**What do they do?**

- **Ensemble modeling activities**
- **Systematic model inter-comparisons**

**What's their topics of interests?**

- **Ambitious climate change mitigations**
- **Land use and dietary change**



# Ensemble Modeling Activity

**How is it being done?**

- **Diverse modeling groups**
- **Share interoperable model data**
- **Manual data cleaning & harmonization**

**Can we improve it?**



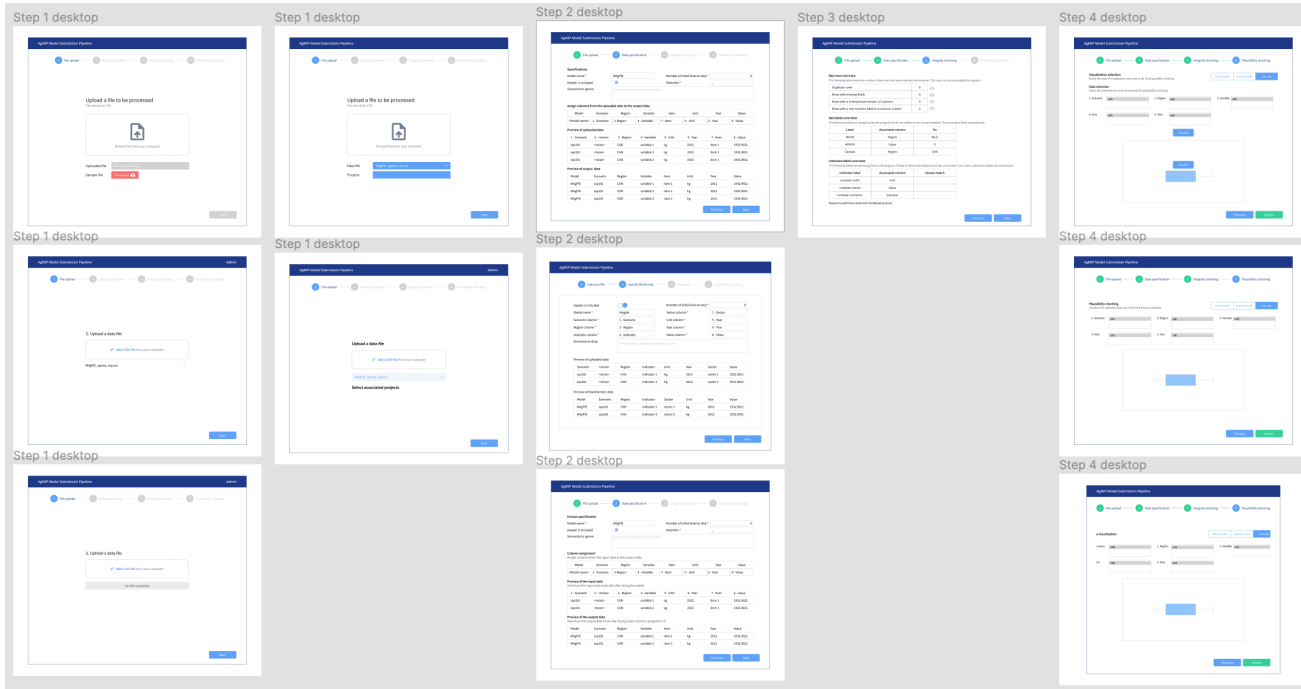
**“How can we design a robust data submission pipeline that can streamline the ensemble modeling activities conducted by the AgMIP Global Economics team?”**

# Approach

**What features should we provide?**

**Create mockups and prototypes and iterate**





Some UI mockups that we have created during this process



# Result

## User stories?

- Private data submission
- Data validation & cleaning
- Error diagnosis
- Easy protocol update
- Outlier visualization

## End-product?



# Web Application

**Jupyter notebook & MyGeoHub**

**Ease of development and deployment**

**Hosts existing data exploration and analysis tool**



# Application Overview

**Accept & transforms model data**

**Table (model name, scenario, region, variable, item, unit, year, value)**

**Harmonizes model data**



# Application Overview

**How do we represent the data format protocols?**

**Example of protocols:**

- **List of valid regions**
- **Minimum/maximum value for a given variable**

**Why this form of representation?**




# File Upload Design

AgMIP GlobalEcon Data Submission

1 File Upload — 2 Data Specification — 3 Integrity Checking — 4 Plausibility Checking

1) Upload a data file <sup>⌵</sup>

 Add a CSV file from your computer

DemoFile.csv ✕

2) Select associated projects

- agmipglobalecondata
- agmipglobaleconagclim50iv

Next

# Data Specification Design

AgMIP GlobalEcon Data Submission

1 File Upload — 2 Data Specification — 3 Integrity Checking — 4 Plausibility Checking

**Specify the format of the input data**

Model name \*  Delimiter \*

Header is included \*  Number of initial lines to skip \*

Scenarios to ignore

**Assign columns from the input data to the output data**

Model	Scenario	Region	Variable	Item	Unit	Year	Value
FARM	b) Scenario	c) Region	e) Variable	d) Item	g) Unit	f) Year	h) Value

**Preview of the input data**

a) Model	b) Scenario	c) Region	d) Item	e) Variable	f) Year	g) Unit	h) Value
FARM	SSP2_NoMt	World	Tot	Popt	2010	million	7370.274372
FARM	SSP2_No...	SEA	RICE	POPT	2010	million	#Div/0!

**Preview of the output data**

Model	Scenario	Region	Variable	Item	Unit	Year	Value
FARM	SSP2_NoMt	World	Popt	Tot	million	2010	7370.274372
FARM	SSP2_No...	SEA	POPT	RICE	million	2010	#Div/0!

[Previous](#) [Next](#)

# Integrity Checking Design



### Rows overview

The table shows an overview of the uploaded data's rows. The rows can be downloaded to be analyzed.

Number of rows with structural issues (missing fields, ...)	776	⬇
Number of rows containing an ignored scenario	0	⬇
Number of duplicate rows	0	⬇
Number of accepted rows	229,589	⬇

### Bad labels overview

The table lists labels that are recognized by the program but do not adhere to the correct standard. They will be fixed automatically.

Label	Associated column	Fix
#Div/0!	Value	0
Tot	Item	TOT
Popt	Variable	POPT

### Unknown labels overview

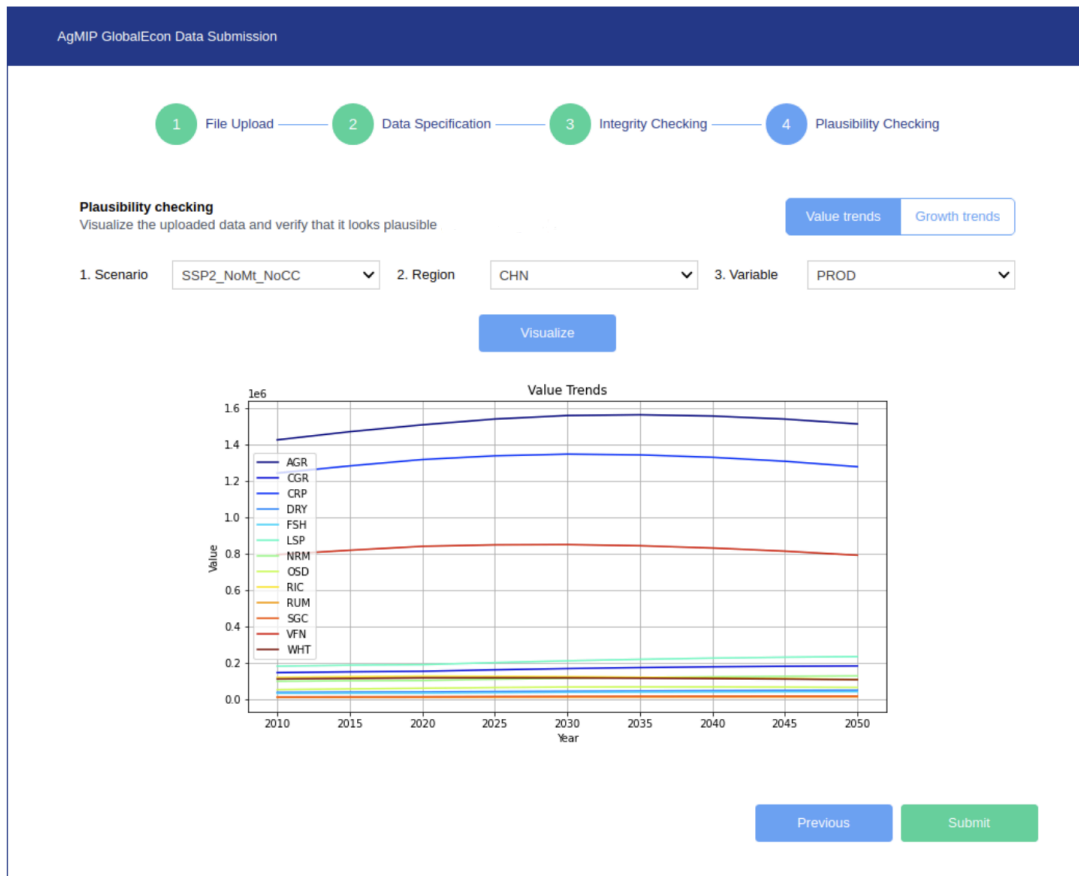
The table lists labels that are not recognized by the program. Please fix or override the labels, otherwise records containing them will be dropped.

Label	Associated column	Closest Match	Fix	Override
SSP2_NoMt	Scenario	SSP2_NoMt_NoCC	⌵	<input checked="" type="checkbox"/>
RICE	Item	RIC	RIC ⌵	<input type="checkbox"/>
-	-	-	⌵	<input type="checkbox"/>

Previous

Next

# Plausibility Checking Design



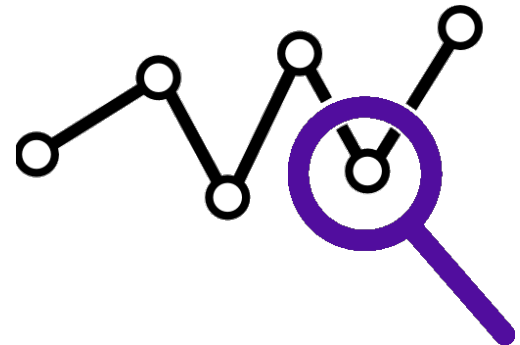


# Conclusion

# Future Improvement

**More sophisticated statistical analysis method**

**Manual vs automatic outlier detection**



# Acknowledgement

**My mentors, Lan Zhao and Rob Campbell from Purdue RCAC**

**The AgMIP Global Economics team**

**Discovery Undergraduate Research Internship (DURI)**



**The End**