Machine Learning Estimation of the Future Climate Risk Amplification of Food Security Induced Conflict

Climate change (e.g. global warming) and climate variability (e.g. La Nina and El Nino) amplify the risk of potential conflict between parties with existing tensions. For example, the Arab Spring uprising in 2010/2011 was triggered when existing political tensions within Egypt were heightened due to excessive bread prices, brought on by climate-induced wheat production reductions in exporting nations (e.g. Australia, China, Russia). We aim to understand, visualize and model the chain of influence from the: climate; to agricultural production; commodity prices; trade; and finally, the risk of potential conflict. A better understanding of the interactions between the physical climate and these human social systems is key to better assessing and proactively dealing with the associated climate impacts. This is particularly pertinent in a future world with increasing population and resource scarcity.

Client Information

- host name: Dr Vassili Kitsios
- host LinkedIn: https://www.linkedin.com/in/vassilikitsios/
- host organization CSIRO: https://www.linkedin.com/company/csiro/

Expected outcomes

There are several sub-components to addressing the above task. We would work with the interests of the students to ensure they are working on the sub-component most compatible with their interests and abilities:

1. Population: An understanding and visualization of existing population demographic projections per nation from now until the end of the century, and the curation of measures of historical social unrest and conflict.
2. Production: Assessment of the appropriate machine learning hyper-parameter settings in determining the relationships between a suit of climate variables and the agricultural production of commodities key to food security (e.g. wheat, rice, maize, soy).
3. Price: Assessment of the appropriate machine learning hyper-parameter settings in determining the relationships between agricultural yield and price for the above commodities, or alternatively between the climate variables and price directly.
4. Trade: Visualisation of the trade of these agricultural commodities between producing and consuming nations (including local consumption), and characterization of the stability of this trade network in time.
5. Climate: Comparison of climate model simulations to historical observations per scale (i.e. sized vortex), and hence determining which of these scales are trustworthy in future climate projections out to the end of the century.

External links

Internships & Project: https://industry.science.unimelb.edu.au/index.php?r=user/home

2023 Project List: https://canvas.lms.unimelb.edu.au/courses/152096/files/14426626?module_item_id=4658243

Contact Methods

Team information

Full name		LinkedIn Profile	Email Address
Ritwik Giri	RG	https://www.linkedin.com/in/ritwik-giri/	rgiri@student.unimelb.edu.au
Xiangyi He	Emma	https://www.linkedin.com/in/xiangyi-he-7a15a2269/	xiangyih1@student.unimelb.edu.au
Jiaqi Hu	Nick	https://www.linkedin.com/in/hu-jiaqi-498578202/	jiaqih7@student.unimelb.edu.au

Sunchuangyu Huang	Rin	https://www.linkedin.com/in/rin-huang/	sunchuangyuh@student.unimelb.edu.au
Jihang Yu	Jonathan	https://www.linkedin.com/in/jihang-yu-1022151a5/	jihangy@student.unimelb.edu.au

Staff information

Title	Name	Email
Subject coordinates	Michael Kirley	mkirley@unimelb.edu.au
Subject coordinates	Joyce Zhang	lele.zhang@unimelb.edu.au
Industry consultant	Marika Benetti-Hille	m.benetti@unimelb.edu.au
Project Supervisor	Feng Liu	feng.liu1@unimelb.edu.au
Project Host	Vassili Kitsios	Vassili.Kitsios@csiro.au

Discord: Join the DS Capstone - Group 36 Discord Server!

Slack: https://team36mast90106.slack.com

File sharing

Most of our files will be shared via google drive, it will include general documentation, meeting records, etc.

Google Drive: https://drive.google.com/drive/folders/1jOl69VxlQHf852pdPQAE0giSAc0qm9T8?usp=share_link Connect your Google account

Code repository: https://bitbucket.org/ds-cap-team36/ds-cap-team36/src/master/

Project management: 2023 Data Science Capstone Project

Project overleaf report: wait for the update

Project Milestone

1. Conducting a literature review of relevant prior work on the problem
2. Establishing a research question based on the literature review (in collaboration with the client)
3. Initial analysis of data
4. Development and evaluation of baseline method
5. Investigation and study design
6. System development and data analysis (perhaps including additional data collection)
7. Detailed analysis
8. The final write-up of the report

https://canvas.lms.unimelb.edu.au/courses/152096/pages/project-milestones?module_item_id=4594769

Working with BitBucket

There are two plausible ways to work with BitBucket.

- SourceTree
- Git command line tool

If you are not familiar with the command line tool, then SourceTree will be a good choice, this is similar to GitHub Desktop. Otherwise, please use AccessToken or SSH to clone the repository.

Clone a repository using SourceTree

Use these steps to clone from SourceTree, our client for using the repository command-line free. Cloning allows you to work on your files locally. If you don't yet have SourceTree, download and install it first.

1. You'll see the clone button under the **Source** heading. Click that button.
2. Now click **Checkout in SourceTree**. You may need to create a SourceTree account or log in.
3. When you see the **Clone New** dialog in SourceTree, update the destination path and name if you'd like to and then click **Clone**.
4. Open the directory you just created to see your repository's files.

Now that you're more familiar with your Bitbucket repository, go ahead and add a new file locally. You can push your change back to Bitbucket with SourceTree, or you can add, commit, and push from the command line.

Clone a repository using SSH

Before cloning, need to set up a valid ssh key, please follow Set up personal SSH keys on Linux | Bitbucket Cloud | Atlassian Support, then simply clone with the command:

```
git clone git@bitbucket.org:ds-cap-team36/ds-cap-team36.git
```

Clone repository with Access Token

DO NOT share this access token with someone else due to this key having full access to the current repository.

```
git clone https://x-token-auth:ATCTT3xFfGN0rLkuXi9RQzjpo3hNz9wSsjf_j9S53WIcjoRAcJ0r5Ah6NDXffEHLz0uiD4R7IS5BBLJEXj
```

After cloning, the repository simply edits or updates any scripts via the preferred editor. For example, use vs code via

```
cd ds-cap-team36  # change working directory
code .            # open vscode
```

Python Environment

The conda environment uses Python version 3.11.

- Performance improvements: Python 3.11 was expected to have faster interpreter startup times and improved performance for certain types of operations, such as list and dictionary manipulation.
- Improved error messages: Python 3.11 was expected to have more informative and user-friendly error messages for common mistakes.
- Structural pattern matching: A new feature called structural pattern matching was planned for Python 3.11. This would allow developers to match complex patterns in data structures using a syntax similar to the switch statement in other programming languages.
- Additional standard library modules: Several new modules were planned to be added to the Python standard library in Python 3.11, including the graphic module for working with graph structures and the zones module for handling time zones.
- Changes to typing annotations: The typing annotations in Python 3.11 were expected to be simplified and made more consistent with the rest of the language.

Install scripts

```
# Create a conda environment from conda environment.yml file
conda create --name newenv --file environment.yml

# e.g
conda create --name mds-g36 --file environment.yml

# After installation, activate conda env via conda activate
conda activate mds-g36

# install required python dependencies
pip3 install -r requirements.txt
```

Pre-build packages

In the environment.yml and requirements.txt, including:

- Jupyter, Jupyter Lab
- Matplotlib, Seaborn, Plotly, Folium, hvplot
- Pandas, Numpy, Geopandas, Polars
- Pytorch, Scikit-learn, statsmodel

Integrated CI/CD

Integrated CI/CD | Bitbucket Data Center and Server 8.14 | Atlassian Documentation

Integrate your CI/CD pipeline | Bitbucket Data Center and Server 8.14 | Atlassian Documentation

> In the project, we don't construct a full CI/CD pipeline due to we are unlikely to have any test on our code. Therefore, CI/CD only use to generate PR to avoid any potential coding conflict.

With Integrated CI/CD you can create a seamless idea to production workflow. It enables you to link Bitbucket Data Center and Server with Bamboo or Jenkins, and to combine this with a link to Jira Software to get a centralized view of your CI/CD pipeline.

Configuring, maintaining, and monitoring your pipeline can then be done from Bitbucket, allowing you to see build statuses and deployment information all within the context of your code.

CI/CD stands for Continuous Integration and Continuous Delivery (or Continuous Deployment). It is a software development approach that emphasizes automating the building, testing, and deployment of software applications.

The reason why CI/CD is important is that it enables developers to deliver high-quality software quickly and efficiently. With traditional software development methods, developers would write code and then wait for a long time before integrating their code with other developers' code, testing, and finally deploying the software. This process was slow, error-prone, and time-consuming.

With CI/CD, developers integrate their code frequently, often multiple times a day, with an automated process that tests the code for errors and ensures that it integrates smoothly with other code. Once the code passes the tests, it can be automatically deployed to production environments, reducing the time between code changes and deployment.

CI/CD helps to catch and fix errors early, reduces the risk of introducing bugs, and improves collaboration among developers. It also allows organizations to respond to changing customer needs and market demands more quickly, which is essential in today's fast-paced business environment.

Meeting notes in space

Create meeting note

Incomplete tasks from meetings

Description	Due date	Assignee	Task appears on
☐ Train models by extracting CPI inflation rate without deseasonalisation, the correlation aim should be around 0.6-0.7 +- 0.05			2023-09-04 Client Meeting Notes
☐ Incorporate additional CPI factor ← require deseasonalized			2023-08-21 Client Meeting Notes
☐ SARIMAX and another model for the next			2023-08-07 Client Meeting notes
☐ Modify the existing model to incorporate new date data, not solely relying on t-1 information.			2023-07-27 Client Meeting Notes
☐ Aim to forecast three months ahead to improve prediction capabilities.			2023-07-27 Client Meeting Notes
☐ Conduct an analysis to identify the best lag for the model (standard AR/ARIMA) without considering any ENSO/environmental factors. This model will be the baseline for future model development.			2023-07-27 Client Meeting Notes
☐ Use statistical techniques to select the lag that yields the most accurate results based on RMSE/RMAE.			2023-07-27 Client Meeting Notes
☐ Utilize VARMAX to check for cointegration among the variables in the model.			2023-07-27 Client Meeting Notes
☐ Perform a Granger causality test to evaluate the causal relationship between an exogenous factor and the model's dependent variable.			2023-07-27 Client Meeting Notes
☐ Compare the results of the two models to determine if the effect of the exogenous factor is Granger causal or not.			2023-07-27 Client Meeting Notes

Decisions from meetings

Page Title	Decisions
2023-03-23 Supervisor Meeting notes	Supervisor Feng suggests completing the second objective first, then extending the model to the rest of the topics and finding the influence/result of machine learning models. Regular meeting time with supervisor Feng fortnightly, from 11:00 am to 12:00 pm, at Feng's office (room 108 Old Geology)
2023-04-06 Team Meeting notes	The project focuses on the "price" section Create documentation for each sub-topic Create a slide for Friday's host meeting, demonstrate group understanding on the topic
2023-04-14 Client Meeting notes	Use AR as a baseline model.
2023-05-04 Team Meeting notes	S1 Deliverable Job Allocation
2023-05-11 Client Meeting notes	In the report, reference the ENSO circulation plot from Vassile's paper.
2023-07-27 Client Meeting Notes	The team aims to define the baseline model (without MEI) in the next week. Meeting time reschedule with Feng.
2023-08-03 Client Meeting Notes	Task distribution on four types of commodities: wheat, maize, soy, rice

All meeting notes

Title	Creator	Modified
Climate Change Risk Assessment	Sunchuangyu HUANG	Oct 09, 2023
2023-10-02 Meeting notes	Jiaqi Hu	Oct 02, 2023
2023-07-27 Client Meeting Notes	Sunchuangyu HUANG	Sep 28, 2023
2023-05-11 Client Meeting notes	Sunchuangyu HUANG	Sep 28, 2023
2023-04-27 Client Meeting notes	Sunchuangyu HUANG	Sep 28, 2023
2023-04-14 Client Meeting notes	Sunchuangyu HUANG	Sep 28, 2023
2023-03-30 Client Meeting notes	Sunchuangyu HUANG	Sep 28, 2023
2022-09-25 Client Meeting Notes	Sunchuangyu HUANG	Sep 28, 2023
2023-08-28 Client Meeting notes	jihangy	Sep 28, 2023

2023-09-11 Client Meeting Notes	Sunchuangyu HUANG	Sep 28, 2023
2023-08-21 Client Meeting Notes	Sunchuangyu HUANG	Sep 11, 2023
2023-09-04 Client Meeting Notes	Sunchuangyu HUANG	Sep 11, 2023
2023-08-14 Client Meeting notes	Jiaqi Hu	Aug 22, 2023
2023-08-03 Client Meeting Notes	Sunchuangyu HUANG	Aug 22, 2023
2023-08-02 Supervisor Meeting Notes	Sunchuangyu HUANG	Aug 22, 2023
2023-08-07 Client Meeting notes	Sunchuangyu HUANG	Aug 22, 2023
2023-08-01 General Meeting Notes	Sunchuangyu HUANG	Aug 01, 2023
2023-05-04 Team Meeting notes	Sunchuangyu HUANG	May 07, 2023
2023-04-13 Team Meeting notes	Sunchuangyu HUANG	Apr 29, 2023
2023-04-06 Team Meeting notes	Sunchuangyu HUANG	Apr 29, 2023

Find more results

2023-10-02 Meeting notes

📅 Date

Oct 2, 2023

👥 Participants

- @Jiaqi Hu @rgiri @Sunchuangyu HUANG @jihangy @Xiangyi He
-

🥅 Goals

-

🗣 Discussion topics

Time	Item	Presenter	Notes
14:30	Presenting the current report progress	@rgiri	
14:32	Asking if the RMSE and correlation graph should be included in the report	@rgiri	Vasili said that the graphs could be included in the appendix of the report, Feng said we can have the graphs be aside of the tables
14:35	Presentation on Oct 12th	@rgiri	Consider give audience some background information (reasoning on predicting the commodity price on climate information) on the Oct 12th
14:38	Giving suggestions of the thesis paper	Feng Liu	Give some more motivation examples on the starting of the paper. Section 7 and 8 can be merged together, @rgiri said the naming “methodology“ can be changed to “research structure“, Feng gave some suggestions on changing use of words on the paper
14:44	Consider the Neural Network to be a parallel part of the ARIMA model in the report	Feng Liu	Be mindful about the reference style \bibliographystyle{apalike}
14:56	Deciding the time to send the report to	everyone	Discuss the report on Thursday

	our client and meet in person		

Action items

- [ ]

Decisions

2022-09-25 Client Meeting Notes

📅 Date

Sep 25, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri
- vassili.kitsios@csiro.au

🗣 Discussion topics

Focus on Report Structure

- **Rolling Window Forecast**
  - The focus is on step-ahead forecasting.
  - It should not be interpreted as a long time period time series forecast.
- **Comparison with Other Teleconnections**
  - ENSO doesn't provide significant benefits.
  - Table of other teleconnections should be generated, focusing on the most recent decades for each.
- **Limitations and Suggestions**
  - Limitations of introducing exogenous variables into the forecast as linear terms.
  - Suggest incorporating non-linearity with ENSO.
  - Using neural networks might enhance the prediction capabilities.
- **Factors Influencing Price**
  - Climate doesn't directly influence the price.
  - Food shortages or insecurity are major influencers.
  - The quantity (total kilograms) can also influence the price.
- **Neural Networks**
  - Consider section10 & 11 together, focusing solely on current neural network methodologies and ignoring DBN.

https://www.chathamhouse.org/sites/default/files/2021-09/2021-09-14-climate-change-risk-assessment-summary-quiggin-et-al_0.pdf

The above link is a good reference as the supporting evidence to express why climate change is important to food security.

✅ Action items

⤴ Decisions

2023-09-11 Client Meeting Notes

📅 Date

Sep 11, 2023

👥 Participants

- @jihangy @Sunchuangyu HUANG @Jiaqi Hu @rgiri @Xiangyi He
- vassili.kitsios@csiro.au
- feng.liu@unimelb.edu.au

🗣️ Discussion topics

- **Result Demonstration**
  - Lower correlation is aligned with expectation.
- **Correlation**
  - There's a statistical difference between the correlations.
  - Types of correlation discussed:
    - Pearson
    - Cross-correlation
- **De-seasonalization**
  - The process involves removing the seasonal cycle.
- **Modeling with SARIMAX**
  - No seasonal value set, hence it performs as ARIMAX.
  - Refitting is required for the models.
  - If the model structure remains the same, then they are comparable.
- **Adding ENSO**
  - This will lead to an increase in the correlation.
- **Statistical Testing**
  - Two time series are to be considered.
  - Calculation of Residual Sum of Squares (RSS) is required.
  - Number of parameters are to be considered.
  - Calculation of the F-score is essential.
  - If value reduces by a significant (sufficient) amount, influence testing is needed.
  - Run statistical tests on neural networks for potentially better results due to more parameters.
- **Report Recommendations**
  - The report should emphasize changes observed from decade to decade.
- **Wheat and ENSO**
  - Wheat isn't necessarily related to ENSO.
  - Wheat production seems to be less affected by ENSO.
  - Non-linear function of ENSO might yield better results due to the absolute value.

✅ Action items

⤴️ Decisions

2023-09-04 Client Meeting Notes

📅 Date

Sep 6, 2023

👥 Participants

- @jihangy @Sunchuangyu HUANG @Jiaqi Hu @rgiri @Xiangyi He
- vassili.kitsios@csiro.au
- feng.liu@unimelb.edu.au

🥅 Goals

- Weekly update

🗣 Discussion topics

- High correlation issue: Current models perform a nearly perfect prediction with a correlation score > 0.99
  - Per our client's suggestion, do not remove the seasonal cycle when extracting the inflation rate
  - The target correlation should be around 0.6-0.7 → The financial market is not stable and the price is constantly increasing due to inflation
  - Hence, the problem appears in a seasonal cycle, and we can't (impossibly) obtain the seasonal cycle in real life → the Current real price log return (ratio) contain future information? needs double check
- Assessing the model in the recent decades and previous decades
  - For different commodities, we don't expect commodities will share exactly the same lag
  - Again, the financial market is not stable hence the lag for different commodities depends on the time
    - e.g. for 2000-2010, the optimal lag may be 4 months, when proceeding to 2011 - 2020, the optimal lag would be 3, etc.
    - e.g. for a weak market, we expect a short lag
  - We need to evaluate the model that:
    - Are the testing statistics getting worse when the lag changes?
    - understanding the relationship of how the price changes in time → NOTE: Our target is to explain the changes instead of generate an accurate prediction
- Seasonal information → Keep 12 months of seasonal info
  - Check is lag 1 contains the exact information for future (at least 1 month)

✅ Action items

- [x] Check code logic on deseasonalisation, double deseasonalisation might produce overfitting results ← model is too good
- [ ] Train models by extracting CPI inflation rate without deseasonalisation, the correlation aim should be around 0.6-0.7 +- 0.05
- [x] Rerun code after the above progress

⤴ Decisions

2023-08-28 Client Meeting notes

📅 Date

Aug 28, 2023

👥 Participants

- @jihangy @Jiaqi Hu @rgiri @Xiangyi He
- vassili.kitsios@csiro.au
- feng.liu@unimelb.edu.au

🥅 Goals

- code diagnostic
- report formatting

🗣 Discussion topics

- Introduced the newly fitted model using log returns of price with inflation included and a 10 year time period with 5 year window to our client.
- The correlations of the models happen to be too high. should be around 0.5, 0.6(could be because of the wrong inflation calculation used in the model which is already fixed during the meeting.)
- For the SARIMAX seasonality order, we can set it to be 12 and when considering lags we only need to consider 1-12 lags since in reality a highly correlated lag of 2 always makes more sense to be chosen than a highly correlated lag of 14 as lag 14 is too outdated.
- Need a basic template on NN model.
- Need to start on the basic structure of the report.

✅ Action items

- [x] Rerun the code and see if the correlation issue is fixed. (If not then try on the model with old time window with inflation added first, then change time window, monitor correlations at each stages to diagnose the code.) correlation should be around 0.5.
- [x] set SARIMAX seasonality order to 12 and use lag 1-12 in models.
- [x] complete code for a NN model template.
- [x] Start on the basic structure of the report + intro + literature review

⤴ Decisions

2023-08-21 Client Meeting Notes

📅 Date

Aug 22, 2023

👥 Participants

- @Sunchuangyu HUANG @Jiaqi Hu @rgiri @Xiangyi He @jihangy
- feng.liu1@unimelb.edu.au
- kitsios.vassili@csiro.com.au

🥅 Goals

- Weekly update
- Report structure (team

🗣️ Discussion topics

- Update meeting results
  - Previously the team used ENSO lagged 0 (current value) as an exogenous factor for model prediction → alter to model use lagged value starting from lag 1
  - The model has the capability to find the best ENSO lag based on the evaluation matrix, however, the user needs to manually select the ENSO value.
- Lag combinations for commodity log return and ENSO lags
  - There is no such way to stop the grid search early → absolutely brutal force and in the worst case we need to explore 24! models for each commodity
- Rolling window issue → 4 years potentially caused an unexpected result e.g. the lowest RMSE/STD happens at lag 24 (typical ENSO cycle happens every 7 months
- Add CPI data to exclude the global market influence
- Confidence interval issue
  - Confidence intervals for autocorrelation function this function calculates the CI for ACFS which does not align with our target
  - In Vassili's report, the grey zone indicates the model with ENSO or no ENSO is statistically indistinguishable around 0 ← this is an F-stats test, see additional information in the update email
- Model sensitivity analysis
  - Granger Causality test ← incomplete

✅ Action items

- [x] Sensitivity analysis to 10 (120 data points?) years window → 5 years training + 5 years analysis?
- [x] Try to find the best ENSO combinations of 2 and log return combinations for a commodity
- [x] F-stat test on model results significance
- [ ] Incorporate additional CPI factor ← require deseasonalized

2023-08-14 Client Meeting notes

📅 Date

Aug 14, 2023

👥 Participants

- @Jiaqi Hu @jihangy @rgiri @Xiangyi He @Sunchuangyu HUANG
- feng.liu1@unimelb.edu.au
- Vassili.Kitsios@csiro.au

🥅 Goals

-

🗣 Discussion topics

Time	Item	Presenter	Notes
14:30	Briefing about the findings after removing seasonal factors of the commodity model	@Sunchuangyu HUANG	•
14:35	talks about that lag = 1 has the most correlation for price models	Vassili	
14:45	suggests that try to also use lag = 1 when exogenous factor MEI is included	Vassili	
14:47	the host states that historical data does not affect recent results very much in time series	Vassili	
14:50	would include ways to exclude sensitivity (economic shocks) after further exploration	Vassili	

14:52	Remember to shift the exogenous factors by *p* if we want to include lag with them	Vassili	
14:57	TODO: find out how to plot the confidence interval of every predictions	Everyone	

Action items

- [ ]

Decisions

2023-08-07 Client Meeting notes

📅 Date

Aug 7, 2023

👥 Participants

- @rgiri @Xiangyi He @Jiaqi Hu @Sunchuangyu HUANG @jihangy
- feng.liu1@unimelb.edu.au
- Vassili.Kitsios@csiro.au

🥅 Goals

- Go through the current work progress
- Question about the model construction

🗣️ Discussion topics

- The historical problem is don't use data that are too old (no more than 13 months)
- Order issue: lag 0 similar to lag 1, double-check with the number 0 (code behaviour check)
- The log return issue needs to be de-seasoned, remove the average from the standard log return (**take away seasonal cycle first**)
- Seasonal ARIMA
- Standard order AR model ← 12 exogenous variables, same thing for each month
- Confidence interval issues → Confidence interval overlapping
- removing economic spikes from GDP stuff

✅ Action items

- [x] Train a model with no seasonal cycle in log return by subtracting the mean value.
- [x] Add one climate feature to see the effect of ENSO/MEI, etc.
- [ ] SARIMAX and another model for the next

⤴️ Decisions

2023-08-03 Client Meeting Notes

📅 Date

Aug 3, 2023

👥 Participants

- @rgiri @xinagyih1 @Jiaqi Hu @jihangy @Sunchuangyu HUANG
- vassili.kitsios@csiro.au

🥅 Goals

- Confirm the final approach for the baseline AR model
- To simplify the progress, focuses on lag (p) for the AR model, leaving the moving average and other methods as the future work

🗣 Discussion topics

- For future forecasting, the training data should use a fixed length time interval
- Denoising problem
  - The team observed a spike (noise point) and this might have resulted by the economic crisis during that time
  - Denoising could improve the result but it simplifies the process → unexpected trend and the result will be close to a mean value,
    - Denoising also might incorporate future data which we don't want to use intentionally
- Current model shows a common problem: off-by-one bug
- statsmodels issue: this library improves the performance when facing a complex problem (increasing the dimension of data input) but we might be over-training the data points
- For future work, we could compare the influence caused by the smoothing and denoising techniques
- Capstone work direction: autoregression → moving average
- Use time series to predict the size of change instead of the actual point value

✅ Action items

- [x] Base model construction, ETA end of the week
- [x] Base model ← simple AR model with no smoothing or denoising techniques

⤴ Decisions

Task distribution on four types of commodities: wheat, maize, soy, rice

2023-08-02 Supervisor Meeting Notes

📅 Date

Aug 2, 2023

👥 Participants

- @rgiri @Jiaqi Hu @jihangy @Sunchuangyu HUANG
- feng.liu1@unimelb.edu.au

🥅 Goals

- Go through the last client meeting /page
- Go through the current stage of AR model selection and model construction issue

🗣 Discussion topics

- General stuff of the last meeting
- AR model training issue:
  - Training data using all historical information might result in the model converging to the mean value.
  - Smoothing/Normalisation does improve the model performance
  - Rolling update is easy to achieve but this consider as a multi-step forecast
- Future group meeting date: every Monday at 2:30 PM is preferred

✅ Action items

- [x] Build model focuses on the one-step-ahead forecast
- [x] The baseline model shouldn't be complex but it might be hard to beat
- [x] Noise point might be an issue and smoothing techniques does improve the performance

⤴ Decisions

2023-08-01 General Meeting Notes

months'📅 Date

Aug 1, 2023

👥 Participants

- @rgiri @Xiangyi He @Jiaqi Hu @jihangy @Sunchuangyu HUANG
- lele.zhang@unimelb.edu.au

🥅 Goals

- General meeting about the group process

🗣 Discussion topics

- Team information: **Team 36**
- Team members information
- Capstone Topics: **Machine Learning Estimation of the Future Climate Risk Amplification of Food Security-Induced Conflict**
- Focused subtopic: **Price estimation based on El Niño/La Niña Southern Oscillation (ENSO) factor MEI for commodity price**
- Project description page: https://canvas.lms.unimelb.edu.au/groups/285662/wiki
- Client Brief:
  - Client information: Vassili Kitsios https://www.linkedin.com/in/vassilikitsios/
  - Client Organisation: CSIRO
- Currently, the team focuses on wheat price prediction with a one-step-ahead forecast based on the last 48 months' data
- Data Privacy Issues: Open-sourced project, all data are available from a public website.
- Findings in semester 1:
  - ENSO does affect the commodity price prediction
  - It is a cyclic event for around 8 months (suggested in Vassili's report)
  - Involving more climate information does help to generate a better commodity price forecast model for most commodities (using 4 types of wheat as examples.)
- Next goal:
  - Construct the baseline model using commodity price only
  - Find the best lag for the above model (probably involved enumeration techniques)
- Expected Final Deliverable:
  - A one-step-ahead commodity price model with ENSO and other climate factors

Other items:

- Bitbucket Repository: 🔒 https://bitbucket.org/ds-cap-team36/ds-cap-team36/src
- Confluence Documentation: 2023 Data Science Capstone Team 36
- Group Discord: 🔗 Discord - A New Way to Chat with Friends & Communities

Action items

Decisions

2023-07-27 Client Meeting Notes

📅 Date

Jul 27, 2023

👥 Participants

@rgiri @Jiaqi Hu @Xiangyi He @Sunchuangyu HUANG @jihangy

🥅 Goals

- update the current state to host

🗣️ Discussion topics

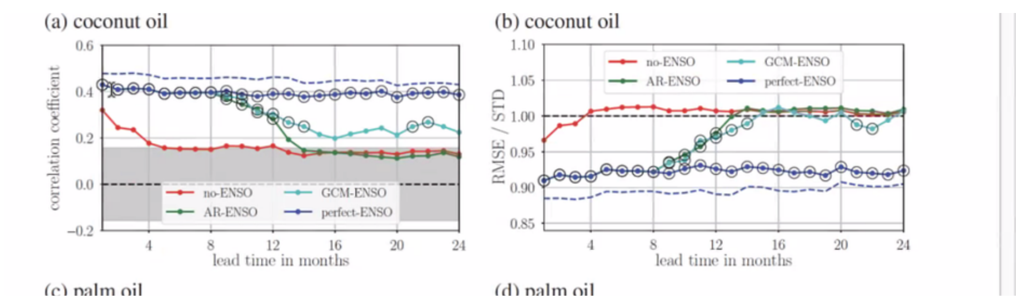
Naive Forecast → simplest forecast model

- Standardization on wheat log return
- Causality test
- Cointegration test
- VARMAX (Vector Auto Regression)

Problem with the date range:

- should be monthly data
- double-check wheat log return calculation

The scale of one-step ahead forecast problem, model usage problem.



- no-ENSO (red line) only using the information based on the commodity price.
- no hardcode, prediction should be moving based on the current date datapoint
- 1 lag at a time → if looking over a year, we will have 12! possible combinations for Lags
- circle → granger causality test → which is statistically significantly better than others

Don't need to worry about ENSO prediction → assume we have the perfect ENSO historical data for prediction

✅ Action items

- [x] Review and validate the data range to ensure accuracy and completeness.

- [x] Update the x-axis to reflect the most recent date data.
- [ ] Modify the existing model to incorporate new date data, not solely relying on t-1 information.
- [ ] Aim to forecast three months ahead to improve prediction capabilities.
- [ ] Conduct an analysis to identify the best lag for the model (standard AR/ARIMA) without considering any ENSO/environmental factors. This model will be the baseline for future model development.
- [ ] Use statistical techniques to select the lag that yields the most accurate results based on RMSE/RMAE.
- [ ] Utilize VARMAX to check for cointegration among the variables in the model.
- [ ] Perform a Granger causality test to evaluate the causal relationship between an exogenous factor and the model's dependent variable.
- [ ] Compare the results of the two models to determine if the effect of the exogenous factor is Granger causal or not.

Decisions

- The team aims to define the baseline model (without MEI) in the next week.
- Meeting time reschedule with Feng.

2023-05-11 Client Meeting notes

📅 Date

May 11, 2023

👥 Participants

- @rgiri @xiangyih @jihangy @Jiaqi Hu @Sunchuangyu HUANG
- Client: vassili.kitsios@csiro.au
- Supervisor: feng.liu@unimelb.edu.au

🥅 Goals

- Semester Presentation Delivery

🗣 Discussion topics

Time	Item	Presenter	Notes
11:05	Presentation Start	@jihangy	• Introduction & methodology
11:07	Reason behind our focus area	@Xiangyi He	
11:10	Research in ENSO based on Vasilli's paper	@rgiri	
11:12	Deploy base model	@Sunchuangyu HUANG	
11:14	Summary & Future expectations	@Jiaqi Hu	
11:16	Q&A Section	@Vassili Kitsios and @feng.liu	

✅ Action items

Project related

- ENSO effects appear for at least 8 months, looking one month ahead may return nothing.
- JRA55 and NNR1 use different inference methods but they should provide identical observation result
- El Niño and La Niña

Presentation Suggestions

- For the audience, need to provide more background knowledge
- Figures might be a good start, a good first impression with less explanation.

⤴ Decisions

In the report, reference the ENSO circulation plot from Vassile’s paper.

2023-05-04 Team Meeting notes

📅 Date

May 7, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri @xiangyih @jihangy

🥅 Goals

- Presentation preparation outlines
  - Pre-host and structure → This structure also suit for report
    - Johnathan: Intro and talk about individual team member roles & methodology
    - Ritwik: Coconut oil example (As methodology will be a part of the example)
    - Rin: base ar model for all the crops
    - Emma: Literature Review (Will connect with Nick for his side of the research papers)
    - Nick: Future Workflow Explain
- Canva Link: Group 36 Part 1 Presentation
- Deliverable: https://docs.google.com/document/d/1icNJTOKH_iBzuDBh2CkNla8IJK3qO_NyXi8si_QyunE/edit Connect your Google account
- Transcript: https://docs.google.com/document/d/1NWXoruqcSYX47CzD0YVeawDtiqQG4MYwcOB7poeNh9c/edit# Connect your Google account

⤴ Decisions

S1 Deliverable Job Allocation

2023-04-27 Client Meeting notes

📅 Date

Apr 27, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri @xiangyih @jihangy @Jiaqi Hu
- Client: vassili.kitsios@csiro.au
- Supervisor: feng.liu@unimelb.edu.au

🥅 Goals

- Received comments about coconut oil and wheat EDA.
- Confirmed with client about the presentation on May.11th.

🗣 Discussion topics

Time	Item	Presenter	Notes
11:02	Demonstrated the EDA about log return of coconut oil price incorporate with MEI using AR model.	@rgiri	• To find the best lag variable, the aim is to find a point MSE or RMSE approximate to 0. (for AR model ideally should not over 12.) • Given the sample size N is small, large RSS(penalized) means the model does not fit well.
11:20	Demonstrate the EDA on wheat price and factors including MEI, CO2, etc. using ARIMA model.	@Sunchuangyu HUANG @Jiaqi Hu	• Use log returns on wheat price when modelling. • Directions of MEI may not be a key factor but its magnitude. • Since the sample size is limited, focus on only wheat price and MEI for now.(excluding other factors like CO2.) • Choose to analyze one type of wheat first if results on other kinds are not significant.
11:40	Asked questions and confirmed future actions with client.	@Everyone	• In Vassili's paper 2.4 equation3: a_l^{EE} is a 1*12 binary vector capturing seasonality. • First presentation on May.11th and make modifications after base on client's suggestion.

✅ Action items

- [x] Use log return of price on wheat problem
- [x] Try using percentage error in evaluating predictions(Mean Average Percentage Error along with RMSE)
- [x] Focus on only MEI for now and one type of wheat if others don't provide significant results.
- [x] Prepare for the presentation on May.11th
- [x] Start to find literatures for the final report
- [x] Add Feng and Vassili to JIRA

Decisions

2023-04-26 Team Meeting notes

📅 Date

Apr 26, 2023

👥 Participants

- @Sunchuangyu HUANG @Jiaqi Hu @rgiri @Xiangyi He @jihangy

🥅 Goals

- Discussed the topics and questions that we should cover in our next meeting with the host.
- Discussed about the final assignment.

🗣 Discussion topics

Time	Item	Presenter	Notes
9:00-9:15	Finalized the content for our upcoming meeting with the host, and discussed the agenda and update the further analysis of the coconut oil model.	@rgiri @jihangy	RG explained his notebook about the MEI and coconut oil analysis to Jihang Yu for the meeting on Thursday as Jihang is the host, listed his questions about the log return of the model, …, etc to be brought to the client for answering.
9:15-9:30	Explained some of the parameters used in the wheat analysis	@Sunchuangyu HUANG	SARIMA standas for Seasonal ARIMA. SARIMA is an extension of the ARIMA model that incorporates seasonality. • p (AutoRegressive order) • d (Integrated order) • q (Moving Average order) • P (Seasonal AutoRegressive order) • D (Seasonal Integrated order) • Q (Seasonal Moving Average order) • s (Seasonal period) # Define the SARIMA parameters p, d, q = 1, 1, 1 P, D, Q, s = 1, 1, 1, 12 These are the time series parameters to be used in the model, we pay extra attention on *p* and *s* since *p* is the step we take when doing predictions and *s* is the seasonal period, setting to 12 stands for one year.
9:30-9:40	Discussed about the presentation and report for the final assignment	Everyone	Do more research on climate change and food security, and need for more sources and references related to the models we're using. Start preparing for the REPORT and PRESENTATION.

✅ Action items

- [ ]

Decisions

2023-04-23 Team Meeting notes

📅 Date

Apr 23, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri @Xiangyi He @jihangy @Jiaqi Hu

🥅 Goals

- Discuss the results on exploring the factors that are potentially affecting crops's price
- Present the current models we have built on coconut oil and wheat and compare its prediction with the history price

🗣 Discussion topics

Time	Item	Presenter	Notes
17:00-17:15	Demonstrate the notebook about the time series model which is built on the history price and MEI (Multivariate ENSO Index)	@rgiri	Explored the MEI factor using boxplot, found out multiple outliers that may potentially affect the accuracy of predicting the price. After trying different length of time window for the model and choose 10 months as the optimal one. Out of expect, the p-value of the MEI shows that it is insignificant to the price of the coconut oil, which contradicts to the conclusion of Vasili's paper.
17:15-18:00	Illustrate the exploration of the factors and price data about US and Canadian wheat and the outcomes	@Sunchuangyu HUANG	'WHEAT_CANADI' has NA value in the dataset between 1950's to 1980's and after 2012, need to find ways to fill out these blanks (e.g. averaging). Plotted boxplot and correlation of every factor, 7 in total, found out that there are only 3 to 4 factors which strongly relates to the price. The model that involves all 7 factors has a great result upon the price of "US_SRW" and "US_HRW", but problematic on the price of "WHEAT_CANADI", which may relate to the recent inflation caused by Ukrainian War.
18:00-18:10	Discuss the time for the next group meeting	Everyone	On 25th of April, time TBD

2023-04-14 Client Meeting notes

📅 Date

Apr 14, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri @xiangyih @jihangy @Jiaqi Hu
- Client: vassili.kitsios@csiro.au
- Supervisor: feng.liu@unimelb.edu.au

🥅 Goals

- Why did we choose/focus on the price? (Motivation
- EDA results and need to explain factor name, also data description doesn't match the true data we have.
- Base model selection.

🗣️ Discussion topics

Time	Item	Presenter	Notes
11:02	Specify the reason and inform the host the team choose topic 2, price prediction as the focus area	@Xiangyi He	https://docs.google.com/document/d/1gS-ta80HE7fBkjfmY1faEd9YyVqTQCcWKbXUHtI5VqI/edit Connect your Google account
11:08	Terminology of the datasets	@rgiri	e.g. 'WHEAT_US_SRW','WHEAT_CANADI', 'WHEAT_US_HRW'
	How to handle NaNs (can we use some filling techniques without hampering the analysis)		
	Spike (inflation) is a maximum in 2008 - 2010. Could that because of the recession? Is data centric to US economic conditions.		
11:25	Are we consider GCM or AR as a base model?	@Sunchuangyu HUANG	Forecasting commodity returns by exploiting climate model forecasts of El Niño Southern Oscillation

	Should we follow the formula in the paper or customize it based on it.		
	Should we consider `ENSO` as an exogenous factor (although there are existing studies on it)?		

Summary

- Discussed the motivation: “Why Choose **Price** as the main focus area“.
- Explain terminologies in datasets (e.g. WHEAT_US_SRW, etc.). The team will consider these features as different crops to construct the time series model.
- In addition, according to Vassili’s suggestion, choose the Auto Regressive model as the base model and focus on the effect caused by jra55.
- Consider ENSO as an exogenous factor.

Action items

- [x] Build an AR model based on the jra55-MEI dataset as a base model, consider using the existing algorithm and mathematical formula mentioned in the paper (*Forecasting commodity returns by exploiting climate model forecasts of El Niño Southern Oscillation*). Consider ENSO as an exogenous factor. At the current stage, the AR model should particularly focus on wheat crops.
- [x] For missing values in datasets. try to find the historical inflation rate, use interpolation or build a statistical model to find the price.
- [x] Further study on background knowledge, find underlying relationships between climate change and commodities' prices.
  - [x] Reading: Harries2020_JAMES.pdf
  - [x] Reading: JMA2013_techreport.pdf
  - [x] Reading: JRA-55_handbook_LL125_en.pdf
  - [x] Reading: JRA-55_handbook_TL319_en.pdf
  - [x] Reading: Kobayashi2015_JMSJ.pdf
- [x] Generate a summary with the next action items, and send the summary to the host, supervisor, and industrial consultant.
- [x] Create a regular meeting schedule in both the calendar and Zoom. Set it in a recurrent mode.

Decisions

Use AR as a baseline model.

2023-04-13 Team Meeting notes

📅 Date

Apr 13, 2023

👥 Participants

- @Sunchuangyu HUANG @rgiri @xiangyih @jihangy @Jiaqi Hu

🥅 Goals

- Make a slide for the client meeting on 2023-04-14 Client Meeting notes
- Go through EDA results, find data patterns
  - However, there are features that need more explanation, for example, `WHEAT_US_SRW`, `WHEAT_CANADI`, `WHEAT_US_HRW`
- From the reading, choose either GCM or AR as a base model ← for late decision

🗣️ Discussion topics

Time	Item	Presenter	Notes
9:05	EDA	@rgiri	• EDA on world bank data price data • Price of wheats increase from early 1960-01-01 • A peak shows up in 2008 to 2009 this may results by economic crisis in 2008
9:10	Model Selecting	@Sunchuangyu HUANG	• Client used two models: GCM and AR • Both model show a relatively good performance • Consider AR as a basemodel
9:08	Meeting Agenda Summarised	@Xiangyi He	• Specify the reason that the team choose commodities price as the focus area • Meeting agenda

✅ Action items

- [x] Create meeting agenda 2023-04-14 Client Meeting notes
- [x] More exploring on data sets, ask client to explain key factors (abbreviations)

⤴️ Decisions

2023-04-06 Team Meeting notes

📅 Date

Apr 6, 2023

👥 Participants

Team: @Jiaqi Hu @rgiri @Sunchuangyu HUANG @xiangyih @jihangy

🥅 Goals

- Figure out the main direction and scope of the project

🗣 Discussion topics

Time	Item	Presenter	Notes
5min	Making the final decision for the sub stream of the project	All members	**Expected outcomes for the project upon completion** There are several sub-components to addressing the above task. We would work with the interests of the students to ensure they are working on the sub-component most compatible with their interests and abilities: 1. Population: An understanding and visualisation of existing population demographic projections per nation from now until the end of the century, and the curation of measures of historical social unrest and conflict. 2. Production: Assessment of the appropriate machine learning hyper-parameter settings in determining the relationships between a suit of climate variables and the agricultural production of commodities key to food security (e.g. wheat, rice, maize, soy). 3. Price: Assessment of the appropriate machine learning hyper-parameter settings in determining the relationships between agricultural yield and price for the above commodities, or alternatively between the climate variables and price directly. 4. Trade: Visualisation of the trade of these agricultural commodities between producing and consuming nations (including local consumption), and characterisation of the stability of this trade network in time. 5. Climate: Comparison of climate model simulations to historical observations per scale (i.e. sized vortex), and hence determining which of these scales are trustworthy in future climate projections out to the end of the century. Here are the 5 sub streams, as we have discussed in last week, the host suggests to focus on the “Price” section.
10min	Allocate which team member to write the documentation for the reasons of choosing this sub stream	TBD	Briefly describe the problem and perhaps include some investigation and thoughts, according to our supervisor’s words, we can build a simple 1 step time series model for Price and Production, then discuss the possibilities to explore other sub streams or derive some other findings.
5min	Play around with some prepacked model getting a better understand for datasets	All members	Should be focused on the term “time series”, look for potential models, such as ANN, RNN, LSTM, XGBoost etc. (maybe we need to do the EDA to get to know the data first?)
5min	Summarise the documentations, videos and data our host sent to us	All members	Go over the documents that the hosts shared with us and make sure we keep on track and be ready for the next meeting

			Causal inference in complex multiscale systems - AI for Missions Program

Action items

- [x] Update README.txt, fix a wrong description
- [x] Start a new email session
- [x] Create a slide for the next meeting
- [x] Create the next team meeting Schedule
- [x] Create the next team meeting and host meeting notes

Decisions

- The project focuses on the “price” section
- Create documentation for each sub-topic
- Create a slide for Friday's host meeting, demonstrate group understanding on the topic

2023-03-30 Client Meeting notes

📅 Date

Mar 30, 2023

👥 Participants

Team: @rgiri @xiangyih @Sunchuangyu HUANG @Jiaqi Hu @jihangy

Client: vassili.kitsios@csiro.au

Supervisor: feng.liu@unimelb.edu.au

Consultant: m.benetti@unimelb.edu.au

🥅 Goals

Project Spec:

- Since this is a really big project, the host suggests the team focus on a single task, e.g. Price or Production. (the decision is upon the team)
- The project mainly uses time series analysis with a monthly interval.
  - Feng has past experience in time-series forecasting
- Data is available via dropbox: 20230330_student_pack.tgz
  - All data are formatted in CSV
  - If any missing data, inform the client
- Project website: Causal inference in complex multiscale systems - AI for Missions Program
- Seminar related to the project: Seminar Presentation - Macquarie University And CSIRO - Part 2/2
- YouTube video prepared for financial audiences: [KEYNOTER] Vassili Kitsios - Influence of Climate Variability on Financial Markets & Health Sector
- Free to use public dataset but cited and provide download link
- The client prefers the programming language Python

Project deliverable:

- The S1 final presentation and report rubric is not released yet
- Presentation conducted involving host
- The final product should be a report comparing different types of models among given data features.

Administrative:

- No more agreement signed issue but potentially requires a policy check.
- The client is happy to meet fortnightly on Thursday from 11:00-12:00 with an online meeting preference. If Vassili is able to conduct the on-site meeting, then meet on campus.
- Documentation, code storage, project management, and communication tools depend on the team preference -> probability integrate JIRA, confluence, GitLab/bit-bucket for future projects.

🗣️ Discussion topics

Time	Item	Presenter	Notes

Action items

- [x] No action on signed agreements, UoM legal team will reach CSIRO
- [x] Create a regular meeting calendar including a zoom link
- [x] Check the next meeting time (14 April 2023
- [x] Go through the project website and do some research, write notes
- [x] Conduct basis EDA on given datasets
- [x] Setup JIRA, Confluence, BitBucket

Decisions

2023-03-23 Supervisor Meeting notes

📅 Date

Mar 23, 2023

👥 Participants

Team: @xiangyih @Jiaqi Hu @rgiri @Sunchuangyu HUANG @jihangy

Supervisor: feng.liu1@unimelb.edu.au

🥅 Goals

- Decide the primary subject from 5 subtopics
- Send an invitation email to client Vassili Kitsios for the first group meeting
- Inviting Feng to join the slack group
- Question about the word “undergraduate” in the CSIRO agreement
- Decide a regular meeting time with the supervisor
- Prepare the first host/client meeting.

🗣 Discussion topics

Time	Item	Presenter	Notes
			•

✅ Action items

- [x] Sending client team brief and meeting invitation email @xiangyih
- [x] Background research on CSRIO @rig @xiangyih @Sunchuangyu HUANG @Jiaqi Hu @jihangy
- [x] Capstone topic decision @rig @xiangyih @Sunchuangyu HUANG @Jiaqi Hu @jihangy

⤴ Decisions

Supervisor Feng suggests completing the second objective first, then extending the model to the rest of the topics and finding the influence/result of machine learning models.

Regular meeting time with supervisor Feng fortnightly, from 11:00 am to 12:00 pm, at Feng’s office (room 108 Old Geology)

2023-03-23 Team Meeting notes

📅 Date

Apr 23, 2023

👥 Participants

Team: @rgiri @xiangyih @Sunchuangyu HUANG @Jiaqi Hu @jihangy

🥅 Goals

- Decide the client meeting time

🗣 Discussion topics

Time	Item	Presenter	Notes
			•

✅ Action items

- [x] Each team member updates their Linkedin Profile before Jan 16, 2023
- [x] Gather information about the Client and CSIRO
- [x] Create an Elevator Pitch (self-introduction) of about 100 words
- [x] Send first meeting invitation to the client @xiangyih

⤴ Decisions

2023-03-05 Team Meeting notes

📅 Date

Mar 5, 2023

👥 Participants

Team: @xiangyih @Sunchuangyu HUANG @rgiri @Jiaqi Hu

🥅 Goals

1. Discuss the project preferences of individual team members
2. Develop a spreadsheet with the reasoning on why we should choose the particular project
3. Fill out the Project Selection form according to the mutually decided projects
4. Set up the information-sharing platforms (Google Drive, LMS Group, Discord, Slack)
5. Discuss the unavailability of the 5th Team Member @jihangy

🗣 Discussion topics

Time	Item	Presenter	Notes
			•

✅ Action items

- [x] Set up google drive, LMS Group introduction page, and discord @Sunchuangyu HUANG
- [x] Set up slack account @rgiri
- [x] Analyze and select individual project preferences @xiangyih @Sunchuangyu HUANG @rgiri @Jiaqi Hu
- [x] Fill out the project selection survey @Sunchuangyu HUANG
- [x] Submit the deed agreement @xiangyih @Sunchuangyu HUANG @rgiri @Jiaqi Hu @jihangy
- [x] Follow up on @jihangy

⤴ Decisions

CSIRO: Commonwealth Scientific and Industrial Research Organisation - chatGPT

CSIRO is an Australian government agency that is responsible for scientific research to benefit the nation and promote innovation.

CSIRO conducts research across various fields including agriculture, health, energy, mining, and manufacturing, with a focus on developing practical solutions to real-world problems.

In terms of corporate social responsibility, CSIRO is committed to operating in an ethical and sustainable manner. This includes reducing its *environmental impact*, promoting diversity and inclusion, and ensuring that its research is sued to benefit society.

CSIRO also works to engage with the broader community, including through education and outreach programs, to promote awareness of and understanding of science and its applications.

History of CSIRO

CSIRO was founded in 1916 as the Advisory Council of Science and Industry, with the aim of applying scientific research to support Australian economic development. In 1926, the organization was restructured and renamed the Council for Science and Industrial Research (CSIR).

Initially, CSIR focused on agricultural research, but over time it expanded into other areas such as chemistry, physics, and engineering. In 1949, CSIR was renamed again as the Commonwealth Scientific and Industrial Research Organisation (CSIRO) to reflect its growing mandate and scope of activities.

Over the years, CSIRO has made significant contributions to Australian industry and society. For example, it played a key role in the development of the polymer banknote, a revolutionary new type of currency that is now used in several countries around the world.

CSIRO has also been involved in major scientific discoveries, such as the identification of the cause of the “dieback” disease affecting eucalyptus trees in the 1970s, and the development of Wi-Fi technology in the 1990s.

Today, CSIRO remains a leading scientific research organization in Australia, with a mission to use science and innovation to address the country’s most pressing challenges and opportunities.

CSIRO contributions to climate change / climate crisis

CSIRO has been a significant contributor to climate change research and policy development in Australia and globally. Some of its key contributions in this area include:

1. Climate Modeling: CSIRO’s climate scientists have developed some of the world’s most sophisticated climate models to better understand how the climate system works and how it might change in the future. This includes models that project future climate change scenarios based on different greenhouse gas emissions scenarios.
2. Carbon Accounting: CSIRO has developed methods for measuring and accounting for carbon emissions from various sources, including forests, land use, and energy production. This has helped to inform policy decisions around carbon pricing and emissions reduction targets.
3. Adaptation Strategies: CSIRO has worked with communities and governments to develop adaptation strategies to help people and ecosystems cope with the impacts of climate change. This includes research on topics such as water management, coastal erosion, and biodiversity conservation.
4. Renewable Energy: CSIRO has conducted research on a wide range of renewable energy technologies, including solar, wind, and bioenergy. This research has helped to support the development of renewable energy in Australia and around the world.
5. Climate Policy Advice: CSIRO has provided advice to governments and policymakers on climate change mitigation and adaptation strategies. This includes advice on topics such as emission reduction targets, renewable energy policies, and carbon pricing mechanisms.

CSIRO contributions to agriculture

CSIRO has a long history of research and innovation in agriculture and has made significant contributions to the development of Australia's agricultural industry. Some of its key contributions in this area include;

1. Crop Improvement: CSIRO has conducted extensive research on plant genetics, breeding, and biotechnology to develop new crop varieties that are more productive, resilient, and adapted to Australian conditions. This includes the development of wheat and barley varieties tha re resistant to disease and drought.
2. Soli and Water Management: CSIRO has developed technologies and practices to improve soil and water management in agriculture. This includes research on topics such as soil carbon sequestration, water use efficiency, and precision agriculture.
3. Biosecurity: CSIRO has developed technologies and strategies to manage pests and diseases in agriculture, including the development of new methods for detecting and controlling invasive species and diseases.
4. Sustainable Agriculture: CSIRO has conducted research on sustainable agriculture practices while maintaining or improving productivity. This includes research on topics such as conservation agriculture, agroforestry, and integrated pest management.
5. Food Security: CSIRO has conducted research on food security issues, including the development of technologies and strategies to improve food production and distribution, reduce food waste, and increase access to nutritious food.

Some readings we may interest in

Climate Risk and Climate Change:

"Climate Change in Australia" report: https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/Climate-change-in-Australia

This report provides a comprehensive analysis of the current and future impacts of climate change in Australia, including changes to temperature, rainfall, sea level, and extreme weather events. It also explores the potential impacts on Australia's natural resources, ecosystems, and human communities.

"Climate Change Risks to Australia's Coasts" report: https://www.csiro.au/en/Research/OandA/Areas/Coasts-and-oceans/Climate-change-risks-to-Australias-coasts

This report examines the potential risks of climate change to Australia's coastal communities and infrastructure, including sea-level rise, storm surges, and coastal erosion. It provides recommendations for coastal planning and management to adapt to these risks.

"Climate Projections for Australia": Redirect Climate Projections

This report provides climate projections for Australia based on climate modeling. It includes projections for changes in temperature, rainfall, and extreme weather events, and can help inform decision-making for climate adaptation and mitigation.

Food Security and Agriculture:

"Future of Food" report: https://www.csiro.au/en/Research/OandA/Areas/Food-and-Agriculture/Future-of-food

This report explores the future of food production and consumption in Australia, including the challenges and opportunities related to food security, sustainability, and nutrition. It includes analysis of current trends and future scenarios for food production, distribution, and consumption, and provides recommendations for policy and investment.

"Food Loss and Waste in Australia" report: https://www.csiro.au/en/Research/OandA/Areas/Food-and-Agriculture/Food-loss-and-waste-in-Australia

This report examines the extent and causes of food loss and waste in Australia, and provides recommendations for reducing this problem. It includes analysis of the economic, social, and environmental costs of food waste, and suggests strategies for prevention and recovery.

"Agriculture and Food Security in a Changing Climate": https://www.csiro.au/en/Research/OandA/Areas/Food-and-Agriculture/Agriculture-and-food-security-in-a-changing-climate

This research project examines the impacts of climate change on global food security and identifies strategies for adaptation and mitigation. It includes analysis of the potential impacts on agricultural productivity, food distribution, and nutrition, and suggests strategies for building resilience in food systems.

Population:

"Population Futures for Australia": https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/Climate-change-in-Australia

This report provides projections for population trends in Australia, including demographic changes and their implications for social, economic, and environmental sustainability. It includes analysis of fertility, mortality, migration, and urbanization trends, and provides recommendations for policy and planning to manage these changes.

"Australia's Demographic Challenges" report: https://www.csiro.au/en/Research/OandA/Areas/Population-and-communities/Australias-demographic-challenges

This report examines the challenges and opportunities related to Australia's ageing population, including the implications for health care, workforce participation, and social cohesion. It includes analysis of demographic trends, economic impacts, and policy responses.

Economics:

"Economic Impacts of Climate Change": https://www.csiro.au/en/Research/OandA/Areas/Assessing-our-climate/Economic-impacts-of-climate-change

This report analyzes the potential economic impacts of climate change on Australia, including the costs of extreme weather events, sea-level rise, and changes in agricultural productivity. It includes analysis of the potential impacts on sectors such as tourism, energy, and infrastructure, and suggests strategies for adaptation and mitigation.

"Australian National Outlook" report: https://www.csiro.au/en/Research/National-Outlook

This report provides a long-term outlook for the Australian economy, society, and environment, and identifies strategies for achieving a sustainable and prosperous future. It includes analysis of trends and scenarios for key sectors such as energy, water, land use, and innovation, and provides recommendations for policy and investment.

Some People we may interest in

Topic	Researcher	Expertise	Recent Publications
Agriculture	Dr. Michael Robertson	Sustainable agricultural systems, crop modeling, climate adaptation	"The global land rush: can it yield sustainable results?", "Sustainable intensification: a case study in Australian grain farming"
Agriculture	Dr. Lisa Lobry de Bruyn	Soil biology, soil health, microbial communities	"Incorporating soil microbial diversity metrics into soil health assessments", "Plant root-microbe communication in shaping root microbiomes"
Population	Dr. Tom Wilson	Demography, population health, epidemiology	"Dementia in Australia: Nature, Prevalence and Burden", "Population Ageing, Health and Aged Care: A Data Booklet"

Economics	Dr. Paul Graham	Environmental economics, resource management, scenario planning	"Scenario Planning in a Post-Pandemic World", "Natural Resource Management for Sustainable Development"
Climate Risk	Dr. Asif Gill	Climate modeling, extreme weather events, climate adaptation	"Climate Change and Extreme Weather Events: Implications for Disaster Management in Australia", "Assessing the sensitivity of Australian climate extremes to variations in regional warming"

Team Contact information

Name		Email	LinkedIn
Ritwik Giri	RG	rgiri@student.unimelb.edu.au	https://www.linkedin.com/in/ritwik-giri/
Xiangyi He	Emma	xiangyih1@student.unimelb.edu.au	https://www.linkedin.com/in/xiangyi-he-7a15a2269/
Jiaqi Hu	Nick	jiaqih7@student.unimelb.edu.au	https://www.linkedin.com/in/hu-jiaqi-498578202/
Sunchuangyu Huang	Rin	sunchuangyuh@student.unimelb.edu.au	https://www.linkedin.com/in/rin-huang/
Jihang Yu	Jonathon	jihangy@student.unimelb.edu.au	https://www.linkedin.com/in/jihang-yu-1022151a5/

Team Individual Introduction

Ritwit Giri

I am a final year Master of Data Science student at the University of Melbourne. My journey so far in the Master's course has been an astonishing one. My course subjects helped me to gain both practical and theoretical perspectives of the different components involved in the Data Science course. Being an international student, with skills in machine learning, data visualization, problem-solving, programming, and creative thinking, I have demonstrated success in building solutions to real-world business problems. I have three years of experience at Deloitte and MatrixCare.

I believe this data science project will expand my limits beyond what I will learn in my Master's degree. It will provide me with a much-needed leap, which will allow me to become a part of an Australian workforce and contribute effectively and efficiently.

Xiangyi He

This is Xiangyi He, you can also call me Emma. I'm in my first year of the Master of Data Science in Unimelb. I am a Chinese international student with a Bachelor's degree in the same field and have gained in programming languages such as R and Python during my three years of undergraduate studies.

My passion for data analysis and management led me to pursue a higher degree in this field, where I aim to gain deeper knowledge and skills in data science, machine learning, and data visualization. I am a fast learner and enjoy challenging myself to expand my knowledge base and explore new technologies. I hope I can be able to develop my research skills and make meaningful contributions to the field of Big Data.

Jiaqi Hu

I am Jiaqi Hu, you can also call me Nick. I am in my final year of the Master of Data Science degree in the University of Melbourne. A claim has been planted in my head during all these years of studying statistics and computer science subjects ever since I got into my university when I was a freshman, that is practice is the best way to learn. The best way to understand some new knowledge is to apply them to a problem by myself, rather than watching the professor presenting it in front of us students one thousand times.

I have been using statistical models in Python and R to analyze data and make predictions and inferences since my undergraduate. It has been thrilling to me to have a chance to apply what I have learned in these years, such as how to preprocess the data, what should be included when writing a report about the understanding of the data, and what to consider when choosing models to build on the data, etc., to a real industry problem and hope to create some meaningful results in this project.

Jihang Yu

Sunchuangyu Huang

From the first day at unimelb, I have been passionate about data and set my sights on a career as a data engineer. I drove by a critical mind to find insights from complex data, build powerful products and provide data-driven solutions to clients.

I achieve this via:

1. learn essential knowledge for instance statistical modeling, machine learning, etc.
2. apply the knowledge in the real-world setting to see the impact in action.

Before pursuing my master's degree, I had the opportunity to participate in several industrial projects. One project involved working with the CSL lab to drive digital transformation. To succeed in this endeavor, our team drew upon our collective knowledge and skills, conducting thorough exploratory data analyses and creating effective visualizations. What's more, we were able to stay agile and adapt to new machine-learning methods as needed.

Based on my past experience, I am confident that the skills and knowledge I have developed, make me a strong asset for any data-driven endeavor. I am excited to work with CSRIO and make a positive impact on future climate risks.

Team Individual Introduction Table

Ritwit Giri	I am a final year Master of Data Science student at the University of Melbourne. My journey so far in the Master's course has been an astonishing one. My course subjects helped me to gain both practical and theoretical perspectives of the different components involved in the Data Science course. Being an international student, with skills in machine learning, data visualization, problem-solving, programming, and creative thinking, I have demonstrated success in building solutions to real-world business problems. I have three years of experience at Deloitte and MatrixCare. I believe this data science project will expand my limits beyond what I will learn in my Master's degree. It will provide me with a much-needed leap, which will allow me to become a part of an Australian workforce and contribute effectively and efficiently.

Xiangyi He	This is Xiangyi He, you can also call me Emma. I'm in my first year of the Master of Data Science in Unimelb. I am a Chinese international student with a Bachelor's degree in the same field and have gained in programming languages such as R and Python during my three years of undergraduate studies. My passion for data analysis and management led me to pursue a higher degree in this field, where I aim to gain deeper knowledge and skills in data science, machine learning, and data visualization. I am a fast learner and enjoy challenging myself to expand my knowledge base and explore new technologies. I hope I can be able to develop my research skills and make meaningful contributions to the field of Big Data.
Jiaqi Hu	I am Jiaqi Hu, you can also call me Nick. I am in my final year of the Master of Data Science degree in the University of Melbourne. A claim has been planted in my head during all these years of studying statistics and computer science subjects ever since I got into my university when I was a freshman, that is practice is the best way to learn. The best way to understand some new knowledge is to apply them to a problem by myself, rather than watching the professor presenting it in front of us students one thousand times. I have been using statistical models in Python and R to analyze data and make predictions and inferences since undergraduate. It has been thrilling to me to have a chance to apply what I have learned in these years, such as how to preprocess the data, what should be included when writing a report about the understanding of the data, and what to consider when choosing models to build on the data, etc., to a real industry problem and hope to create some meaningful results in this project.
Jihang Yu	
Sunchuangyu Huang	Based on my past experience, I am confident that the skills and knowledge I have developed, make me a strong asset for any data-driven endeavour. I am excited to work with CSRIO and make a positive impact on future climate risks.

Terminology

Representattive Concentration Pathway (RCPs)

W Representative Concentration Pathway

Representative Concentration Pathways (RCPs) are a set of scenarios developed by the Intergovernmental Panel on Climate Change (IPCC) to project future greenhouse gas concentrations and associated climate change impacts. RCPs provide a standardized framework for climate models to evaluate the potential impacts of different levels of greenhouse gas emissions.

There are four RCPs, ranging from RCP 2.6 to RCP 8.5, which represent a range of possible greenhouse gas concentration pathways. RCP 2.6 represents a low-emissions scenario, in which emissions peak in the next few years and then decline rapidly, leading to a global temperature increase of 1.5°C by 2100. RCP 8.5 represents a high-emissions scenario, in which emissions continue to increase throughout the century, leading to a global temperature increase of 4.5°C by 2100.

- **RCP 1.9** is a pathway that limits global warming to below 1.5 °C, the aspirational goal of the Paris Agreement
- **RCP 2.6** is a "very stringent" pathway. According to the IPCC, RCP 2.6 requires that carbon dioxide (CO_2) emissions start declining by 2020 and go to zero by 2100
- **RCP 4.5** is described by the IPCC as an intermediate scenario. Emissions in RCP 4.5 peak around 2040, then decline
- **RCP 6**, emissions peak around 2080, then decline
- **RCP 7** is a baseline outcome rather than a mitigation target
- **RCP 8.5** emissions continue to rise throughout the 21st century

AR5 global warming increase (°C) projections[24]

Scenario	2046–2065	2081–2100
	Mean (*likely* range)	Mean (*likely* range)
RCP2.6	1.0 (0.4 to 1.6)	1.0 (0.3 to 1.7)
RCP4.5	1.4 (0.9 to 2.0)	1.8 (1.1 to 2.6)
RCP6	1.3 (0.8 to 1.8)	2.2 (1.4 to 3.1)
RCP8.5	2.0 (1.4 to 2.6)	3.7 (2.6 to 4.8)

Across all RCPs, global mean temperature is projected to rise by 0.3 to 4.8 °C by the late 21st century.

According to a 2021 study in which plausible AR5 and SSP scenarios of CO_2 emissions are selected,[12]

AR5 and SSP Scenarios and temperature change projections

SSP Scenario	Range of Global Mean Temperature Increase (Celsius) – 2100 from pre-Industrial baseline
RCP 1.9	~1 to ~1.5
RCP 2.6	~1.5 to ~2
RCP 3.4	~2 to ~2.4
RCP 4.5	~2.5 to ~3
RCP 6.0	~3 to ~3.5
RCP 7.5	~4
RCP 8.5	~5

AR5 global mean sea level (m) increase projections[24]

Scenario	2046–2065	2081–2100
	Mean (*likely* range)	Mean (*likely* range)
RCP2.6	0.24 (0.17 to 0.32)	0.40 (0.26 to 0.55)
RCP4.5	0.26 (0.19 to 0.33)	0.47 (0.32 to 0.63)
RCP6	0.25 (0.18 to 0.32)	0.48 (0.33 to 0.63)
RCP8.5	0.30 (0.22 to 0.38)	0.63 (0.45 to 0.82)

Across all RCPs, global mean sea level is projected to rise by 0.26 to 0.82 m by the late-21st century.

The RCPs are based on various assumptions about future trends in population growth, economic development, energy use, and land-use change. These assumptions are used to develop scenarios of greenhouse gas emissions and concentrations, which are then used to project future climate change impacts such as sea level rise, changes in precipitation patterns, and increases in the frequency and intensity of extreme weather events.

The RCPs are widely used in climate science and policy research to evaluate the potential impacts of different levels of greenhouse gas emissions on the climate system and to inform decision-making about mitigation and adaptation strategies.

Shared Socioeconomic Pathways

W Shared Socioeconomic Pathways

Shared Socioeconomic Pathways (SSPs) are a set of scenarios developed by the IPCC that describe plausible future socioeconomic developments and their associated greenhouse gas emissions. SSPs are designed to be used in conjunction with the RCPs to provide a comprehensive framework for exploring future climate change and its impacts.

There are five SSPs, ranging from SSP1 to SSP5, which represent a range of possible future socioeconomic and environmental conditions. SSP1 represents a sustainable development scenario, in which there is a strong focus on social and environmental sustainability, while SSP5 represents a fossil-fuel-driven development scenario, in which there is a continued reliance on fossil fuels and a lack of attention to social and environmental issues.

- SSP1: Sustainability (Taking the Green Road)

- SSP2: Middle of the Road
- SSP3: Regional Rivalry (A Rocky Road)
- SSP4: Inequality (A Road Divided)
- SSP5: Fossil-fueled Development (Taking the Highway) [6]

They have been used to help produce the IPCC Sixth Assessment Report on climate change, published on 9 August 2021.[7]

The SSPs are based on assumptions about future trends in population growth, economic development, technological innovation, governance, and environmental policy. These assumptions are used to develop scenarios of socioeconomic development, which are then used to project future greenhouse gas emissions and associated climate change impacts.

The SSPs are intended to provide a common framework for researchers to explore the potential impacts of different socioeconomic and environmental conditions on the climate system, and to inform decision-making about mitigation and adaptation strategies. By combining the SSPs with the RCPs, researchers can explore a wide range of possible future scenarios and evaluate the potential effectiveness of different policy interventions in reducing greenhouse gas emissions and limiting the impacts of climate change.

SSP	Scenario	Estimated warming (2041–2060)	Estimated warming (2081–2100)	Very likely range in °C (2081–2100)
SSP1-1.9	very low GHG emissions: CO_2 emissions cut to net zero around 2050	1.6 °C	1.4 °C	1.0 – 1.8
SSP1-2.6	low GHG emissions: CO_2 emissions cut to net zero around 2075	1.7 °C	1.8 °C	1.3 – 2.4
SSP2-4.5	intermediate GHG emissions: CO_2 emissions around current levels until 2050, then falling but not reaching net zero by 2100	2.0 °C	2.7 °C	2.1 – 3.5
SSP3-7.0	high GHG emissions: CO_2 emissions double by 2100	2.1 °C	3.6 °C	2.8 – 4.6
SSP5-8.5	very high GHG emissions: CO_2 emissions triple by 2075	2.4 °C	4.4 °C	3.3 – 5.7

Time Series

What is Time Series Analysis?

A time series is a sequence of data points or observations collected over time. These data points are usually taken at regular intervals and are ordered in chronological order. Time series analysis involves examining patterns, trends, and changes over time to understand and predict future behavior.

Time series data can be found in various fields, such as finance, economics, engineering, social sciences, and environmental studies. Examples of time series data include stock prices, weather data, population growth rates, and sales figures. Time series analysis techniques, such as trend analysis, seasonal analysis, and forecasting, are used to extract meaningful insights and make predictions based on the patterns observed in the data.

The time series contains Four components:

1. Trend
2. Seasonality: the repeating pattern of data over a set period of time
3. Repeating non-seasoning data
4. Variations: unpredictable trends cannot be explained by other components

Forecast Model:

- ARIMA: auto-regressive integrated moving average
- Exponential Smoothing: forecast model doesn’t have specific trends

Climate Measurements and Standard Initiative (CMSI)

Climate Measurement Standards Initiative (CMSI) | Home

Summary Report

- CMSI - Executive summary
- CMSI - Technical Summary

Climate measurements are essential for understanding the state of the planet's climate and tracking changes over time. There are various types of climate measurements, including temperature, precipitation, humidity, atmospheric pressure, wind speed and direction, and carbon dioxide concentration.

To ensure that climate measurements are consistent and reliable across different locations and over time, there are various standards initiatives that have been developed. These initiatives establish guidelines for measuring and reporting climate data, ensuring that the data collected is accurate, comparable, and reliable.

- GCOS - Global Climate Observing System
- WMO - World Meteorological Organization
- IOC - Intergovernmental Oceanographic Commission
- ICSU - International Council for Science
- UNEP - United Nations Environment Programme

CMSI - Technical Summary

https://uploads-ssl.webflow.com/5f1bdaf710347301b0c01fd4/5f5c2f4cb000cab9c03025d8_CMSI - Climate Science Technical Summary.pdf

There are 3 factors that affect uncertainty in future climate scenarios:

1. Ongoing natural climate variability
2. Global socioeconomic development results in emissions of greenhouse gases and aerosols
3. Ongoing natural variability

Climate change in Australia Reports

The climate science report presents projected hanges for two categories of climate harzards that can damage buildings and infrastructure :

1. Acute - extreme weather evenst identified by the CMSI science committee associated with building and infrastructure damage.
2. Chronic - gradually emerging aspects of climate risk, including changes to annual average temperature, rainfall and sea level, time in drought and days over 35 degrees.

Gaps and needs

Projections for some harzards have wide ranges of plausible changes and low confidence levels. The scientisifc needs include:

- improved understanding of physical processes associated with climatc harzards, especially bushfires, thunderstorms, floods, hail, tropical cyclones, east coast lows, storm surges and drought
- evaluation of present-day climate model harzard simulartions with a focus on extreme hazards that cause damage
- examination of compound events (occurring at the same time or in close succession) and cascading events where one harzard or impact triggers another (e.g. a bushfire triggers water quality and security issues in the months and years following the fire)
- nationally consistent high-resolution projections (rather than the inconsistent high-resolution projects for regional domains currently available), so that regional insights from high-resolution modelling are compareable across the country
- use of up-to-date climate models - projections overwhelmingly use the CMIP3/CMIP5 generation of international models, which are 7-15 years old, but CMIP6 models are now emerging.

CMSI - Executive summary

https://uploads-ssl.webflow.com/5f1bdaf710347301b0c01fd4/5f5c2f4cc0d8e1f3b153e99c_CMSI - Summary for Executives.pdf

> Task Force on Climate-Related Financial Disclosures (TCFD) has developed a set of recoomendations that companies can use to disclose information about climate risks. The disclosures include physical, liability and transition risks.
>
> The TCFD recognises that incorporating scenario analysis into strategic planning processes will improve over time as organizations 'learn by doing'. To facilitate progress, TCFD encourages the following:
>
> ***- further developing 2°C or lower transition scenarios that can be applied to specific industries and locations along with supporting outputs, tools, and user interfaces***
>
> - developing broadly accepted methodologies, data sets, and tools for scenario-based evaluation of physical risk by organisations
>
> - making these data sets and tools publicly available to facilitate use by organisations, reduce costs, minimise expertise differences between jurisdictions, enhance comparability of climate-related risk assessments, and help ensure comparability for investors
>
> - creating more industry specific (financial and non-financial) guidance for preparers and users of climate-related scenarios.

CMSI aims to develop opensource standards and guidelines that will provide companies with:

- a consistent approach to disclosures under the TCFD (Task Force on Climate-Related Financial Disclosures), so that disclosures
- increased confidence in disclosures, as the standards will be supported by science from Australia's leading climate experts
- a potential framework should regulators decide to mandate disclosures
- a roadmap for future research and development aligned with disclosure requirements
- guidance on issues relevant to specific industries
- guidance that will allow smaller companies to disclose their climate change scenarios
- lower likelihood of unintentionally adopting non-standard approaches
- modelling requirements that support provisions of transparent advice in the industry

Table ES1: Scope of CMSI Phase 1

	In scope	Likely future scope
Purpose	• Disclosure of TCFD scenario analyses	• Stress testing and vulnerability testing • Other types of analysis
Analysis	• Scenario specification	• Stress testing including compound events • Sensitivity analysis • Modelling exposure changes • Modelling vulnerability changes • Developing data sets
Climate-related risks	**Physical risk**	**Transition risk**
Hazards	**Acute physical risks:** • Tropical cyclones • East coast lows • Extreme rainfall and riverine floods • Extreme sea level events • Large hail • Extreme bushfire events **Chronic physical risks:** • Average temperature and extreme heat events • Average rainfall • Sea level rise • Drought	**Acute physical risks:** • Storm surge and coastal flooding **Transition risks:** • Technology • Policy and legal • Market • Reputation
Impacts	• Damage to property (buildings and infrastructure)	**Physical risks:** • Loss of use of asset • Loss due to cross-dependency on other assets • Health and human impacts • Agriculture and other sectors Macroeconomic impacts from both physical and transition risks

Three factors determine the climate, and hence physical climate risks, in future scenarios:

- *ongoing climate variability*
- *global socio-economic development and resulting emissions of greenhouse gases and aerosols*
- *regional climate responses to these emissions*

Using the lowest and highest international-standard scenarios for atmospheric greenhouse gas concentrations, known as *representative concentration pathways (RCPs)* provides the range of future climate possibilities, and meets international TCFD guidelines.

There is an emerging system of *socio-economic pathways (SSPs)* that can be related to different RCPs.

The climate science report presents projected changes for 2 categories of climate hazards that can damage buildings and infrastructure:

1. Acute - extreme weather events including changes n tropical cyclones, east coast lows, extreme rainfall, hail, storm surges and fire weather, with confidence ratings presented for RCP 2.6 and RCP 8.5 for 2030, 2050, and 2090.
2. Chronic - gradually emerging aspects of climate risk, including changes to annual average temperature, rainfall and sea level, time in drought and days over 35 degrees. These are also presented for RCP 2.6 and RCP 8.5 for 2030, 2050 and 2090.

Chronic physical risks, such as sea-level rise and temperature increases, are likely to exacerbate losses from acute physical risks such as tropical cyclones, storms, flooding, coastal inundation and erosion, as well as heat waves and drought-caused soil contraction. These risks, and the compounded multiple hazards, will increase with climate change.

Figure ES1: Projected changes in climate hazards that influence physical risks for Australian buildings and infrastructure. Confidence estimates are provided in parentheses.

Scenario analysis of climate-related physical risk for buildings and infrastructure: Financial disclosure guidelines & climate science guidance

CMSI - Technical Summary

Short-term actions

- Collect feedback on CMSI guidelines and update them, including the use of the most recent climate observations and projections.
- Compile and quality-check observations of climate hazards, including their extremes.
- Evaluate present-day simulations of extreme climate hazards.
- Analyze future changes in extreme climate hazards.
- Perform Coupled Model inter-comparison Project phase 6 regional climate projects for Australia.

Longer-term actions

- Improve observation systems for extreme climatic hazards
- Improve understanding and modelling of extreme climatic hazards
- Assess likely future changes in compound events

- Implement climate services that support uptake of hazard projections in impact assessment, planning and action.

GCOS - Global Climate Observing System

The Global Climate Observing System (GCOS) is an international program that was established in 1992.

GCOS focuses on the following main objectives:

1. Define and coordinate global climate observation requirements.
2. Develop and implement strategies to design and maintain climate-observing networks and systems.
3. Advocate for improved access to, and utilization of, climate data and information.
4. Facilitate the development of climate services that support national and international climate policy and decision-making.

WMO - World Meteorological Organization

The World Meteorological Organization (WMO) is an intergovernmental organization established in 1950 and became a specialized agency of the United Nations in 1951. Its primary objective is to promote international cooperation in the field of meteorology, hydrology and related geophysical sciences.

The WMO plays a crucial role is facilitating the exchange of meteorological data and information among its members, improving the accuracy of weather forecasts, and promoting research and development in meteorology and related fields. Its main activities include:

1. Standardization: The WMO sets standards and protocols for meteorological and related observations, ensuring that data is accurate, reliable, and comparable across different countries and regions.
2. Data exchange: The organization facilitates the global exchange of meteorological, climatological, and hydrological data through its World Weather Global Telecommunication System, and the Global Data-Processing and Forecasting System.
3. Research and development: The WMO research initiatives aimed at improving the understanding of Earth's atmosphere and climate, enhancing weather and climate prediction capabilities, and developing new applications for meteorological data.
4. Capacity building: The organization provides technical assistance, training, and education to its members, helping to improve meteorological and hydrological services and capacities.
5. Public awareness: The WMO works to raise awareness about the importance of meteorology and related sciences in everyday life, as well as their relevance to addressing global challenges like climate change and natural disasters.
6. Coordination and collaboration: The WMO collaborates with other international organizations, such as the IOC - Intergovernmental Oceanographic Commission, UNEP - United Nations Environment Programme, and the ISC - International Science Council, on initiatives related to climate, water, and environmental issues.

ISC - International Science Council

The International Science Council (ISC) is a non-governmental organization that was established in 2018 through the merger of two pre-existing organizations: the International Council for Science (ICSU) and the International Social Science Council (ISSC).

The main objectives of the ISC are to:

1. Promote international scientific cooperation and collaboration across disciplines, countries, and regions.
2. Advocate for the importance of science in addressing global challenges and informing policy decisions.
3. Support the development of scientific capacity and the exchange of knowledge and best practices among its members.
4. Foster scientific integrity, responsibility, and ethics in the conduct of research.

The ISC brings together a diverse range of scientific organizations, including international scientific unions and associations, national academies, research councils, and other scientific organizations from more than 140 countries. By facilitating collaboration and dialogue among its members, the ISC aims to strengthen the role of science in society and contribute to addressing global challenges such as climate change, sustainable development, and disaster risk reduction.

IOC - Intergovernmental Oceanographic Commission

The Intergovernmental Oceanographic Commission (IOC) of UNESCO is an international organization that was established in 1960 to promote scientific research, services, and capacity-building in the field of oceanography

The main objectives of the IOC are to:

1. Foster international cooperation and collaboration in marine scientific research to enhance our understanding of the oceans, their resources, and the processes that govern their behaviour.
2. Develop and coordinate global ocean observation systems to improve our ability to predict and monitor ocean-related phenomena, such as tsunamis, storm surges, and climate change.
3. Support the development and application of marine scientific knowledge to inform policy decisions and promote the sustainable management of marine resources.
4. Strengthen the capacity of its member states, particularly developing countries, to conduct oceanographic research, implement ocean observation systems and use marine scientific knowledge for decision-making.

UNEP - United Nations Environment Programme

The United Nations Environment Programme (UNEP) is the leading global environmental authority within the United Nations system. Established in 1972, UNEP's mission is to provide leadership and encourage partnership in caring for the environment by inspiring, informing, and enabling nations and peoples to improve their quality of life without compromising that of future generations.

UNEP's primary objectives include:

1. Assessing global, regional, and national environmental conditions and trends.
2. Developing international and national environmental policies and instruments, including conventions, protocols, and guidelines.
3. Strengthening institutional capacities to address environmental issues at all levels of governance.
4. Facilitating the coordination of UN environmental activities and providing support to relevant institutions and organizations.
5. Promoting the integration of environmental considerations into social and economic policies and planning processes.
6. Encouraging and supporting environmental education, public awareness, and participation in sustainable development.

ICSU - International Council for Science

The International Council for Science (ICSU) was a non-governmental organization established in 1931 to promote international scientific cooperation and research. It was dedicated to advancing scientific knowledge, supporting scientific collaboration, and advocating for the use of science in the development of public policy. In 2018, ICSU merged with the International Social Science Council (ISSC) to form the International Science Council (ISC).

Before the merger, ICSU had a broad membership base, including national scientific organizations, research councils, and international scientific unions. Its objectives were to:

1. Encourage and facilitate interdisciplinary and international scientific cooperation.
2. Promote the development of scientific capacity, particularly in developing countries.
3. Advocate for the use of science in policy-making and the development of evidence-based solutions to global challenges.
4. Support scientific integrity, responsibility, and ethics in the conduct of research.

Granger Causality

Before understanding Granger causality, we need to understand what the term causality means in time series analysis.

Causality in time series refers to the concept of understanding the cause-and-effect relationship between variables in a time-ordered dataset. In time series analysis, **the goal is often to find if changes in one variable lead to changes in another mechanism**, enhancing predictive models, and making informed decisions on those relationships.

However, establishing causality in time series data can be challenging due to several reasons.

1. Correlation vs. causation: Correlation between two time series does not necessarily imply causation. It's essential to differentiate between mere associations and true cause-and-effect relationships.
2. Confounding variables: Unobserved or unmeasured variables might be affecting both the variables under consideration, leading to a spurious relationship.
3. Reverse causality: The direction of causality may be the opposite of what is initially assumed, i.e., variable Y may be causing changes in variable X, rather than the other way around.
4. Feedback loops: The relationship between variables might be bidirectional, where both variables influence each other over time.

W Granger causality

Granger causality is a statistical hypothesis test used to determine whether one time series can predict another time series, based on the premise that if a variable X Granger causes variable Y, then the past value of X will contain information that helps predict Y.

Granger causality is not the same as the traditional notion of causality in the sense that it doesn’t establish a cause-and-effect relationship. Instead, it tests whether the past values of one time series can be used to improve forecasts of another time series. Granger causality relies on linear regression models and lags of the time series data to determine whether one variable has predictive poewr over the other:

To perform a Granger causality test, you would:

1. Estimate a regression model of the current value of Y, using its own past values (lags) and the past values (lags) of X.
2. Estimate another regression model of the current value of Y, using only its own past values (lags).
3. Compare the goodness of fit of these two models, typically by using an F-test or a likelihood ratio test.

Including the past values of X significantly improves the prediction of Y, then X is said to “Granger-cause“ Y. It’s important to note that Granger causality is sensitive to the choice of the lag length, and it assumes that the relationship between the variables is linear and time-invariant.

Data Processing

- Raw Data Descriptions

Raw Data Descriptions

From initial observation, data in all datasets are monthly data.

- jra55: date: 1958-01-01 to 2018-12-01
- nnr1: date: 1948-01-01 to 2020-05-01
- CMOHistoricalDataMonthly: 1960-01-01 to 2020-02-01

jra55

	jra55.AO	**jra55.IOD**	**jra55.MEI**	**jra55.NHTELE**	**jra55.PNA**	**jra55.PSA**	**jra55.SAM**
n_modes	20	10	24	10	10	20	20
eofs_start_date	19790101	19790101	19790101	19790101	19790101	19790101	19790101
eofs_end_date	20011230	20011230	20011230	20011230	20011230	20011230	20011230
base_period_start	19790101	19790101	19790101	19790101	19790101	19790101	19790101
base_period_end	20011230	20011230	20011230	20011230	20011230	20011230	20011230
calendar	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian

nnr1

	nnr1.AO	**nnr1.IOD**	**nnr1.MEI**	**nnr1.NHTELE**	**nnr1.PNA**	**nnr1.PSA**	**nnr1.SAM**
n_modes	20	10	24	10	10	20	20
eofs_start_date	19790101	19790101	19790101	19790101	19790101	19790101	19790101
eofs_end_date	20011230	20011230	20011230	20011230	20011230	20011230	20011230
base_period_start	19790101	19790101	19790101	19790101	19790101	19790101	19790101
base_period_end	20011230	20011230	20011230	20011230	20011230	20011230	20011230
calendar	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian	proleptic_gregorian

World bank CMOHistoricalDataMonthly

World Bank Commodity Price Data (The Pink Sheet)
monthly prices in nominal US dollars, 1960 to present
(monthly series are available only in nominal US dollars)
Updated on March 03, 2020

Readings

- Forecasting commodity returns by exploiting climate model forecasts of El Niño Southern Oscillation
- Causal inference in complex multiscale systems
- Vassili Kitsios - Influence of Climate Variability on Financial Markets & Health Sector
- Dynamic Bayesian Networks for Evaluation of Granger Causal Relationships in Climate Reanalyses
- Economic forecasting in agriculture
- Can exchange rates forecast commodity prices?
- Forecasting the Yield and the Price of Cotton

Forecasting commodity returns by exploiting climate model forecasts of El Niño Southern Oscillation

Author: Vassili Kitsios, Lurton De Mello, and Richard Matear.

In reality, both human and geophysical systems are complex and vastly multiscale, comprising timescales ranging from seconds to centuries.

A historical study of prices and yields of agricultural commodities.

Numerous studies focused on macroeconomic factors more when building a forecasting commodity model.

When considering the climate change factor or climate projection factors, the system is primarily focuses on specified boundary conditions of CO2 and other forcings associated with prescribed scenarios of future economic development.

> Initial condition is a crucial factor when contributing to forecast skills.e

ENSO has in fact played a significant role in the human response to climate, in particular with regard to agriculture, water, and health. → which shown to have the largest impact on the prices of vegetable oils, grains, and some industrial commodities. Particularly commodities produce in tropical regions.

The climate forecast dataset adopted within was generated using the climate reanalysis and forecast ensemble (CAFE) system.

Main Machine Learning models

- General circulation model (GCM) (GCM)-ENSO
- CAFE - Climate reanalysis and forecast ensemble system → generate datasets
- Autoregressive Integrated Moving Average (ARIMA) (AR) → Economic time series commodity forecasting.

Evaluation Metrics: BIC for model comparison, RMSE for model evaluation

Commodity forecast → building general AR models including the log returns of G7 averaged GDP as an exogenous variable.

AR-ENSO or GCM-ENSO model produces better commodity forecasts depending on the lead time, skill measure selected, and commodity in question.
Initial difference in the ability of the AR model and the GCm to forecast ENSO is small in comparison to the error associated with the commodity forecasting model as a whole.

Causal inference in complex multiscale systems

Causal inference in complex multiscale systems - AI for Missions Program

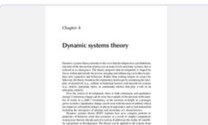
The *Causal inference and prediction in high dimensional multi-scale systems projects* seek to identify robust relationships between **climate and socio-economic** impacts.

Working with petabytes of short- and long-term weather, economic and geo-political data generated globally.

Project aims:

- identify robust relationships between climate and socioeconomic impacts
- produce robust estimates of climate-induced risks to the social and economic structures that underpin the nation's security and enhance the ability to successfully navigate oncoming impacts
- capability to produce climate data for user-defined emissions scenarios.

Model related to *dynamic system theory.*



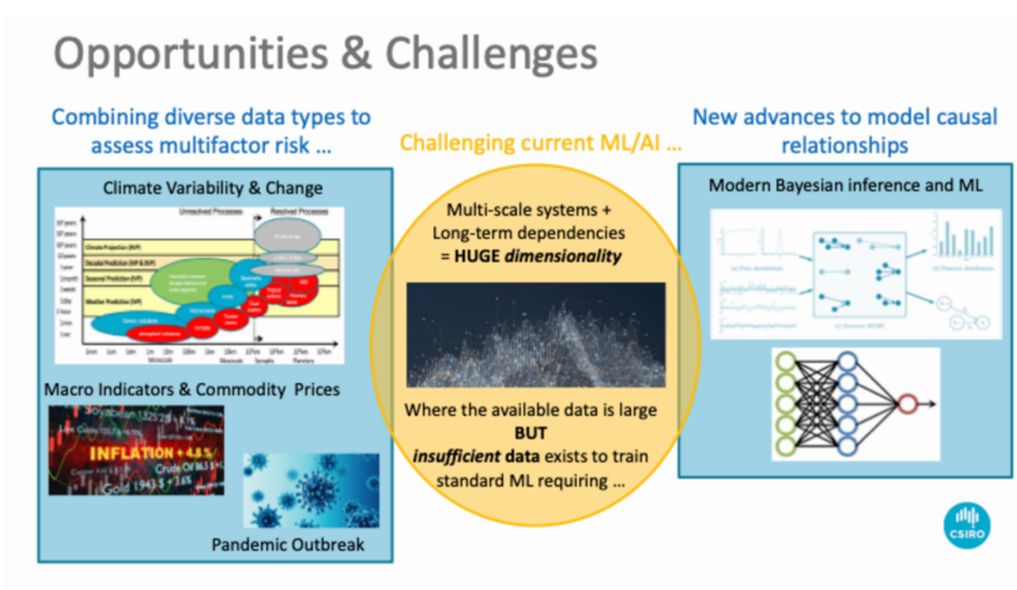
Chapter-4---Dy… ent.pdf
06 Apr 2023, 04:22 AM

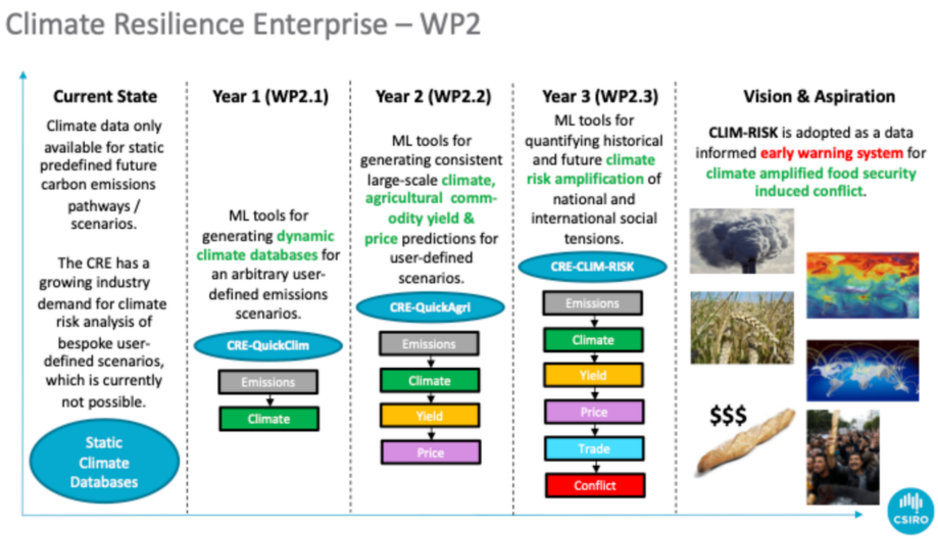
Integrated assessment models (IAMs) are the primary tool used to evaluate the technological and economic feasibility of climate goals such as the Paris Agreement's long-term temperature goal to hold global warming well below 2˚C.

- the NGFS (Network for Greening Financial System) emissions pathways inferred from the economic activity simulated in the current generation of IAMs, are not the emissions used to force climate model projects of future climatic states.
- an area having generated a massive data resource of the climate over the past six decades comprising close to one hundred replicant “digital” earths [9,10] generated through assimilation of a comprehensive set of atmospheric, ocean and sea-ice observations and a state-of-the-art global climate model (CAFE60 reanalysis - Registry of Open Data on AWS)

Climate impacts:

- agricultural and food production
- commodity prices





Methods:

- FEM-BV-VARX: Finite element bounded variation vector autoregressive with external factors
- eSPA+ entropy-optimal scalable probabilistic approximation
- Dynamic Bayesian Network (DBN)
- NODEs Neural Ordinary Differential Equations
- Free energy Gaussian state space models
- Variation Network Inference
- Deep convolutional gaussian processes

[1] Dezfouli, A., Bonilla, E. V., & Nock, R.. (2018). Variational Network Inference: Strong and Stable with Concrete Support. Proceedings of the 35th International Conference on Machine Learning, in Proceedings of Machine Learning Research.

[2] Harries, D., & O'Kane, T. J. (2021). Dynamic Bayesian networks for evaluation of Granger causal relationships in climate reanalyses. Journal of Advances in Modeling Earth Systems, 13, e2020MS002442. https:// http://doi.org/10.1029/2020MS002442

[3] I. Horenko, D. Rodrigues T.J. O'Kane and K. Everschor-Sitte (2021) Scalable detection of latent relations and their applications to magnetic imaging, Commun. Appl. Math. Comput. Sci., 16(2), 267—297, Communications in Applied Mathematics and Computational Science Vol. 16, No. 2, 2021

[4] I. Horenko (2020) On a scalable entropic breaching of the overfitting barrier for small data problems in machine learning, Neural Computation 32(8), 1563-1579

[7] Tran, G., Bonilla, E.V., Cunningham, J., Michiardi, P. & Filippone, M. (2019). Calibrating Deep Convolutional Gaussian Processes. Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research.

[8] Zhi W., Lai T., Ott L., Bonilla E. V., & Ramos F. (2021). Learning ODEs via Diffeomorphisms for Fast and Robust Integration. arXiv:2107.01650.

[11] E. Vecchi, L. Pospisil, S. Albrecht, T.J. O'Kane, I. Horenko (2021) eSPA+: Scalable entropy-optimal machine learning classification for small data problems (2022) 34 (5): 1220–1255.

Vassili Kitsios - Influence of Climate Variability on Financial Markets & Health Sector

[KEYNOTER] Vassili Kitsios - Influence of Climate Variability on Financial Markets & Health Sector

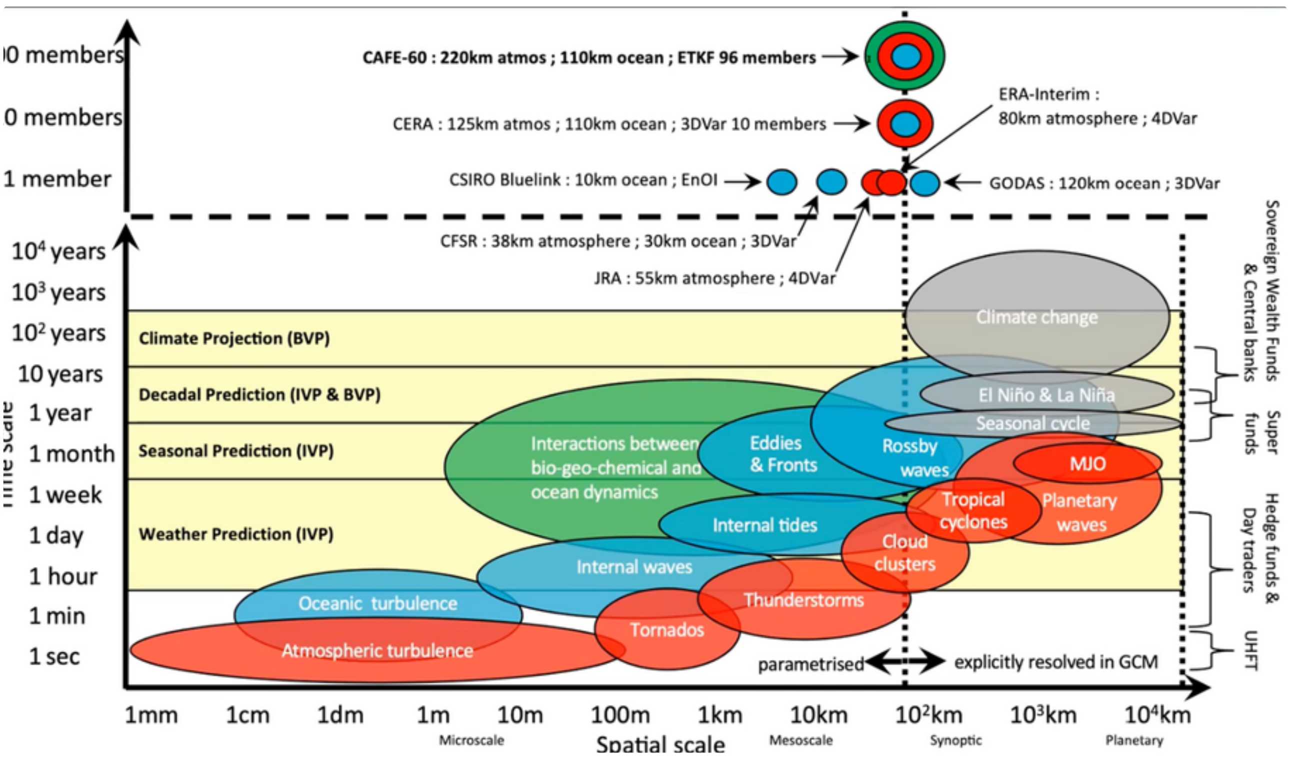
The intention is to provide a background on the climate system, modelling approaches, datasets, and examples of how one can make better-informed decisions concerning climate risk-exposed activities.

Specific questions addressed:

- What is a climate model?
- What is the distinction between the various climate datasets of forecasts, projections and reanalyses?
- What is the multi-scale association between climate and human systems?
- What is the appropriate framework to assess climate risk?
- How does multi-year climate variability influence commodity markets and the healthcare sector?

Earth system is complex and multi-disciplinary

Each system is multi-scale



Older historical data seems irrelevant to the current state.

Forecast largely depends on the boundary conditions

The difference scale makes it difficult to assess physical climate risk of individual dwellings or other assets

Framework for assessing climate risk

- Harvard

- Harvard
- Vulnerability
- Exposure

Climate Measurements and Standard Initiative (CMSI)

- Chronic - pervasive economy-wide impacts
- Acute - associated with the extreme weather events

> **Exogenous factors**: are external variables or influences that affect a system, process, or organism from outside. These factors are typically outside the control. Of the system or organism being affected may include various environment, economic, political, or social factors.

> **Endogenous**: refers to variables or factors that are internal to the system or process, rather than being influenced by external factors.

Potential model: auto regress model

Search on climate models -> more details provided by climate does generate a better model

Concludes:

- The climate and human systems are both complex and multi-scale in their own right.
- However, there are informative relationships between the climate hazard and exposed assets for certain patio-temporal scales
- Relationships learnt between climate variability and human systems, can inform how these human systems might respond under various future climate change scenarios

CSIRO climate resilience enterprise

Better enable the industry to assess, mitigate and/or adapt to their climate risk. Initiatives include (but are not limited to):

- Developing a climate data marketplace in conjunction with Microsoft to host a curated set of forecasts, projections and reanalysis datasets
- Developing analytical methods (it sounds hard) and an example workflow to enable the assessment of physical chronic and acute climate risk
- Undertaking integrated assessment modelling for a broader range of economic scenarios to better assess the breadth of transition risks catered toward the Australian economy and its trading partners

Dynamic Bayesian Networks for Evaluation of Granger Causal Relationships in Climate Reanalyses

Dylan Harries and Terence J.O'Kane

Dynamic Bayesian Network (DBN)

> Due to deficiencies in the parameterization of unresolved processes. Limitations manifest as systematic baiases in the output of the model compared to the obvseration.

In this paper, Harries focuses on Regression Analysis using DBN, considering lag factors to generate a climate reanalysis model.

In the DBN model, mathematically, observations are collections of individual, dynamic and non-linear data. This is caused by the nature of the environmental system, in fact, a change in one factor could result in a change in another.

In addition, theoretically, there must be a full model (beauty mathematics equation) that represents the climate system. But it is impossible, as mentioned above because variables/observations are not identical and independent variables, also climate system is a large-scale problem and is potentially affected by unknown factors. Hence, most of the climate models approach by decomposing the problem into sub-systems, trying to find the relationship between one more and another.

In practice, Harries built DBN via a combination of directed acyclic graphs (DAGs), Bayesian inference and Linear Auto-Regressive Models. Since we focus on the sub-systems (particularly with JRA-55). Each node in DBN represents a random event state. All random events are connected by edges, in that case, the model becomes a Markov chain model in which edges represent the interaction and the probability of state transition. To improve the model efficiency, removed less correlated edges using the mutual information criterion.

There are two major limitations:

- we have limited samples but they all have non-zero values,
- although events are non-linear and dependent on some other factors, it's likely to have a low correlation, as a result, it's hard to exclude weak correlations in BN.

For the project, at the current stage, we understand that we're facing a complex, large-scale problem. Since we don't have enough data plus unknown measurement accuracy and insufficient domain knowledge, we should focus on solving sub-problems.

→ thus, per Vassili's advice, we should focus on an Auto-Regressive model with factor JRA-55 and ENSO to predict crops' prices, especially WHEAT.

As a result, we should use this AR model as a baseline to measure the performance of other models.

In addition, we should use this AR model to find the Causality, moreover, Granger Causality and extend this model to other fields.

Economic forecasting in agriculture

International Journal of Forecasting 10, 81-135. Allen PG (1994)

Can exchange rates forecast commodity prices?

Quarterly Journal of Economics 125, 1145-1194.

Forecasting the Yield and the Price of Cotton

Moore HL, New York, NY: The Macmillian Company

Time series models

Time series models are statistical or machine learning models used to analyze and forecast values in a sequence of data points, measured over time. These models take into account the temporal ordering of observations and often exploit patterns, trends, seasonality, and other temporal structures in the data to make predictions.

Algorithms suitable for climate forecasting and commodity price forecasting include:

1. Autoregressive Integrated Moving Average (ARIMA) : A linear statistical model that combines autoregression (AR), differencing (I), and moving average (MA) components to capture trends, seasonality, and noise in the data.
2. Seasonal Decomposition of Time Series (STL): A technique that decomposes a time series into its seasonal, trend, and residual components, allowing for the modeling and forecasting of each component separately.
3. Exponential Smoothing State Space Model (ETS): A family of forecasting models that use weighted averages of past observations with an exponential decay to capture various types of trends and seasonality in the data.
4. SARIMA (Seasonal ARIMA): An extension of the ARIMA model that specifically accounts for seasonal patterns in the data.
5. Prophet: An open-source forecasting tool developed by Facebook that combines additive regression models with time series decomposition to handle seasonality, trends, and holidays.
6. Long Short-Term Memory (LSTM) networks: A type of recurrent neural network (RNN) that is designed to capture long-term dependencies in time series data. LSTM networks can be used to model complex patterns and relationships in climate and commodity price data.
7. Gated Recurrent Units (GRU): Another type of RNN that is simpler and computationally more efficient than LSTM networks while still being effective at capturing long-term dependencies in time series data.
8. Convolutional Neural Networks (CNN): CNNs can also be used for time series forecasting, as they can capture local patterns and dependencies in sequential data by using convolutional layers.
9. Vector Autoregression (VAR): A multivariate time series model that captures the linear dependencies between multiple time series and can be used to forecast multiple climate variables or commodity prices simultaneously.
10. Bayesian Structural Time Series (BSTS): A probabilistic approach to time series modeling that combines structural time series models with Bayesian inference techniques. BSTS can incorporate various components, such as trends, seasonality, and external regressors, to model complex time series data.

These algorithms can be adapted and fine-tuned to address the specific challenges and requirements of climate and commodity price forecasting tasks.

Autoregressive Integrated Moving Average (ARIMA)

A linear statistical model that combines autoregression (AR), differencing (I), and moving average (MA) components to capture trends, seasonality, and noise in the data.

The *AR* component captures the relationship between the current observation and its lagged values (previous observations).

The *I* component is used to make the time series stationary by differencing it, while

The *MA* component captures the relationships between the current observation and past errors.

> Example generated by GPT4

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

# Load your time series data
# data = pd.read_csv('your_data.csv', index_col='date', parse_dates=True)

# Generate example data
np.random.seed(42)
data = pd.Series(np.random.randn(100).cumsum(), index=pd.date_range('2000-01-01', periods=100))

# Plot the time series data
data.plot()
plt.show()

# Determine the order of differencing (d) using ACF and PACF plots
plot_acf(data)
plot_pacf(data)
plt.show()

# Choose appropriate values for p, d, and q based on ACF and PACF plots
p = 1
d = 1
q = 1

# Fit the ARIMA model
model = ARIMA(data, order=(p, d, q))
results = model.fit()

# Forecast the next 10 data points
forecast, stderr, conf_int = results.forecast(steps=10)

# Plot the forecasted values
plt.plot(data.index, data, label='Observed')
plt.plot(pd.date_range(data.index[-1], periods=11)[1:], forecast, label='Forecast')
plt.fill_between(pd.date_range(data.index[-1], periods=11)[1:], conf_int[:, 0], conf_int[:, 1], color='pink')
plt.legend()
plt.show()
```

This code snippet demonstrated loading a time series dataset, plotting the data, determining the appropriate order of differencing using the autocorrelation function (ACF) and partial autocorrelation (PACF) plots, fitting the ARMIA model, and forecasting future values.

Remember to replace the example data with your own time series data in CSV format, and adjust the `p`, `d`, and `q` values based on the ACF and PACF plots.

General circulation model (GCM)

A General Circulation Model (GCM), also known as a Global Climate Model, is a mathematical representation of the Earth's climate system that simulates the interactions between the atmosphere, oceans, land surfaces, and cryosphere (ice and snow). GCMs are based on fundamental physical, chemical, and biological processes governing the behavior of these components and their interactions. These models are widely used in climate research and forecasting to study and predict long-term climate changes and variations.

GCMs are typically implemented as computer programs and run on high-performance computing clusters due to their complexity and computational demands. They divide the Earth's surface into a grid system, with each grid cell representing a specific geographical area. The vertical dimension is also discretized into multiple layers, representing different altitudes in the atmosphere and depths in the ocean.

The key components of GCMs include:

1. Atmospheric model: Simulates the behavior of the atmosphere, including temperature, pressure, humidity, winds, and precipitation. This component includes representations of various physical processes, such as radiation, convection, cloud formation, and large-scale dynamics.
2. Ocean model: Represents the behavior of the oceans, including temperature, salinity, currents, and sea ice. It simulates ocean circulation, mixing, and the transport of heat and salt. The ocean model also includes representations of sea ice dynamics and thermodynamics.
3. Land surface model: Describes the interactions between the atmosphere and the land surface, including processes such as soil moisture, evapotranspiration, runoff, and vegetation dynamics.
4. Sea ice model: This represents the growth, melting, and movement of sea ice, as well as its interaction with the atmosphere and the ocean.
5. Snow model: Simulates snow accumulation, melting, and redistribution on the land surface and sea ice.
6. Biosphere model: This represents the role of living organisms in the climate system, such as photosynthesis, respiration, and carbon cycling.

GCMs are essential tools for understanding and predicting the Earth's climate and its response to external forcings, such as increasing greenhouse gas concentrations, solar variability, and volcanic eruptions. They are used to study historical climate changes, project future climate scenarios, and assess the potential impacts of climate change on ecosystems, agriculture, water resources, and human societies.

Creating a full General Circulation Model (GCM) is a complex task that requires specialized knowledge and expertise. GCMs are typically developed and maintained by large teams of scientists and programmers, and their codes can be tens of thousands of lines long.

Here is a code snippet of simple 1D Energy Balance Model (EBM):

```python
import numpy as np
import matplotlib.pyplot as plt

# Constants
sigma = 5.67e-8  # Stefan-Boltzmann constant (W m^-2 K^-4)
solar_constant = 1361  # Solar constant (W m^-2)
albedo = 0.3  # Earth's average albedo
emissivity = 1.0  # Earth's emissivity (1.0 for a blackbody)

# Define a simple EBM function
def energy_balance_model(temp):
    absorbed_solar_radiation = (1 - albedo) * solar_constant / 4
    outgoing_longwave_radiation = emissivity * sigma * temp**4
    net_energy_flux = absorbed_solar_radiation - outgoing_longwave_radiation
    return net_energy_flux

```

```python
# Find the equilibrium temperature
equilibrium_temperature = (solar_constant * (1 - albedo) / (4 * sigma * emissivity))**0.25

# Plot the energy balance
temperatures = np.linspace(200, 350, 100)
energy_fluxes = [energy_balance_model(temp) for temp in temperatures]

plt.plot(temperatures, energy_fluxes, label='Net energy flux')
plt.axhline(0, color='k', linestyle='--')
plt.axvline(equilibrium_temperature, color='r', linestyle='--', label=f'Equilibrium temperature: {equilibrium_te
plt.xlabel('Temperature (K)')
plt.ylabel('Net energy flux (W m^-2)')
plt.legend()
plt.show()
```

CAFE - Climate reanalysis and forecast ensemble system

The CAFE (Community Atmosphere and Forecast Ensemble) climate reanalysis and forecast ensemble system is a sophisticated, state-of-the-art weather and climate prediction system. It combines observations, reanalysis, and ensemble forecasting techniques to provide accurate and reliable climate data and predictions. The system is designed to address the challenges associated with weather and climate forecasting, such as uncertainties, biases, and complex interactions between various components of the Earth system

Dynamic Bayesian Network (DBN)

A Dynamic Bayesian Network (DBN) is a probabilistic graphical model that represents a temporal sequence of random variables and their dependencies. DBNs extend traditional Bayesian networks by incorporating the concept of time and modeling the relationships between variables at different time steps. They are widely used in various fields, such as speech recognition, finance, bioinformatics, and robotics, to model and predict time series data with complex dependencies and uncertainties.

A DBN consists of a directed acyclic graph (DAG) where nodes represent random variables, and edges represent conditional dependencies between variables. The model is defined by two sets of conditional probability distributions (CPDs): one for the initial time step (t=0) and one for the transition between consecutive time steps (t to t+1).

```
import numpy as np
from pomegranate import *

# Define the initial probabilities
initial_probs = DiscreteDistribution({'Sunny': 0.5, 'Rainy': 0.5})

# Define the transition probabilities
transition_probs = ConditionalProbabilityTable([
    ['Sunny', 'Sunny', 0.8],
    ['Sunny', 'Rainy', 0.2],
    ['Rainy', 'Sunny', 0.3],
    ['Rainy', 'Rainy', 0.7]
], [initial_probs])

# Define the emission probabilities (observation model)
emission_probs = {
    'Sunny': DiscreteDistribution({'Dry': 0.9, 'Wet': 0.1}),
    'Rainy': DiscreteDistribution({'Dry': 0.2, 'Wet': 0.8})
}

# Create the HMM (DBN) model
states = State.create_multiple([initial_probs, transition_probs], emission_probs)
model = HiddenMarkovModel.from_states(states)

# Generate a sequence of observations
observations = ['Dry', 'Wet', 'Dry', 'Wet', 'Wet']

# Infer the most likely hidden state sequence
state_sequence = model.predict(observations)

# Decode the state sequence
decoded_sequence = [s.name for s in state_sequence]
print("Most likely hidden state sequence:", decoded_sequence)

```

This code snippet demonstrates how to create a simple 2-state HMM (a DBN) using the `pomegranate` library, define its initial, transition, and emission probabilities, and infer the most likely hidden state sequence given a sequence of observations. Note that this example is a basic representation of a DBN, and more complex models can be built by extending the structure and incorporating additional variables and time steps.

FEM-BV-VARX: Finite element bounded variation vector autoregressive with external factors

Autoregressive Integrated Moving Average (ARIMA)

FEM-BV-VARX is a hybrid model that combines the Finite Element Method (FEM), Bounded Variation (BV), Vector Autoregressive (VAR), and external factors (X) in order to create a flexible and powerful framework for time series analysis and forecasting. This method aims to tackle problems with non-stationary, non-linear, and multivariate data while also accounting for external influences.

Finite Element Method (FEM): FEM is a numerical technique that can be used to approximate solutions to differential equations by breaking down the problem domain into smaller, simpler parts called “elements “. This is particularly useful in solving complex problems in engineering, physics, and other fields.

Bounded Variation (BV): In the context of time series analysis, BV refers to a functional space where the total variation (i.e, the sum of the absolute differences between consecutive values) is bounded. This property helps to control the smoothness of the estimated function and prevent overfitting.

Vector Autoregressive (VAR): VAR models are multivariate time series models that aim to describe the joint dynamic behavior of several variables over time. They capture the relationships between multiple variables by considering their lagged values as predictors.

External Factors (X): in VARX models, external factors (X) represent variables that are not part of the core system but may influence it. These factors can be used to enhance the predictive ability of the model by accounting for external influences on the time series of interest.

```python
import numpy as np
import pandas as pd
from scipy.interpolate import BivariateSpline
from statsmodels.tsa.api import VAR
from sklearn.preprocessing import StandardScalar

data = pd.read_csv("data.csv")
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)

# Split your data into endogenous (core system) and exogenous (external factors) variables
endogenous_data = data_scaled[:, :n_endogenous]
exogenous_data = data_scaled[:, n_endogenous:]

# Apply FEM with BV to endogenous_data
# (For the sake of simplicity, we use BivariateSpline, but you should use a FEM implementation that supports BV)
x, y = np.linspace(0, 1, endogenous_data.shape[0]), np.linspace(0, 1, endogenous_data.shape[1])
x_grid, y_grid = np.meshgrid(x, y)
splines = [BivariateSpline(x, y, endogenous_data[:, i]) for i in range(endogenous_data.shape[1])]

# Extract smoothed data for VAR
smoothed_endogenous_data = np.array([spline(x_grid, y_grid) for spline in splines]).T

# Fit a VAR model with exogenous factors (VARX)
lag_order = 2  # Choose an appropriate lag order based on your data
model = VAR(endog=smoothed_endogenous_data, exog=exogenous_data)
results = model.fit(lag_order)

# Forecast
n_forecast = 10  # Choose an appropriate forecasting horizon
exog_future = exogenous_data[-lag_order:]  # Prepare future exogenous data, if necessary
forecast = results.forecast(y=smoothed_endogenous_data[-lag_order:], steps=n_forecast, exog_future=exog_future)
```

```

# Inverse scaling and other necessary transformations, if needed
forecast_original_scale = scaler.inverse_transform(forecast)
```

eSPA+ entropy-optimal scalable probabilistic approximation

eSPA+ is a method that aims to provide an efficient and scalable approximation of complex systems by optimizing a probabilistic model with respect to entropy. eSPA+ can be applied in various domains, such as time series analysis, image processing, and network analysis, among others.

The idea behind eSPA+ is to find an optimal trade-off between accuracy and complexity in representing the system's behavior. By optimizing the entropy, eSPA+ ensures that the model captures the essential characteristics of the system while reducing computational complexity and memory requirements.

```python
import numpy as np
from sklearn.mixture import GaussianMixture
from scipy.stats import entropy
from sklearn.preprocessing import StandardScaler

# Load your data (e.g., time series, image, network)
# data = np.load('your_data.npy')

# Preprocess your data (e.g., normalize)
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)

# Define the eSPA+ optimization function
def optimize_entropy(data, n_components_range, random_state=42):
    best_entropy = float("inf")
    best_gmm = None

    for n_components in n_components_range:
        gmm = GaussianMixture(n_components=n_components, random_state=random_state)
        gmm.fit(data)
        labels = gmm.predict(data)
        probs = gmm.predict_proba(data)

        # Compute entropy
        ent = entropy(probs.T)

        if np.mean(ent) < best_entropy:
            best_entropy = np.mean(ent)
            best_gmm = gmm

    return best_gmm

# Optimize eSPA+ (find the optimal number of components)
n_components_range = range(1, 11)  # Adjust this range based on your problem
optimal_gmm = optimize_entropy(data_scaled, n_components_range)

# Use the optimal Gaussian Mixture Model for further analysis or predictions
# labels = optimal_gmm.predict(new_data)

```

NODEs Neural Ordinary Differential Equations

Neural Ordinary Differential Equations (NODEs) are a type of neural network that combines deep learning with the theory of ordinary differential equations (ODEs). Instead of using discrete layers with fixed operations, NODEs define a continuous-time transformation of the input data through the ODE. This transformation is parameterized by a neural network, which is learned during training.

NODEs can be applied to time series analysis by modeling the continuous-time dynamics of the time series data. This can be beneficial for tasks like forecasting, anomaly detection, and data imputation. NODEs can capture complex dynamics and handle irregularly sampled time series data, making them well-suited for various time series applications.

```
import torch
import torhc.nn as nn
from torchdiffeq import odeint_adjoint as odeint
import numpy as np

class ODEFunc(nn.Module):
    def __init__(self):
        super(ODEFunc, self).__init__()
        self.net = nn.Sequential(
            nn.Linear(2, 50),
            nn.ReLU,
            nn.Linear(50, 2)
        )

    def forward(self, t, y):
        return self.net(y)

class NODEModel(nn.Module):
    def __init__(self, ode_func):
        super(NODEModel, self).__init__()
        self.ode_func = ode_func

    def forawrd(self, t, y0):
        return odeint(self.ode_func, y0, t)

# Create the ODE function and NODE model
ode_func = ODEFunc()
model = NODEModel(ode_func)

# Example time series data
timesteps = torch.linspace(0., 4., 100)
y0 = torch.tensor([[1., 0.]])
output_true = 2 * torch.sin(2 * np.pi * timesteps)

optimizer = torch.optim.Adam(model.parameters(), lr=1e-3)
loss_fn = nn.MSELoss()

num_epochs = 500
for epoch in range(num_epochs):
    optimizer = zero_grad()
    output_pred = model(timesteps, y0).squeeze()
    loss = loss_fn(output_pred, output_true)
    loss.backward()
```

```
    optimizer.step()

    if (epoch + 1) % 50 == 0:

        print(f'Epoch [{epoch + 1}/{num_epochs}], Loss: {loss.item():.4f}')
```

Free energy Gaussian state space models

Free energy Gaussian state space models (FEGSSMs) are a probabilistic modeling framework that represents the generative process of time series data using Gaussian distributions. They are also known as Gaussian dynamic models or linear Gaussian state space models. GSSMs are popular for their simplicity, tractability, and ability to handle uncertainty in data. The models involve an underlying state process that evolves linearly with Gaussian noise, and an observation process that is also linear with Gaussian noise.

Free energy refers to an approximation technique that simplifies complex inferences in the state space models by minimizing an energy-based objective function (called the free energy) to find the optimal Gaussian approximation to the true posterior distribution.

GSSMs can be applied to time series analysis tasks such as filtering, smoothing, prediction, and model selection. The Kalman filter, a popular algorithm for time series analysis, is an instance of GSSMs.

```
import numpy as np
import matplotlib.pyplot as plt
from pykalman import KalmanFilter

# Generate synthetic time series data
num_timesteps = 200
timesteps = np.arange(num_timesteps)
true_state = np.sin(0.1 * timesteps)  # Example target function
measurements = true_state + 0.1 * np.random.randn(num_timesteps)  # Add Gaussian noise

# Define the initial state and covariance
initial_state_mean = measurements[0]
initial_state_covariance = 1.0

# Define the transition matrix and covariance
transition_matrix = np.eye(1)
transition_covariance = 0.0

# Define the observation matrix and covariance
observation_matrix = np.eye(1)
observation_covariance = 0.1

# Create and fit the Kalman filter
kf = KalmanFilter(
    initial_state_mean=initial_state_mean,
    initial_state_covariance=initial_state_covariance,
    transition_matrices=transition_matrix,
    transition_covariance=transition_covariance,
    observation_matrices=observation_matrix,
    observation_covariance=observation_covariance,
)
filtered_state_means, _ = kf.filter(measurements)

# Plot the true state, measurements, and filtered state estimates
plt.plot(timesteps, true_state, label='True State')
plt.plot(timesteps, measurements, label='Measurements', linestyle=':', marker='o', markersize=3)
plt.plot(timesteps, filtered_state_means, label='Filtered State', linestyle='--')
plt.legend()
plt.show()

```

Variation Network Inference

Variation Inference is a technique for approximate Bayesian inference that transforms the problem of computing a complex posterior distribution into an optimization problem. The goal is to find an approximation from a simpler family of distributions that is close to the true posterior. Variational inference can be applied to time series analysis when dealing with complex Bayesian models.

```
import torch
import pyro
import pyro.distributions as dist
from pyro.contrib.examples.util import get_data_loader
from pyro.infer import SVI, Trace_ELBO
from pyro.optim import Adam

# Define the model
def model(climate_factors, commodity_prices):
    num_timesteps = len(commodity_prices)
    # Define the priors for the latent state variables
    mu = pyro.sample("mu", dist.Normal(0, 1))
    sigma = pyro.sample("sigma", dist.HalfNormal(1))

    with pyro.plate("timesteps", num_timesteps):
        # Define the latent state process
        latent_state = pyro.sample("latent_state", dist.Normal(mu, sigma))
        # Define the observation process
        prediction = pyro.sample("prediction", dist.Normal(latent_state + climate_factors, 1), obs=commodity_pri

# Define the guide function (variational distribution)
def guide(climate_factors, commodity_prices):
    mu_loc = pyro.param("mu_loc", torch.tensor(0.0))
    mu_scale = pyro.param("mu_scale", torch.tensor(1.0), constraint=dist.constraints.positive)
    sigma_loc = pyro.param("sigma_loc", torch.tensor(1.0), constraint=dist.constraints.positive)

    pyro.sample("mu", dist.Normal(mu_loc, mu_scale))
    pyro.sample("sigma", dist.HalfNormal(sigma_loc))

    with pyro.plate("timesteps", len(commodity_prices)):
        latent_loc = pyro.param("latent_loc", torch.zeros(len(commodity_prices)))
        latent_scale = pyro.param("latent_scale", torch.ones(len(commodity_prices)), constraint=dist.constraints
        pyro.sample("latent_state", dist.Normal(latent_loc, latent_scale))

# Create synthetic climate factors and commodity prices data
num_timesteps = 100
climate_factors = torch.randn(num_timesteps)
commodity_prices = 3 * climate_factors + torch.randn(num_timesteps)

# Initialize the optimizer, loss function, and inference engine
optimizer = Adam({"lr": 0.01})
svi = SVI(model, guide, optimizer, loss=Trace_ELBO())

# Training loop
num_epochs = 1000
for epoch in range(num_epochs):
    loss = svi.step(climate_factors, commodity_prices)
```

```
    if (epoch + 1) % 100 == 0:
        print(f'Epoch [{epoch + 1}/{num_epochs}], Loss: {loss:.4f}')
```

Deep convolutional gaussian processes

Deep Convolutional Gaussian Processes DCGPs are an advanced machine learning technique that combines the strengths of deep learning and Gaussian Processes (GPs) to model complex data. They can be particularly useful in time series analysis, as they can capture non-linear patterns and make predictions with uncertainty estimates.

```python
import torch
import gpytroch
import numpy as np
import matplotlib.pyplot as plt

# Generate synthetic data
num_data_points = 100
x = torch.linspace(0, 1, num_data_points)
temperature = torch.sin(12 * x) + 0.66 * torch.cos(25 * x) + torch.randn(x.size()) * 0.1
commodity_price = torch.sin(10 * x) + 0.5 * torch.cos(20 * x) + torch.randn(x.size()) * 0.1

# Train-test split
train_frac = 0.8
num_train = int(train_frac * num_data_points)
train_x, test_x = x[:num_train], x[num_train:]
train_y, test_y = temperature[:num_train], temperature[num_train:]

# Create a DCGP model
class DCGPModel(gpytorch.models.ExactGP):
    def __init__(self, train_x, train_y, likelihood):
        super(DCGPModel, self).__init__(train_x, train_y, likelihood)
        self.mean_module = gpytorch.means.ConstantMean()
        self.covar_module = gpytorch.kernels.ScaleKernel(
            gpytorch.kernels.GridInterpolationKernel(
                gpytorch.kernels.RBFKernel(ard_num_dims=1),
                grid_size=50,
                num_dims=1,
            )
        )

    def forward(self, x):
        mean_x = self.mean_module(x)
        covar_x = self.covar_module(x)
        return gpytorch.distributions.MultivariateNormal(mean_x, covar_x)

# Train the model
likelihood = gpytorch.likelihoods.GaussianLikelihood()
model = DCGPModel(train_x, train_y, likelihood)

# Use the Adam optimizer
optimizer = torch.optim.Adam(model.parameters(), lr=0.1)

# "Loss" for GPs is the marginal log likelihood
mll = gpytorch.mlls.ExactMarginalLogLikelihood(likelihood, model)

training_iterations = 50
model.train()
```

```
likelihood.train()

for i in range(training_iterations):
    optimizer.zero_grad()
    output = model(train_x)
    loss = -mll(output, train_y)
    loss.backward()
    print("Iter %d/%d - Loss: %.3f" % (i + 1, training_iterations, loss.item()))
    optimizer.step()

# Make predictions
model.eval()
likelihood.eval()

with torch.no_grad(), gpytorch.settings.fast_pred_var():
    test_preds = likelihood(model(test_x))
    mean = test_preds.mean
    lower, upper = test_preds.confidence_region()

# Plot predictions
plt.figure(figsize=(12, 6))
plt.plot(x, temperature, 'k*')
plt.plot(test_x, mean, 'b')
plt.fill_between(test)
```

Univariate autoregressive models

Univariate autoregressive (AR) models are a class of linear series models used to represent and forecast a single variable based on its own past values. The main idea behind the autoregressive is that the current value of the time series can be expressed as a linear combination of its previous value plus an error term. The error term represents the unpredictable part of the time series.

$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t$

Where:

- Y_t is the value of the time series at time t
- c is a constant term
- φ_i (i = 1, 2, ..., p) are the autoregressive parameters
- Y_{t-i} are the past values of the time series (lags)
- ε_t is a white noise error term with mean 0 and constant variance σ^2

The order p of the AR model indicates the number of lags included in the model. The choice of p is an important aspect of model selection and can be determined using various criteria, such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or by examining the autocorrelation and partial autocorrelation functions of the time series.

Univariate AR models assume that the underlying time series is stationary, meaning that its statistical properties (such as mean and variance) do not change over time. If a time series is non-stationary, it may need to be transformed (e.g., differenced) before applying an AR model.

Bivariate autoregressive model

A bivariate autoregressive (AR) model is a type of multivariate time series model that aims to capture the linear relationship between two variables over time. It considers both the past values of each variable and their interactions to model and predicts the current values of both variables. Bivariate AR models are useful for studying the dynamic relationships between two time series and for jointly forecasting their future values.

A bivariate autoregressive model of order p, denoted as AR(p), can be written as:

$$Y1_t = c1 + \varphi11_1 Y1_{t-1} + ... + \varphi11_p Y1_{t-p} + \varphi12_1 Y2_{t-1} + ... + \varphi12_p Y2_{t-p} + \varepsilon1_t$$

$$Y2_t = c2 + \varphi21_1 Y1_{t-1} + ... + \varphi21_p Y1_{t-p} + \varphi22_1 Y2_{t-1} + ... + \varphi22_p Y2_{t-p} + \varepsilon2_t$$

Where:

- $Y1_t$ and $Y2_t$ are the values of the two time series at time t
- c1 and c2 are constant terms
- φij_i (i, j = 1, 2; i ≠ j) are the autoregressive parameters
- $Y1_{t-i}$ and $Y2_{t-i}$ are the past values (lags) of the two time series
- $\varepsilon1_t$ and $\varepsilon2_t$ are the white noise error terms for each time series, with mean 0 and constant variances $\sigma1^2$ and $\sigma2^2$, respectively

In this model, each equation expresses the current value of a time series as a linear combination of its own past values, the past values of the other time series, and an error term. The order p of the bivariate AR model indicates the number of lags included in the model for each time series. Model selection, including the choice of p, can be guided by criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

Bivariate AR models assume that the underlying time series are stationary. If either or both time series are non-stationary, they may need to be transformed (e.g., differenced) before applying the model. In cases where the relationship between the two variables is more complex, other multivariate models like vector autoregression (VAR) or vector error correction models (VECM) may be more appropriate. Bivariate AR models can be extended to include more than two variables, resulting in a multivariate autoregressive model.

Climate Change Risk Assessment

- The Paris Agreement set the common goal of limiting global average temperature increases (relative to pre-industrial levels) to ‘well below 2 degrees and 'pursuing efforts’ to 1.5 degrees

> The risks of climate change can be understood more clearly when research starts by identifying what it is that we most wish to avoid and then assessing its likelihood as a function of time (Sharp, 2019)

Sharpe, S. (2019), ‘Telling the boiling frog what he needs to know: why climate change risks should be plotted as a probability over time, Geoscience Communication, 2(1), pp. 95–100, doi:10.5194/gc-2-95-2019 (accessed 13 Aug. 2021).

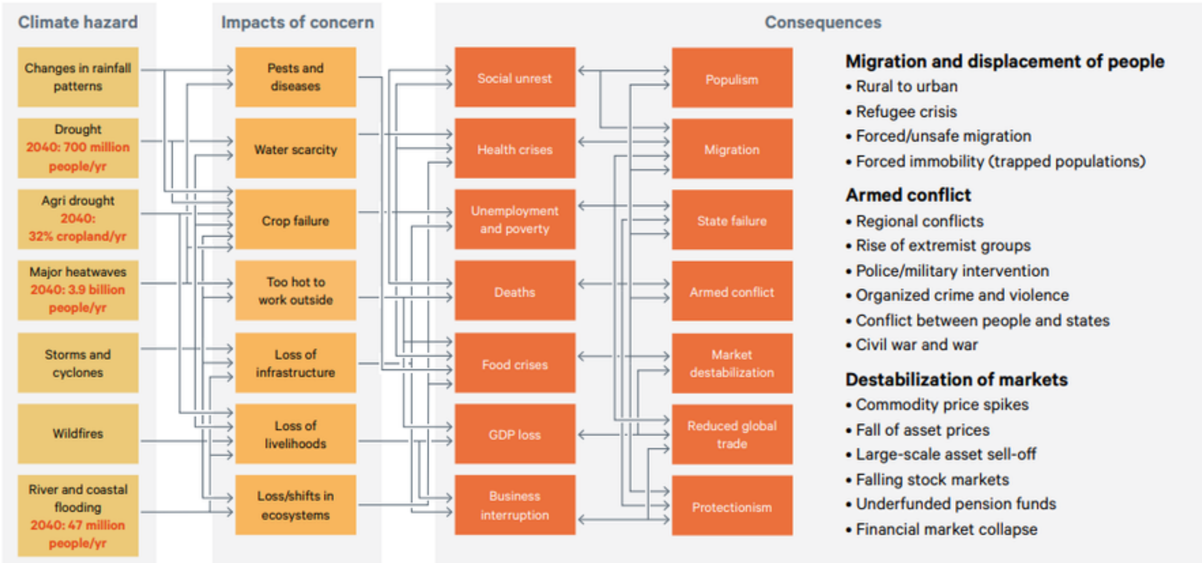
- Increased temperatures and heatwaves are increasingly limiting labour productivity and causing heat-stress-related mortality.

Food Security

- To meet global demand, agriculture must produce almost 50% more food by 2050.
- Between 1980 and 2019, global average crop yield potentials for maise, winter wheat, soybeans and rice have declined, with reductions of 5.6 per cent, 2.1 per cent, 4.8 per cent and 1.8 per cent, respectively.
- **In Australia, severe drought caused a 50% in collapse of wheat harvests two years in a row (2006 and 2007).** 9 Index Mundi (2021), ‘Australia Wheat Production by Year’, Agricultural Production Statistics by Country ? country=au&commodity=wheat&graph=production (accessed 13 Jun. 2021).

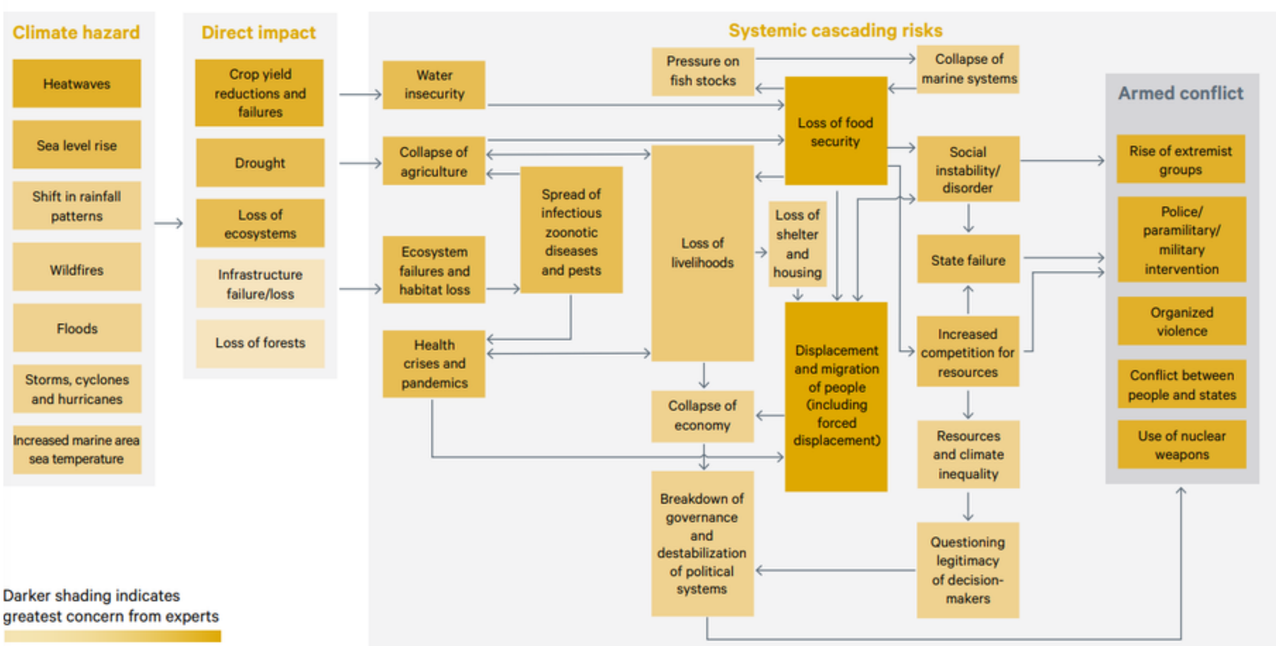
→ some consequence: by 2050, the central estimate indicates that nearly 40% of global cropland area will be exposed to serve drought for three months or more each year.

increasing temperature → crop failure → decreasing agricultural yields → increase demands but decrease supply → crop price increase (put this in conclusion)



climate harzard has direct impact on five sectors: population, price, ….

-> lead to systemic cascading risks



Forecasting commodity returns by exploiting climate model forecasts of the El Niño Southern Oscillation

(Section wise notes)

Introduction

Despite of highly complex human and geophysical systems, there might be a relationship between both. **There can be some relationships between climate and the economy that can be predicted for certain time and space scales.**

The study aims to understand how climate variability affects the accuracy of commodity forecasts, particularly for vegetable oil markets.

In this paper, researchers build models using **exogenous** climate-related factors and use econometric methods to analyze their relationship with commodity prices. It's worth noting that forecasting a geophysical/environmental factor does not necessarily follow the same approach as forecasting the commodity price itself. Various approaches to climate prediction are adopted herein, including both econometric methods, and also more computationally expensive physics-based dynamical simulations of the Earth system.

Weather and seasonal predictions depend on initial conditions, and how well these are known can affect forecast accuracy. Decadal predictions are a combination of both boundary and initial conditions. When making multiyear forecasts, it is crucial to consider physical phenomena that occur over the associated timescales. For example, the El Niño Southern Oscillation (ENSO) is an irregularly periodic variation in sea-surface temperatures and winds over the tropical regions in the Pacific Ocean that dominates variability over the annual to multiyear timescale band.

The relationship between ENSO and commodities and broader macroeconomic indicators has been studied to determine the economic risk associated with ENSO.

Studies have shown that ENSO has a direct impact on gross domestic product (GDP) and the spillovers (indirect impacts) associated with trade and financial markets. The impact of ENSO on commodity prices is most direct for commodities that are significantly sourced from tropical regions, such as vegetable oils, grains, and some industrial commodities. The percentage contribution of **ENSO to the forecast error variance for coconut oil log returns was found to be the largest at a 4-year lead time.**

(To be included in the report and presentation with proof)

Impact of ENSO on commodities: During the La Niña phase, there are close linkages between soybean futures price movements and ENSO behavior for monthly timescales.

The studies have also found that Brazilian corn production has negative correlations with sea-surface temperature anomalies, while corn prices have positive correlations.

For wheat, rice, and maize, the neutral phase is the most productive, with increased magnitude in either phase of ENSO producing below average yield.

However, for soybeans, El Niño produces above average yields, with yields in neutral and La Niña phases being below average.

The studies mentioned used **nonlinear multivariate regression** methods to analyze the relationship between ENSO and vegetable oil prices.

El Niño and La Niña phases have asymmetric impacts on vegetable oil prices, with El Niño phases leading to lower prices and La Niña phases leading to higher prices (negatively correlated).

Additionally, they found that coconut and palm kernel oils are more responsive to ENSO shocks than ground nut and cotton seed oils. The magnitude of the shock response on vegetable oil prices is larger in the El Niño phase than the La Niña phase.

Task of the paper: if the accuracy of econometric forecasts of commodity spot prices, specifically for vegetable oils produced in tropical regions of Asia, **can be improved by incorporating exogenous factors** related to ENSO provided by physics-based climate prediction simulations.

specific vegetable oils analyzed in this study are **coconut oil, palm oil, and soybean oil.**

The aim of the task is to explore if including climate forecast data in the econometric models can improve the accuracy of commodity price forecasts, potentially allowing for more effective risk management and decision-making in the agricultural commodity markets.

The climate forecast dataset adopted within was generated using the

climate reanalysis and forecast ensemble (CAFE) system

Outline of the Paper

Section 2

- provides further information about ENSO,
- the datasets used,
- climate model forecasts, and
- autoregressive (AR) forecasting methods.
- This section will explain **their methodology for forecasting ENSO and commodity prices.**

Section 3

- The most parsimonious AR forecast model governing the evolution of ENSO is constructed on monthly intervals.

- The out-of-sample forecast skill of this AR model is then compared to that of ENSO climate model forecasts from the CAFE system

- Additionally, **AR models are built for the real log returns,** and out-of-sample forecasts are undertaken for the vegetable oil commodities. The exogenous ENSO factors are provided by the CAFE system and the AR model.

- **Test commodity forecasts using no ENSO information as a lower bound and perfect future ENSO knowledge as a reference upper bound.** By comparing the skill of these cases, demonstrate the potential for improved economic forecasts by exploiting climate predictions.

Section 4

- Provide a discussion linking the results back to the physical climate system.
- This section will explain how the results of their economic forecasts relate to the underlying ENSO phenomenon and the broader climate system.

Section 5

- Provide concluding remarks and suggest potential avenues for future research in this area.

Section 2: Background and Methods

2.1. Physical description of the El Niño Southern Oscillation

· ENSO (El Niño-Southern Oscillation) is a naturally occurring climate **mode of variability** that causes **multiyear variations** in sea-surface temperature and winds over the tropical Pacific Ocean. **ENSO is a natural climate variability phenomenon.**

· **ENSO can lead to changes in atmospheric circulation patterns and wind speeds, which can in turn impact weather patterns around the world.**

· It comprises of the mature phases La Niña and El Niño, and also the neutral phase.

· During La Niña, there are cooler than average waters in the eastern Pacific, and warmer than average waters in the west, which is accompanied by enhanced evaporation.

· During El Niño, the opposite occurs, with warm waters in the east and cool waters in the west, and the Walker circulation breaks down.

· ENSO can affect rainfall and temperature in distant regions of the Earth, with the strongest impacts in the tropics.

The phases of ENSO have seasonal, asymmetric, and heterogeneous impacts on rainfall, temperature, and potentially economic activity.

2.2 Climate Observations

$$E_{\text{obs}}(d) = \langle T_{0,\text{obs}}(x,y,d) - \overline{T}_{0,\text{obs}}(x,y,\tau)\rangle,$$

· In this we do the calculation of the phase and magnitude of ENSO using the Niño4 index.

· The Niño4 index is used to classify the ENSO phase as El Niño, La Niña, or neutral.

· The magnitude of the index is proportional to the strength of the event, with magnitudes less than half of the standard deviation classified as neutral.

· These ENSO-related model coefficients and initial conditions are used in the out-of-sample forecasts.

· **positive values indicate warmer than average conditions (El Niño) and negative values indicate cooler than average conditions (La Niña) in the tropical Pacific.**

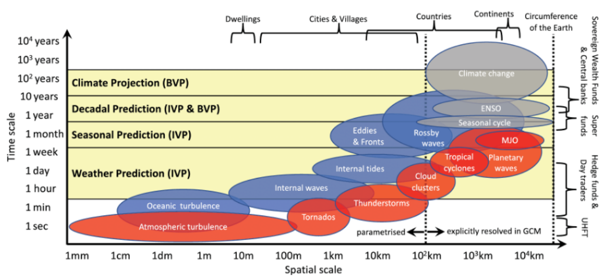
-

2.3 Climate model forecasting

· The general circulation model (GCM) is a numerical model used to simulate and predict the behaviour of the Earth's atmosphere and oceans.

· However, the vast range of scales present in the Earth's climate makes it computationally challenging to have a grid domain large enough to capture the entire globe.

· The CAFE system, used in this study, has been developed to overcome this challenge. The system uses a grid with a 1° resolution in longitude, with higher latitudinal resolution in specific regions, and 50 vertical levels in the ocean grid. **This high spatial resolution allows the system to explicitly resolve the dynamics of ENSO, which is important for this study.**



It should be noted that any physical phenomena to the left of the vertical dotted line in Figure 1 are unresolved and their influence on the large scales is parameterized. **Reason**: The global climate has a vast range of scales, making its numerical simulation challenging. It is not computationally feasible to have a grid that captures the entire globe while having divisions small enough to resolve the finest scale turbulence.

$$E_{\mathrm{gcm}}(d,\Delta t) = \langle T_{0,\mathrm{gcm}}(x,y,d,\Delta t) - \overline{T}_{0,\mathrm{obs}}(x,y,\tau)\rangle, \tag{2}$$

We now find the GCMNiño4 index using the formulae above. **GCMNiño4 index is a model-based representation of the Niño4 index (we found in previous section), tailored to the specific GCM and forecast dataset used in this study.**

· The GCMNiño4 index depends on both the initialization date and the lead time, while the **Niño4 index** are only dependent on the of the Niño 3.4 region for a specific date.

· We are interested to know about relationship between ENSO and some other variable (price) over a longer time period – hence we will use **GCMNiño4** instead of **Niño4.**

· **The GCMNiño4 index is a standardized index that ranges between -1 and 1**

· **Negative** values of the GCMNiño4 index indicate the **La Niña** phase of ENSO, which is characterized by cooler than average temperatures in the eastern and central tropical Pacific Ocean.

· **Positive** values of the GCMNiño4 index indicate the **El Niño** phase of ENSO, which is characterized by warmer than average temperatures in the eastern and central tropical Pacific Ocean.

· **For our econometric model, we use only monthly averaged climate data that are available at the beginning of the month**.

· By using only, the data available at the beginning of the following month, we ensure that our analysis is based on information that was available at the time of the forecast, and not on any information that becomes available after the forecast has been made. This helps to ensure the validity and accuracy of the analysis.

· As we know that GCM Niño4 is tailored to the specific GCM. The **GCM Niño4 indices are subjected to biases in the climate models that generate them**. Therefore, they need to be adjusted or debiased to ensure that they are reliable and accurate representations of the actual climate conditions.

(All the above steps are done, and the data is stored in the MEI.csv)

2.4 Econometric time series forecasting of ENSO

· The aim is to determine the most parsimonious autoregressive (AR) representation of ENSO with additive seasonality and **compare its forecast skill** to that of the GCM forecasts discussed in the previous section.

AR model with additive seasonality for ENSO vs GCM forecasts for the Niño4 index

· The AR model is a statistical model that is fitted to historical ENSO data and is used to make predictions based on the past behavior of the system.

· It takes into account the lagged values of ENSO as well as any seasonal patterns that may be present in the data.

Whereas,

· the GCM forecasts for the Niño4 index are generated using a complex computer model that simulates the behaviour of the Earth's climate system.

· The model takes into account a wide range of physical processes that affect the climate, such as ocean currents, atmospheric circulation, and solar radiation.

Next, we build a AR model with additive seasonality for ENSO

$$E(t) = a^{EE} + \sum_{l=1}^{11} a_l^{EE} D_l(t) + \sum_{k \in \mathbf{I}^{EE}} b_k^{EE} E(t-k), \qquad (3)$$

· Aim is to forecast ENSO using an econometric time series model.

· Assume that the commodity price does not have any impact on ENSO, which is a common assumption in the literature.

· **Equation 3** represents an autoregressive (AR) model for ENSO with additive seasonality.

· The equation starts with the model offset term, which is the baseline ENSO value.

- Then, the equation includes a seasonal term, which captures the additive seasonality of ENSO. Here, Dl(t) is a binary variable that takes the value 1 during month l and 0 otherwise. The coefficient aEE l represents the deviation from the baseline ENSO value during month l.

- Finally, the equation includes a summation term that represents the influence of past ENSO values on the current value.

- Here, lEE is a vector of lags, and bEE k represents the impact of the kth lag on the current ENSO value.

Determining the AR model:

All combinations of lags up to 12 months are assessed building models using all available data from January 1980 to December 2020.

The aim is to find the most parsimonious model, which is the one with the fewest number of parameters while still providing a good fit to the data.

To achieve this, the models are checked for significant serial autocorrelation in the residuals, which indicates that the model is not capturing all of the relevant dynamics in the data.

If any models are found to have significant residual autocorrelations for any lag up to 12 months, they are excluded from consideration.

Likewise, any models that are shown to be heteroskedastic to a 99% confidence level based on Engle tests are also excluded.

Heteroskedasticity occurs when the variance of the errors is not constant over time, which can lead to biased parameter estimates and incorrect inference.

After excluding models with significant residual autocorrelation and heteroskedasticity, the remaining combinations of lags are evaluated based on their ability to **minimize the Bayesian information criterion (BIC).**

The BIC is a criterion for model selection that balances the goodness of fit with the complexity of the model, penalizing models with more parameters. **The model with the lowest BIC value is selected as the most parsimonious model.**

Other information criteria, such as the Akaike information criterion (AIC) and the Hannan-Quinn information criterion (HQIC), **are also evaluated and yield the same result for ENSO**. The selected model is then used to compare forecast skill with the GCM forecasts discussed in the previous section.

2.5 Econometric time series commodity forecasting

$$p(t) = a^{pp} + \sum_{l=1}^{11} a_l^{pp} D_l(t) + \sum_{k \in \mathbf{l}^{pp}} b_k^{pp} p(t-k) + \sum_{k \in \mathbf{l}^{pE}} b_k^{pE} E(t-k), \qquad (5)$$

$$p(t) = (1 + i(t)) \times (1 + \widetilde{p}(t)) - 1,$$

$$\widetilde{p}(t) = \log(P(t)/P(t-1))$$

$$i(t) = \log(C(t)/C(t-1))$$

Section 3: Commodity forecasts

(Note: We are interested in the model with prices as endo and enso as exo factors. Sec 3.1 just forecasts the ENSO. Hence, we shift our interest to 3.2)

3.2 Commodity forecasts

· The authors chose these commodities (coconut oil, palm oil, and soybean oil) because they have been found in previous studies **to have a linear relationship with ENSO** (El Niño-Southern Oscillation).

· The authors **used the log returns of these commodities and the ENSO index as the input variables** to determine the optimal ARIMA model for each commodity.

· Select the endogenous lags of the log returns and exogenous lags of ENSO that minimized the BIC (Bayesian Information Criterion) for each commodity.

· Also check the models for serial correlation and heteroskedasticity. The presented models were found to have no statistically significant serial correlation to a 95% confidence level or heteroskedasticity to a 99% level.

· The models were also found to be stationary based on augmented Dickey–Fuller tests.

· Considered the possibility that these commodities were additionally **dependent upon global macroeconomic growth**, so built more general AR models which also included the log returns of the G7 averaged GDP as an exogenous variable. However, none of these commodities contain GDP at any lag in their most parsimonious form.

Table 1. Autoregressive models of commodity real log returns with exogenous Niño4 factors.

Commodity	Endogenous commodity lags $\mathbf{I}^{pp}$	Exogenous ENSO lags $\mathbf{I}^{pE}$	P_{ARCH}	P_{serial}
Coconut oil	(1,3,12)	(8)	0.663	0.08
Palm oil	(1,2,4,9)	(6)	0.02	0.128
Soybean oil	(1,2,4)	(7)	0.891	0.504