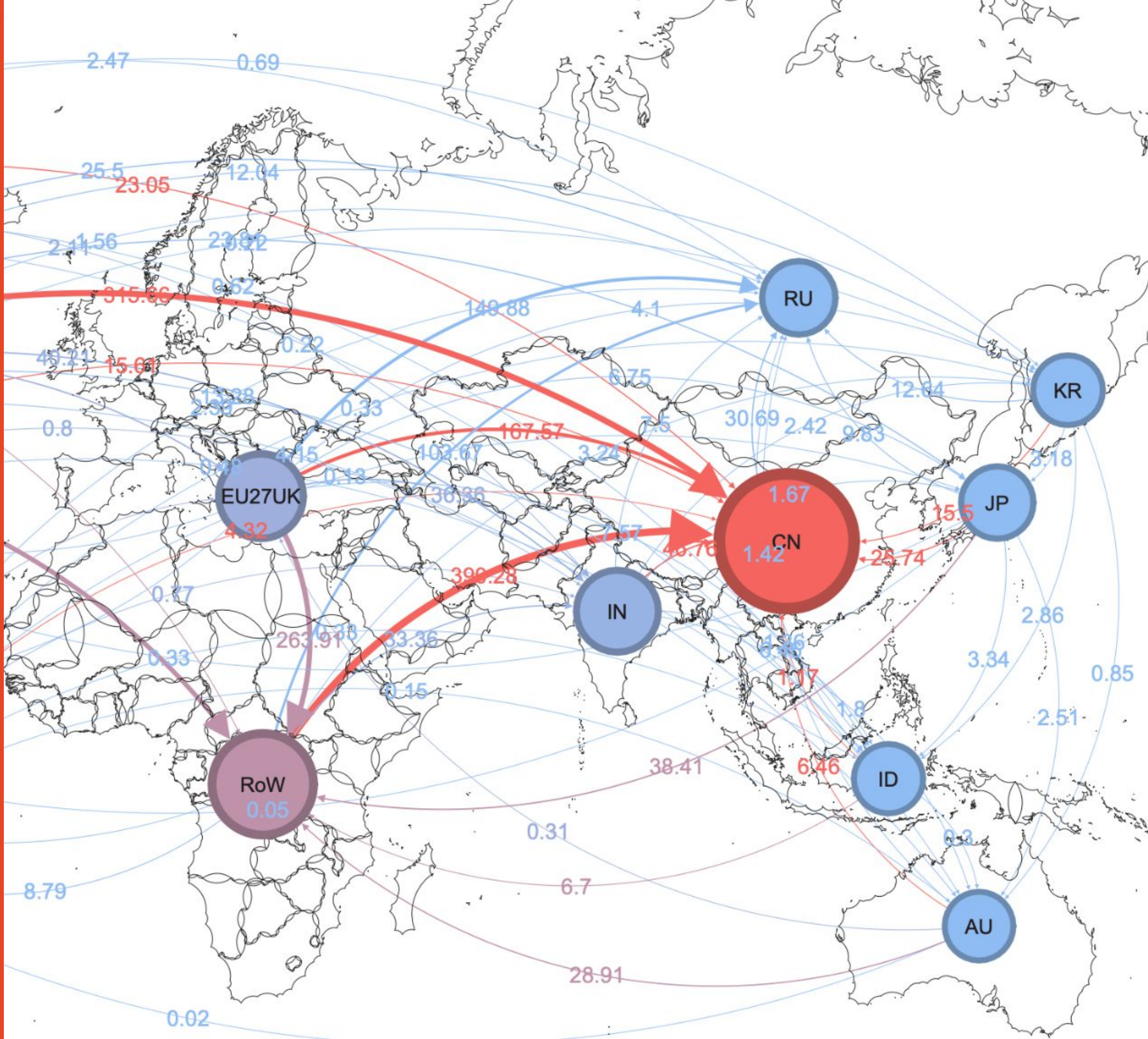
ICLR 2024 Workshop: Tackling Climate Change with Machine Learning

ExioML: Eco-economic dataset for Machine Learning in Global Sectoral Sustainability

Yanming Guo, Jin Ma
yguo0337@uni.sydney.edu.au



Outline

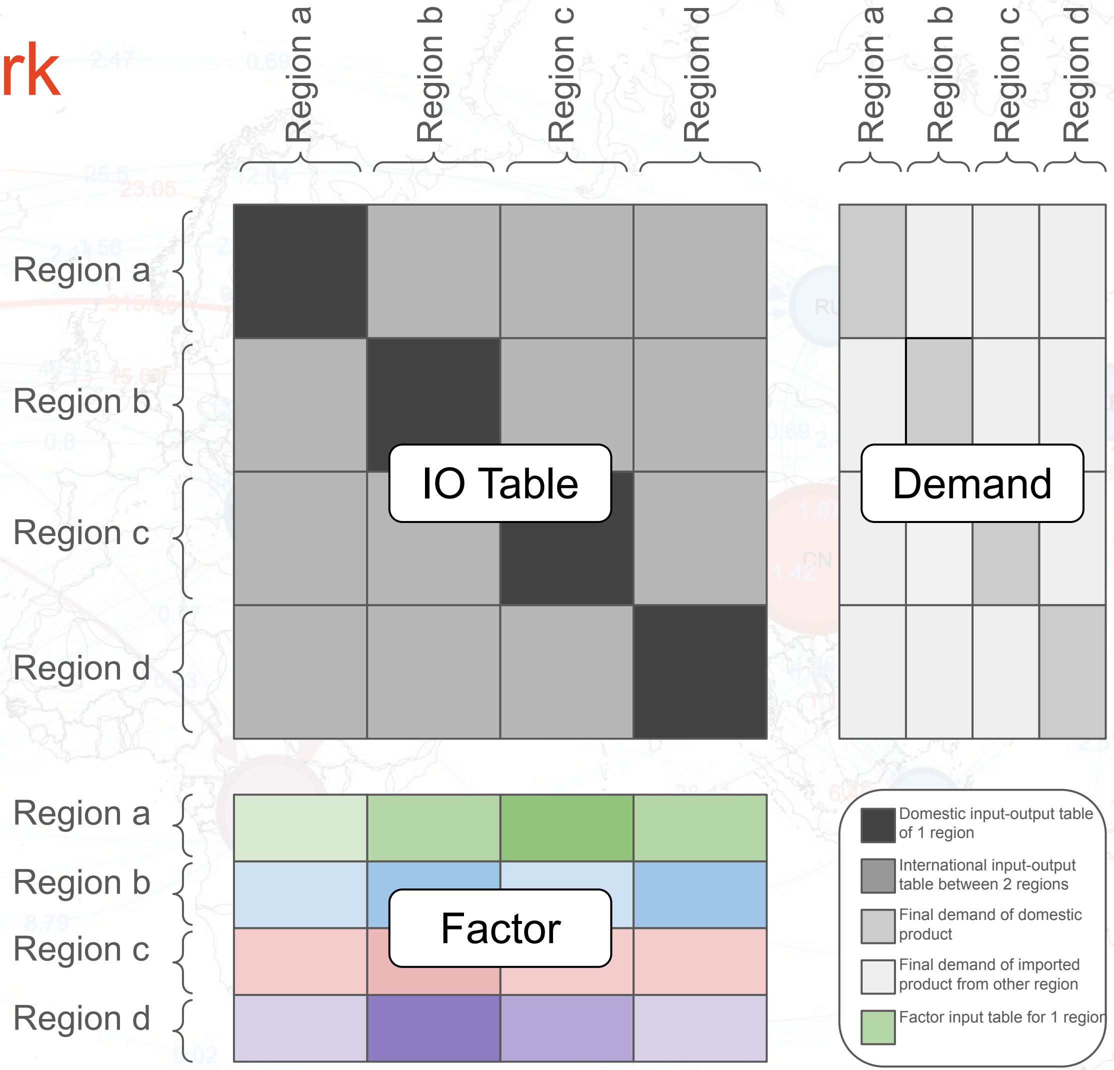
1. Backgrounds
2. Research Gaps
3. ExioML
4. Validation
5. Conclusion

Backgrounds

1. Key Eco-Economic framework is Environmental-Extended Multi Regional Input Output (EE-MRIO)
2. Tracks international resource transfer between international sectors
3. High-dimensional network data in matrix formation
4. Great potential integrating with ML for ***predicting the embodied emission through international trade***, ***estimation of regional sustainability transition***, and the ***topological change of global trading networks based on historical trajectory***

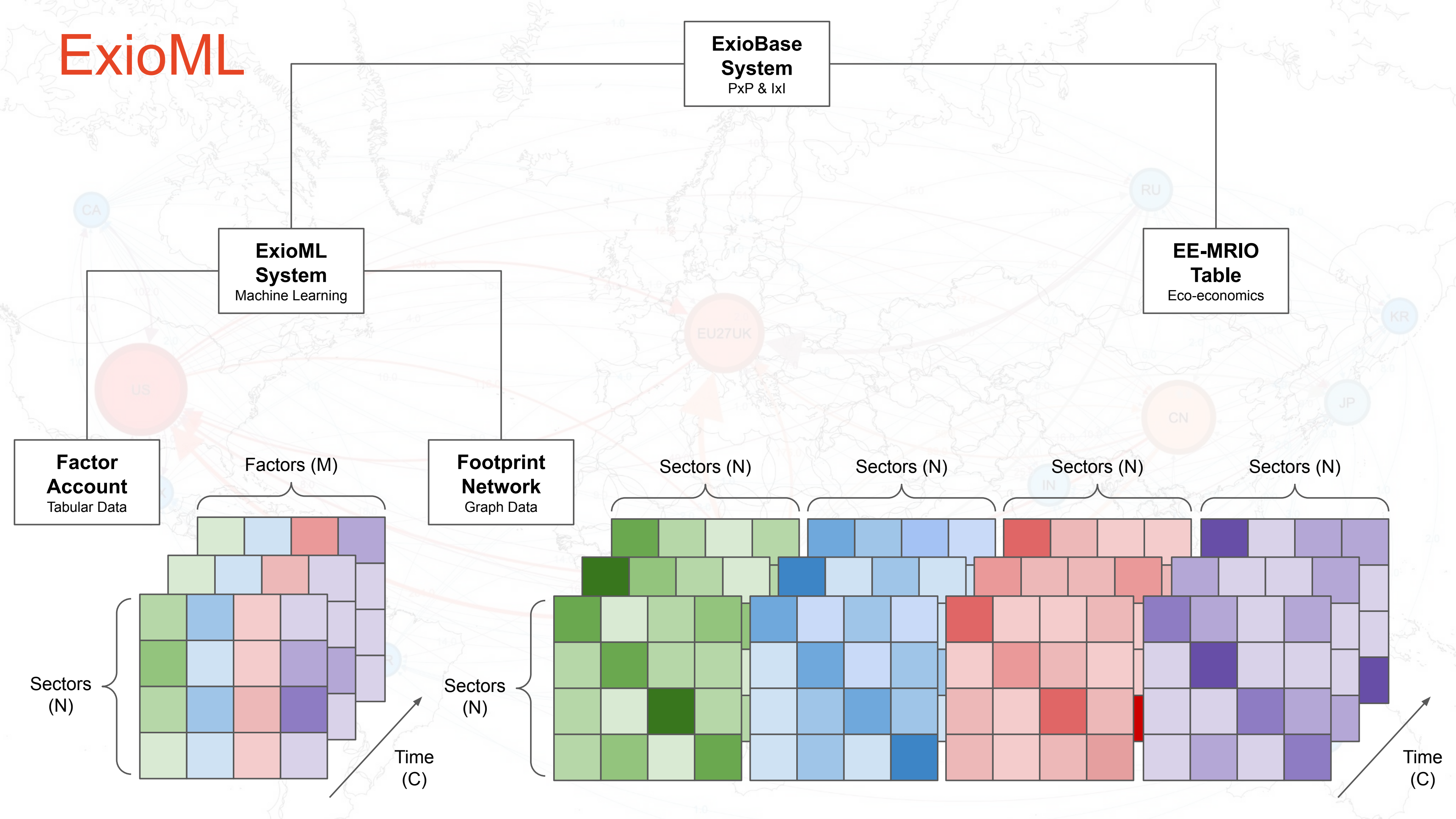
EE-MRIO Framework

1. Input-output Table
2. Demand Table
3. Factor Accounting



Research Gap

1. **EE-MRIO dataset inaccessibility**: Most of the global EE-MRIO (GTAP, Eora) data are closed-source with expensive access and suffer from spatiotemporal resolution (WIOD).
2. **Intensive data pre-processing requires domain knowledge**: ML research requires structured data, and a paradigm shift and data cleaning is required.
3. **Lack of Benchmark datasets and ML models**: ML techniques are less discussed in previous literature. A uniform benchmark dataset should be developed to accelerate the cooperation between ML and Eco-economic research.


ExioML
ExioBase System
PxP & IxI
ExioML System
Machine Learning
EE-MRIO Table
Eco-economics
Factor Account
Tabular Data
Footprint Network
Graph Data
Factors (M)
Sectors (N)
Sectors (N)
Sectors (N)
Sectors (N)
Sectors (N)
Sectors (N)
Time (C)
Time (C)

ExioML Details

Characteristics	**Descriptions**
Component	Factor accounting (Tabular data), Footprint Network (Graph data)
Time frame	Covers 28 annual time steps from 1995 to 2022
Geographical coverage	49 regions (44 countries and 5 rest of the world)
Sectors detail	200 products in PxP, 163 industries in IxI
Key factors	Value added, employment, energy consumption, and GHG emission
Development toolkit	Provide MRIO computation with GPU accelerations for high-level customisation for later researchers

Validation

1. Validated on a Sectoral GHG emission regression task
2. Compare the performance of shallow and deep models
3. Utilise the PxP & IxI factor accounting table in ExioML
4. Achieved low Mean Square Error for 10 runs

	PxP		IxI	
Model	MSE	Time (s)	MSE	Time (s)
KNN	1.071 ± 0.010	0.035 ± 0.001	1.151 ± 0.018	0.026 ± 0.001
Ridge	2.265 ± 0.000	**0.005 ± 0.002**	2.514 ± 0.000	**0.004 ± 0.002**
DT	0.926 ± 0.051	0.316 ± 0.017	0.848 ± 0.070	0.254 ± 0.019
RF	0.356 ± 0.004	21.521 ± 0.157	0.302 ± 0.004	16.511 ± 0.060
GBDT	0.234 ± 0.006	30.276 ± 0.224	0.219 ± 0.007	32.847 ± 0.388
MLP	0.226 ± 0.007	219.218 ± 0.904	0.250 ± 0.092	205.051 ± 24.309
GANDALF	**0.204 ± 0.010**	352.756 ± 7.036	**0.189 ± 0.007**	383.119 ± 3.664
FTT	0.330 ± 0.007	330.578 ± 1.527	0.302 ± 0.023	468.911 ± 7.329

Conclusion

1. ExioML is the first benchmark ML data in EE research
2. Factor accounting table measures relationship between factors
3. Footprint network tracks resource transfer in global trading network
4. ExioML provides high-quality data enabling various ML tasks
5. ExioML development toolkit reduces the barriers for ML researchers
6. Usability is validated by sectoral GHG emission regression task