Introducing RobustiPy: An efficient next generation multiversal library with model selection, averaging, resampling, and explainable AI

September 29, 2025

Daniel Valdenegro Ibarra, **Charlie Rahal**, and Jiani Yan

City+2025@Oxford conference

The plan for today!

We're going to:

- Give a short talk about model uncertainty
- Try to get you up and running with RobustiPy:
  - Simulated examples
  - Empirical examples
- Host a short round-table about the future of model uncertainty tooling more broadly.

RESEARCH ARTICLE | SOCIAL SCIENCES

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Nate Breznau, Eike Mark Rinke, Alexander Wuttke, +162, and Tomasz Żółtak Authors Info & Affiliations

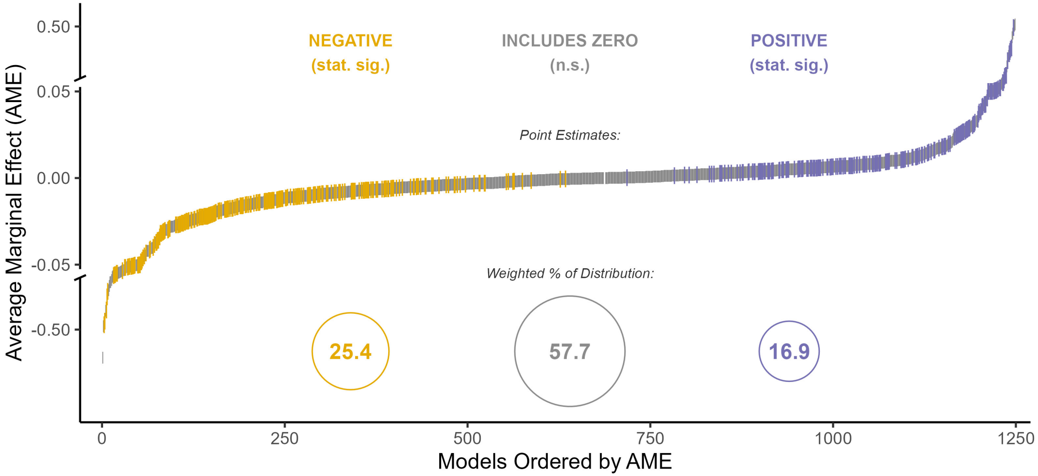
Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

October 28, 2022 | 119 (44) e2203150119 | https://doi.org/10.1073/pnas.2203150119

THIS ARTICLE HAS BEEN UPDATED

THIS ARTICLE HAS BEEN CORRECTED +

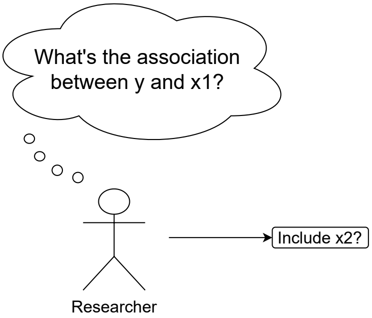
VIEW RELATED CONTENT +

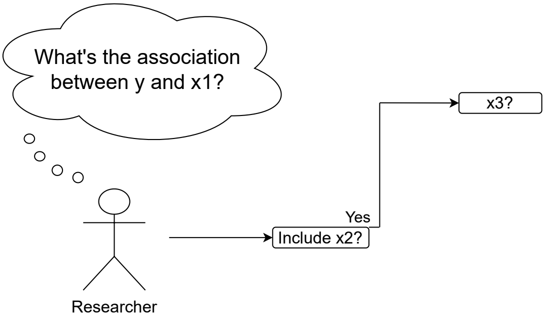

NEGATIVE
(stat. sig.)
INCLUDES ZERO
(n.s.)
POSITIVE
(stat. sig.)
Point Estimates:
Weighted % of Distribution:
25.4
57.7
16.9
Average Marginal Effect (AME)
0.50
0.05
0.00
-0.05
-0.50
0
250
500
750
1000
1250
Models Ordered by AME

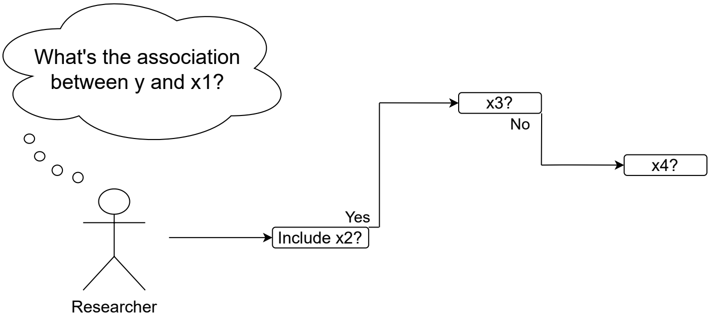
The problem: Types of Model Uncertainty

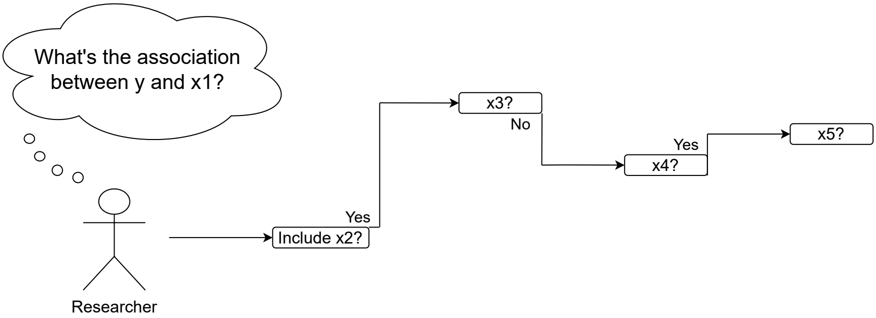
- **Parameter Uncertainty**:
  - Variability due to estimation from limited data.
- **Structural Uncertainty**:
  - Inadequate model to capture system complexities fully.
  - Simplifications that may oversimplify or miss critical aspects.
- **Data Uncertainty**:
  - Errors in data collection or measurement.
  - Outliers affecting model stability or accuracy.
- **Model Formulation Uncertainty**:
  - Choice of mathematical model structure.
  - Assumptions about relationships between variables.

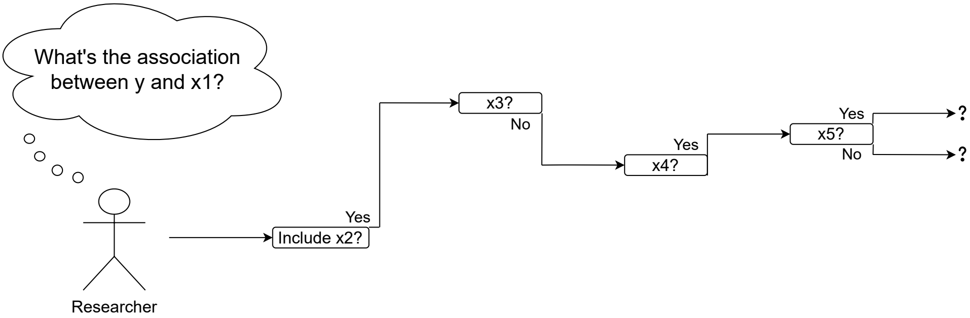
Some types of uncertainty can be handled by applied researchers, some can't.

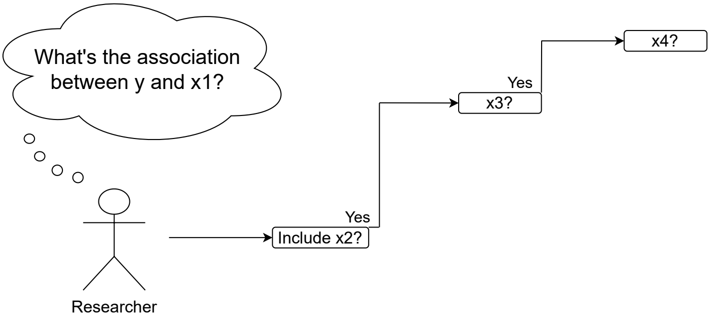

What's the association between y and x1?
Include x2?
Researcher

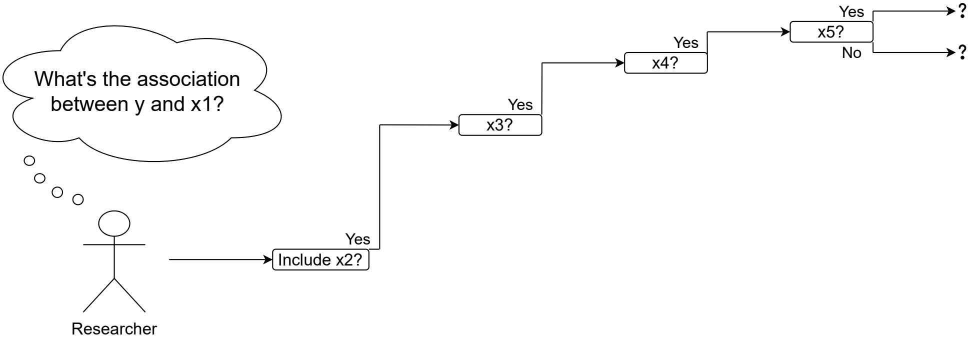

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?

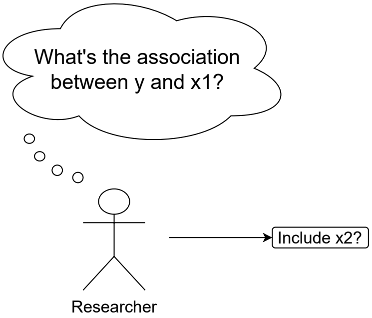

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?
No
x4?

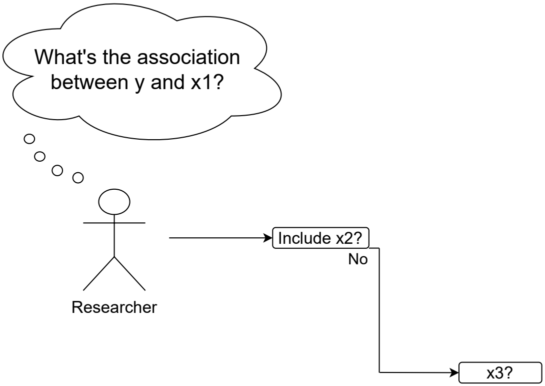

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?
No
x4?
Yes
x5?

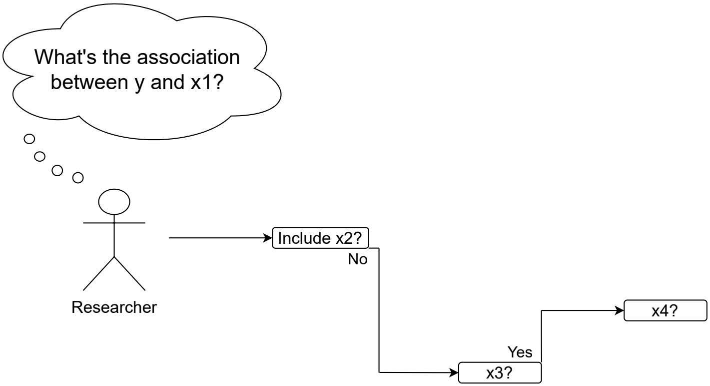

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?
No
x4?
Yes
x5?
Yes
No
?
?

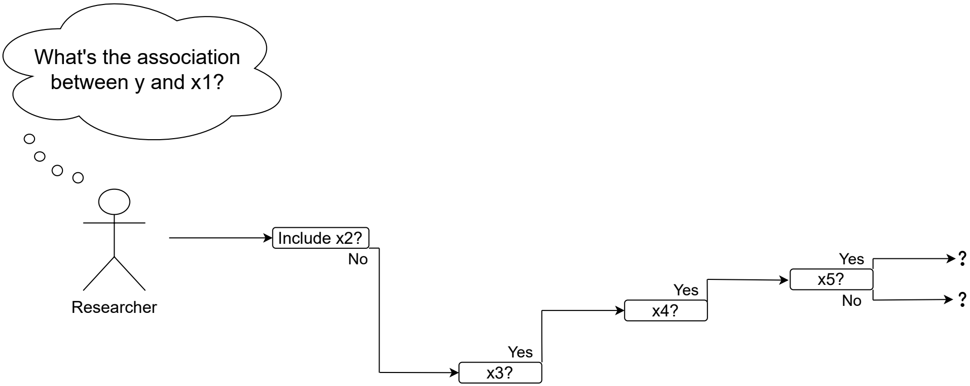

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?
Yes
x4?

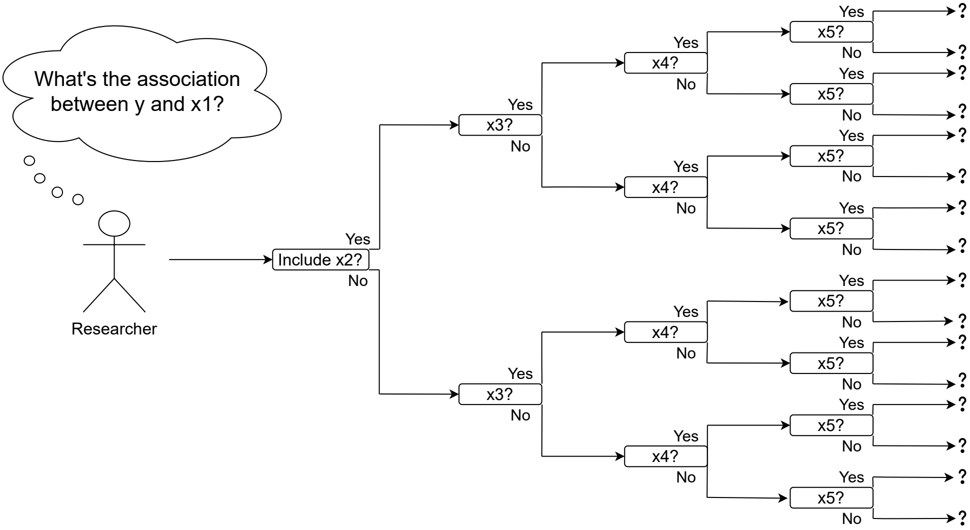

What's the association between y and x1?
Researcher
Include x2?
Yes
x3?
Yes
x4?
Yes
x5?
Yes
No
?
?


What's the association between y and x1?
Include x2?
Researcher


What's the association between y and x1?
Researcher
Include x2?
No
x3?


What's the association between y and x1?
Researcher
Include x2?
No
x3?
Yes
x4?


What's the association between y and x1?
Researcher
Include x2?
No
x3?
Yes
x4?
Yes
x5?
Yes
?
No
?


What's the association between y and x1?
Researcher
Include x2?
Yes
No
x3?
Yes
No
x4?
Yes
No
x5?
Yes
No
?

Introducing RobustiPy

- Multiverse Analysis is ideal to tackle P-Hacking and Harking.
  - Both are big issues for the reproducibility of scientific findings.
- However, multiverse analysis is new: no *established* tool exists to conduct it reliably.
- Additionally, not all researchers are, nor need to be expert programmers.
- So, we set ourselves the task:

Can we create a reliable tool to conduct this/adjacent analysis?

nature human behaviour

View all journals | Search | Log in

Explore content | About the journal | Publish with us

nature > nature human behaviour > resources > article

Resource | Published: 27 July 2020

Specification curve analysis

Uri Simonsohn, Joseph P. Simmons & Leif D. Nelson

Nature Human Behaviour **4**, 1208–1214 (2020) | Cite this article

A Publisher Correction to this article was published on 09 October 2020

This article has been updated

RobustiPy: The formal problem

Consider the key association between variables Y and X, where a set of covariates $\mathbf{Z}$ can influence the relationship between the former as follows:

$$Y = F(X, \mathbf{Z}) + \epsilon. \tag{1}$$

Let's now observe that:

- Usually Y and X are imprecisely defined latent variables.
- Likewise, the set of covariates $\mathbf{Z}$ can also be composed of imprecisely defined variables from which the number of them can be unknown and/or non-finite.
- Finally, $F()$ is an unknown data generating function.

RobustiPy: The formal problem

Let's define the set of *reasonable* operationalisations of Y, X, $\mathbf{Z}$ and $F()$ as $\overleftrightarrow{Y}$, $\overleftrightarrow{Y}$, $\overleftrightarrow{\mathbf{Z}}$ and $\overleftrightarrow{F()}$. Then, we have:

$$\overleftrightarrow{Y}_{k_Y} = \overleftrightarrow{F}_{k_f}(\overleftrightarrow{X}_{k_X}, \overleftrightarrow{\mathbf{Z}}_{k_\mathbf{Z}}) + \epsilon. \tag{2}$$

- Eq. 2 corresponds to a single possible *specification* of the Π set of all possible specifications.
- The total number of specifications can then be calculated as 2^N where N is $n_{\overleftrightarrow{Y}} + n_{\overleftrightarrow{X}} + n_{\overleftrightarrow{\mathbf{Z}}} + n_{\overleftrightarrow{F()}}$.
- For example, an analysis with two functional forms, two target variables, two predictors and five covariates will yield a specification space Π of 2^{10} or 1024 possible specifications.

RobustiPy: The formal problem

We can also split the computation based on the operationalised variables in Eq. 2 by creating a 'reasonable specification space' for each: $\Pi_{\overleftrightarrow{Y}}, \Pi_{\overleftrightarrow{F()}}, \Pi_{\overleftrightarrow{X}}, \Pi_{\overleftrightarrow{Z}}$. The whole specification space can be obtained again as follows:

$$\Pi = \Pi_{\overleftrightarrow{Y}} \times \Pi_{\overleftrightarrow{F()}} \times \Pi_{\overleftrightarrow{X}} \times \Pi_{\overleftrightarrow{\mathbf{Z}}} \tag{3}$$

- Any random and independently selected sample π from Π would lead to a reasonable approximation of $Y = F(X, \mathbf{Z})$.
- The main problem with current research practice is that the sample of Π reported by the researchers is not random.
- **The goal of RobustiPy is to automate this process to make it as easy as possible.**

Vanilla Computation

- Baseline case: single outcome Y, single predictor X, and varying controls Z.
- Data generating process:

$$Y = \beta X + \gamma Z + \epsilon$$

- Specification space reduces to $\Pi_{\overleftrightarrow{Z}}$ only.
- Estimation: OLS with k-fold CV and bootstrap resampling.
- Outputs: distribution of $\hat{\beta}$ across $2^{|Z|}$ models, with null vs full model highlighted.

Array of Never-Changing Variables

- Some predictors are fixed (always included in $\overleftrightarrow{X}$).
- Reduces computational complexity by shrinking the specification space.
- Empirical motivation: theoretically indispensable covariates.
- Specification space:

$$\Pi = \Pi_{\overleftrightarrow{Z}} \quad \text{with fixed } X_f$$

- Results: interpretable estimates with reduced variance and runtime.

Fixed Effects OLS

- Designed for panel/longitudinal data with grouping variable g.
- Within-transformation removes group-specific heterogeneity:

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + \epsilon_{it}$$

- Implemented by demeaning variables by g.
- $\Pi_{\overleftrightarrow{Z}}$ still varied, but estimation is within-groups.
- Useful for causal inference in repeated-measures data.

Binary Dependents

- Logistic regression estimator for binary $Y \in \{0, 1\}$.
- Model:

$$\Pr(Y = 1 \mid X, Z) = \frac{1}{1 + \exp(-(\beta X + \gamma Z))}$$

- $\Pi_{\overleftrightarrow{Z}}$ explored as before, with log-likelihood fit.
- Supports odds ratio interpretation of $\hat{\beta}$.
- Alternative: linear probability model with OLSRobust.

Multiple Dependents

- Allows multiple operationalisations of Y.
- Construct composites by averaging z-scored outcomes:

$$Y_{i,S}^{\mathrm{comp}} = \frac{1}{|S|} \sum_{j \in S} \tilde{Y}_i^{(j)}$$

- Specification space enlarges: $\Pi_{\overleftrightarrow{Y}} \times \Pi_{\overleftrightarrow{Z}}$.
- Provides robustness to measurement uncertainty in the outcome.
- Outputs: specification curves across all dependent definitions and composites.

README | Code of conduct | Contributing | GPL-3.0 license

RobustiPy

Purpose: Research | Python: 3.11 | License: GNU3.0 | DOI: 10.5281/zenodo.15700698

Welcome to the home of `RobustiPy`, a Python library for the creation of a more robust and stable model space. RobustiPy does a large number of things, included but not limited to: high dimensional visualisation, Bayesian Model Averaging, bootstrapped resampling, (in- and)out-of-sample model evaluation, model selection via Information Criterion, explainable AI (via SHAP), and joint inference tests (as per Simonsohn et al. 2019).

Full documentation is available on Read the Docs. The first release is indexed onto Zenodo here.

`RobustiPy` performs Multiversal/Specification Curve Analysis which attempts to compute most or all reasonable specifications of a statistical model, understanding a specification as a single attempt to create an estimand of interest, whether through a particular choice of covariates, hyperparameters, data cleaning decisions, and so forth.

Introducing RobustiPy: An efficient next generation multiversal library with model selection, averaging, resampling, and explainable AI

Daniel Valdenegro Ibarra[1,3], Jiani Yan[1,2,3,4], Duiyi Dai[1,3], and Charles Rahal[1,3,5]

[1]Leverhulme Centre for Demographic Science, University of Oxford
[2]Department of Sociology, University of Oxford
[3]ESRC Centre for Care, University of Oxford
[4]Wolfson College, University of Oxford
[5]Nuffield College, University of Oxford

September 29, 2025

Abstract

Scientific inference is often undermined by the vast but rarely explored "multiverse" of defensible modelling choices, which can generate results as variable as the phenomena under study. We introduce `RobustiPy`, an open-source Python library that systematizes multiverse analysis and model-uncertainty quantification at scale. `RobustiPy` unifies bootstrap-based inference, combinatorial specification search, model selection and averaging, joint-inference routines, and explainable AI methods within a modular, reproducible framework. Beyond exhaustive specification curves, it supports rigorous out-of-sample validation and quantifies the marginal contribution of each covariate. We demonstrate its utility across five simulation designs and ten empirical case studies spanning economics, sociology, psychology, and medicine—including a re-analysis of widely cited findings with documented discrepancies. Benchmarking on ~672 million simulated regressions shows that `RobustiPy` delivers state-of-the-art computational efficiency while expanding transparency in empirical research. By standardizing and accelerating robustness analysis, `RobustiPy` transforms how researchers interrogate sensitivity across the analytical multiverse, offering a practical foundation for more reproducible and interpretable computational science.

Keywords: *Model Uncertainty, Statistical Software, Open Science*

Simulated Example Time!

- A link to the first simulation can be found **here**!
  - This is a simple example which has three control variables and uses OLSRobust.
- A link to the second simulation can be found **here**!
  - This uses grouping data.
- A link to the third simulation can be found **here**!
  - This uses an array of x variables grouped together.

Empirical Example Time!

- A link to the first empirical example can be found **here**!
  - This is a simple example which uses the canonical union.dta dataset.
- A link to the second empirical example can be found **here**!
  - This uses panel data.
- A link to the third empirical example can be found **here**!
  - This estimates a problem with a logistic regression.

Discussion Questions 1: Defining Model Uncertainty

- How do we currently define and quantify model uncertainty in academic research?
- What types of uncertainty does RobustiPy capture?
- How does the necessity of capturing such things vary by academic field?
- What are the major challenges researchers face?
  - Are they more about accurate modelling, or communicating uncertainty?

Discussion Questions 2: Impact on Research Validity and Reproducibility

- How does model uncertainty impact the validity and reproducibility of research findings?
- What steps can researchers take to mitigate the effects of uncertainty on their results?
- How should researchers communicate model uncertainty?
- What role does transparency in uncertainty play in the credibility of academic research?

Discussion Questions 3: Educational and Training Needs

- What is the current status of teaching model uncertainty?
- What is needed to improve understanding and management of uncertainty?
- What role do open-source frameworks and collaborative platforms play?
- How commonly utilised are Open Science tools and ideas in *your* discipline?
  - How widely are *they* taught?

Discussion Questions 4: Ethical Considerations

- How trusted are academic ‘experts’ right now in general?
- Are there ethical implications related to how model uncertainty is managed and communicated in academic research?
- Is RobustiPy a tool that can be used for evil, as well as good?
- How should uncertainty be addressed when research findings influence policy or public perception?

Discussion Questions 5: Impact of Emerging Technologies

- How might emerging technologies (e.g., quantum computing, advanced simulations) impact our ability to handle and reduce model uncertainty?
- How can advancements in AI and machine learning contribute to improving uncertainty estimation?
- How can interdisciplinary collaboration drive innovation in uncertainty modelling and management?
- What are the potential risks and opportunities associated with these technologies in uncertainty modelling?

Thank You!

Thank you so much all for coming!