Abstract

- **Main Thesis**: Traditional philosophical analyses focused on trace-based reasoning in paleosciences inadequately capture the essential role of modeling.
- **Key Points**:
  - Models inform both actual history and possible scenarios.
  - Phenomena-driven modeling ensures relevance by aligning closely with historical data.
  - Emphasizes that historical sciences rely on understanding possibilities (modality) alongside actual events.

1. Introduction

- **Context**: Paleoscientists face epistemic scarcity due to extensive temporal scales and information decay.
- **Argument**: Current philosophical accounts prioritize trace-based reasoning but overlook the pervasive use of models.
- **Objective**: Highlight the critical role of models in understanding the deep past, focusing on their ability to explore possibilities.

2. Models & Traces

2.1. Trace-Based Reasoning

- **Definition**: Inferring past events from traces (fossils, tracks) using theories of trace formation.
- **Philosophical Accounts**:
  - **Carol Cleland**: Emphasizes "smoking gun" reasoning—correlations, hypothesis generation, and trace discrimination.
  - **Derek Turner**: Argues historical knowledge is underdetermined due to trace decay and lack of intervention capabilities.
- **Critique**: Trace-based reasoning alone doesn't account for the extensive use of modeling in paleosciences.

2.2. Modelling as Strategy

- **Philosophical Perspectives**:
  - **Weisberg & Godfrey-Smith**: Models as indirect strategies to understand target systems through proxies.
- **Case Study**:
  - **MBL Model**: Stochastic branching model used to identify macroevolutionary patterns beyond randomness.
- **Contrast**: Traditional modeling is more abstract, whereas paleoscientific modeling is often closely tied to empirical data.

2.3. Philosophy of Paleoscientific Modeling

- **Applications**:
  - **Functional Morphology**: Simulations to infer locomotion of extinct species.
  - **Archaeology**: Models to interpret cultural patterns and integrate legacy data.
  - **Paleoclimatology**: Computational models to reconstruct climatic history.
- **Philosophical Insights**: Models act as epistemic tools that correct, refine, and integrate data, bridging actual and possible historical scenarios.

3. Two Models of the Deep Past

3.1. Dinosaur Tails

- **Study**: Ibrahim et al. (2020) tested propulsion efficiency of Spinosaurus' tail using physical models.

- **Findings**: Spinosaurus' tail generated significantly more thrust, supporting its adaptation for aquatic pursuit.
- **Philosophical Implication**: Demonstrates models testing capacity hypotheses, establishing possibilities about extinct species' behaviors.

3.2. Avalonian Ecological Communities
- **Study**: Mitchell et al. (2019) used Spatial Point Process Analysis (SPPA) to analyze Ediacaran taxa distribution.
- **Findings**: Avalonian communities were predominantly structured by neutral (stochastic) processes, differing from modern niche-structured ecosystems.
- **Philosophical Implication**: Illustrates phenomena-driven modeling, ensuring models are directly informed by empirical data to explore specific historical phenomena.

4. Modality & Modeling in Historical Science

4.1. Possible Pasts
- **Argument**: Historical sciences require understanding possibilities to reconstruct the past.
- **Role of Models**: Explore capacity hypotheses and regularities, complementing trace-based reasoning.

4.2. Phenomena-Driven Modeling
- **Definition**: Models are iteratively refined based on empirical phenomena, ensuring relevance.
- **Contrast with Theory-Driven**: Emphasizes direct alignment with data rather than exploring abstract theories.
- **Benefits**: Enhances model relevance and epistemic robustness by grounding possibilities in empirical evidence.

5. Conclusion
- **Summary**: Models are indispensable for both inferring actual events and exploring possible scenarios in historical sciences.
- **Phenomena-Driven Approach**: Ensures models are tightly integrated with empirical data, enhancing their explanatory power.
- **Philosophical Implications**: Calls for expanded philosophical frameworks that recognize the nuanced role of modeling alongside trace-based reasoning in historical sciences.

---

1. Introduction
- **Philosophical Gap**: Highlights the lack of focus on data models in the philosophy of science.
- **Objective**: Examines how paleontologists construct and correct paleodiversity data models from fossil records.
- **Key Theses**:
  - Fidelity, not purity, determines epistemic reliability.
  - Fidelity is gradual and can be enhanced through model-based corrections.
  - Data models should be assessed based on their adequacy for specific purposes.

2. Historical Context of Paleodiversity Data
- **Early Recognition of Biases**:
  - Charles Lyell and Charles Darwin identified biases and incompleteness in the fossil record.
  - Darwin's concerns about gaps undermining evolutionary theory.
- **Paleobiological Revolution**:
  - 1970s shift towards quantitative methods and computer simulations in paleontology.
  - Jack Sepkoski's contributions, including the Sepkoski curve and large fossil databases.

- **Approaches to Reading the Fossil Record**:
  - Optimistic (literal) reading vs. corrected (generalized) approaches using models.

3. Contemporary Methods for Correcting Paleodiversity Data

- **Subsampling Approaches**:
  - **Classical Rarefaction**: Assumes uniform sample sizes; identified as inadequate by Alroy.
  - **Shareholder Quorum Subsampling (SQS)**: Tracks data coverage, improving on classical methods.
- **Residuals Approaches**:
  - Separates biological signals from geological biases by modeling and subtracting bias effects.
  - Tested through simulations showing improved accuracy over raw data.
- **Phylogenetic Approaches**:
  - Uses cladistic analysis to infer and fill gaps (ghost lineages, Lazarus taxa).
  - Reliability assessed via simulations, showing generally better performance than raw data.

4. Philosophical Implications: Fidelity and Adequacy-for-Purpose

- **Purity vs. Fidelity**:
  - Purity (unprocessed data) is less important than fidelity (accuracy in representing the signal).
- **Fidelity as a Degree**:
  - Data models vary in how well they capture the relevant biological signal.
- **Vicarious Control**:
  - Models improve data fidelity post-collection by correcting distortions and noise.
- **Adequacy-for-Purpose**:
  - Data models must be evaluated based on their suitability for specific scientific hypotheses and purposes.

5. Data Models Across Hierarchical Levels

- **Prepared Fossil Specimens**:
  - Fossil preparation involves model-based decisions to enhance data utility.
  - Preparators distinguish between signal and noise based on theoretical needs.
- **Model-Data Symbiosis**:
  - Models and data are mutually dependent, enhancing each other's scientific utility.
  - Emphasizes the iterative nature of improving data through modeling.

6. Conclusion

- **Summary of Philosophical Themes**:
  - Fidelity over purity, incremental improvement of data, vicarious control, and adequacy-for-purpose.
- **Significance of Model-Based Corrections**:
  - Demonstrates how models enhance the reliability and utility of paleodiversity data.
- **Broader Insights**:
  - Provides a framework for understanding data models' role and evaluation in scientific practice.
- **Final Thoughts**:
  - Highlights the ingenuity of scientists in addressing data incompleteness and bias through sophisticated modeling techniques.

7. Acknowledgements and References

- **Acknowledgements**: Credits contributors and supporters.
- **References**: Comprehensive list of sources cited, underpinning the paper's arguments and methodologies.