UKCEH

Lancaster University

SEMANTIC DECOMPOSITION OF ENVIRONMENTAL OBSERVED PROPERTIES · MSc PLACEMENT RESEARCH

Making environmental measurements machine-comparable

Semi-automatic decomposition of observed-property descriptions into the I-ADOPT semantic framework, tested on environmental sources the methods had never seen

Omar Elaktash · MSc Data Science & Artificial Intelligence, Lancaster University · **Placement project hosted by the UK Centre for Ecology & Hydrology**

WHY THIS MATTERS FOR ENVIRONMENTAL DATA

Comparable metadata

Structured variable descriptions let monitoring datasets be searched, joined and compared across programmes.

Interoperable and reusable

Machine-interpretable observed properties underpin interoperable, reusable environmental data.

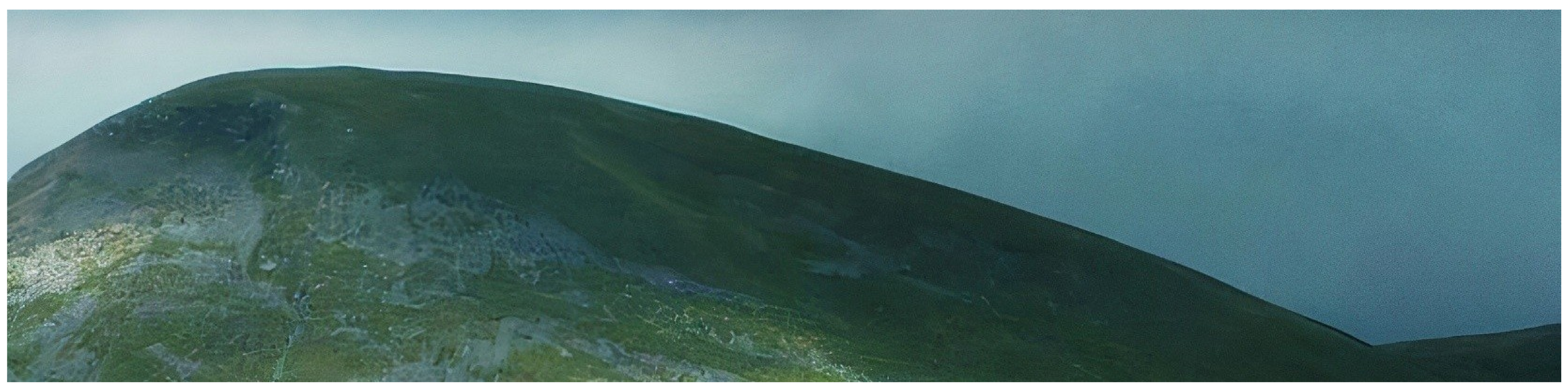
Less manual annotation

Semi-automatic decomposition could reduce the expert effort needed to describe large variable collections.

01 THE PROBLEM

One short phrase, several meanings

A label like **"daily mean soil temperature at 50 cm"** is obvious to a scientist. Software cannot tell which part is the property, which is the entity observed, and which is only a condition, so comparable measurements can be difficult for software to match across catalogues.



02 THE I-ADOPT FRAMEWORK

One label, five semantic roles

"soil temperature at 50 cm depth"	
Property	**temperature**
Object of Interest	**soil**
Matrix	*leave empty*
Constraint	**50 cm depth**
Statistical Modifier	*leave empty*

The hard part is putting each concept in the correct role while leaving unsupported roles empty.

I-ADOPT = InteroperAble Descriptions of Observable Property Terminologies, a Research Data Alliance recommendation.

03 FIVE METHODS COMPARED

From transparent rules to contextual AI

Method 1 · Rule-based baseline
Transparent rules. Auditable, but limited reach.

Method 2 · Vocabulary & ontology matching
Reviewed scientific vocabularies. Precise, low coverage.

Method 3A · Sentence embeddings (MPNet)
Local model. Broad retrieval, weak transfer.

Method 3B · Hosted large language model (LLM)
Frozen source-only prompt. Flexible, must be reviewed.

Method 3C · Sentence embeddings (BGE)
Second local route. Useful control, narrow reach.

04 A FAIR TEST

Development and the final test kept apart

DEVELOPMENT CORPUS

89 clean-core labels from 5 datasets: Auchencorth meteorology; ECN precipitation, stream-water & soil-solution chemistry; COSMOS-UK soil/hydrometeorology.

FROZEN ↓

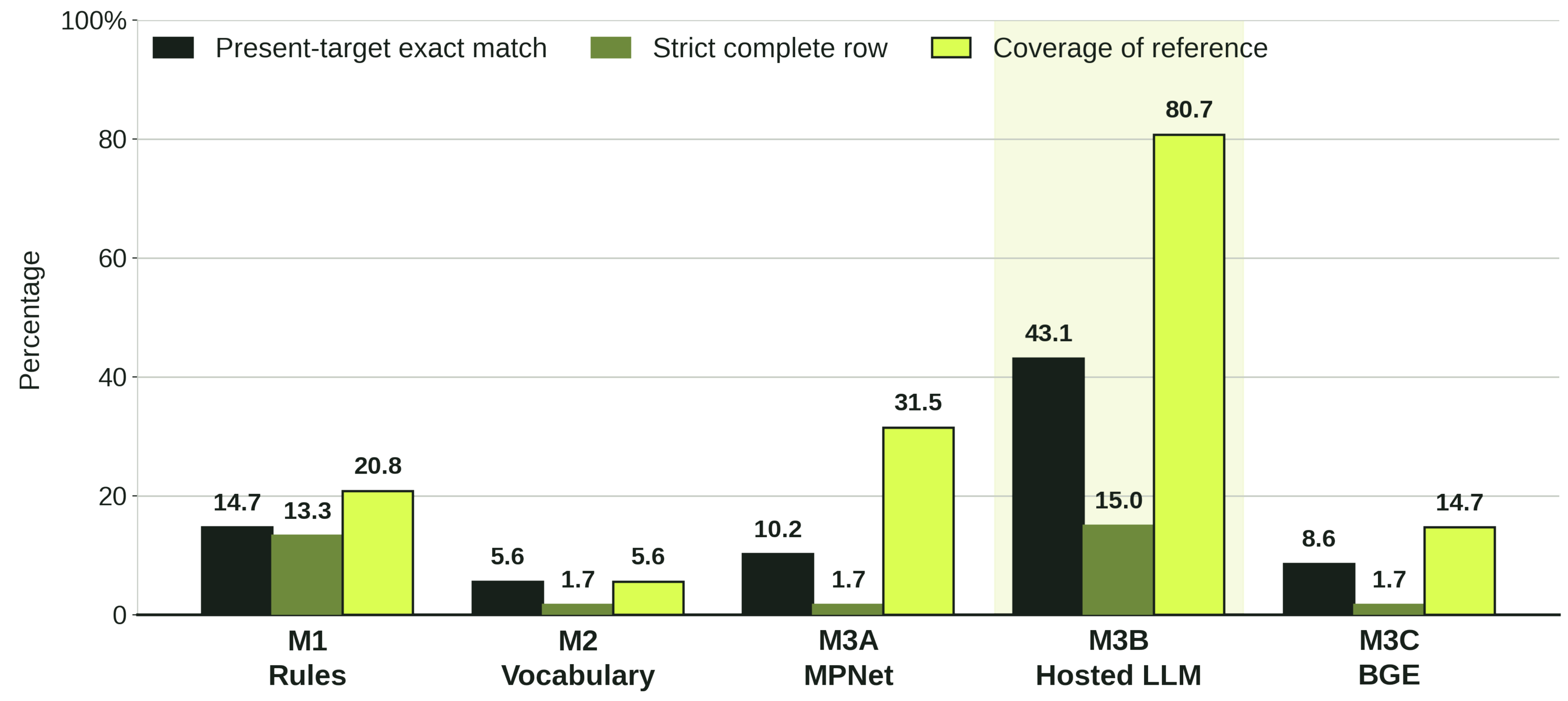
INDEPENDENT TEST CORPUS · UNSEEN SOURCES

60 labels from 3 unseen datasets: 18 Siksik meteorology/hydrology; 18 gannet egg contaminants & shell index; 24 otter/environmental chemistry. Blind annotation, host review, then a frozen 197-component reference.

Nothing from the test was used to tune any rule, mapping, threshold or prompt.

05 WHAT HAPPENED ON UNSEEN DATA

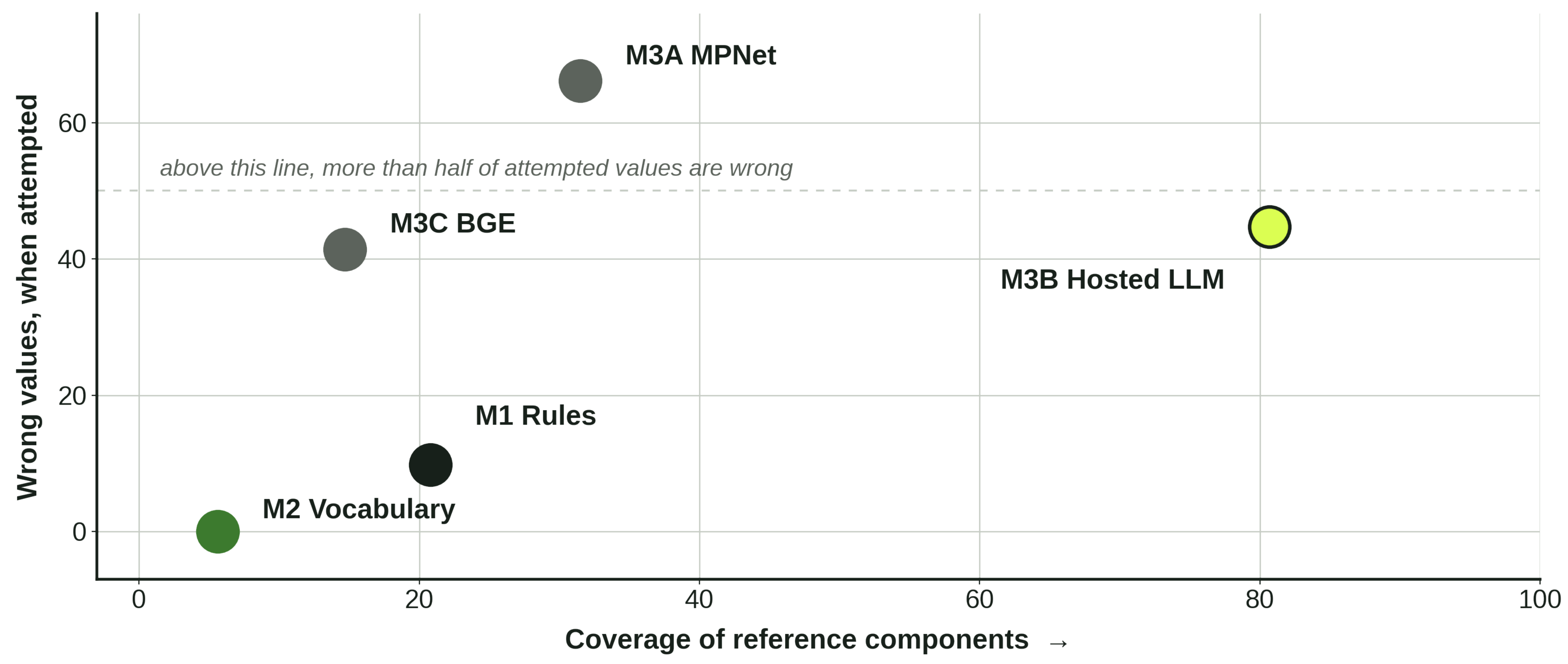
Five frozen methods, one run each



The hosted LLM led on exact match and coverage. The transparent rule baseline still produced 8 complete rows out of 60, only one fewer.

06 COVERAGE ≠ RELIABILITY

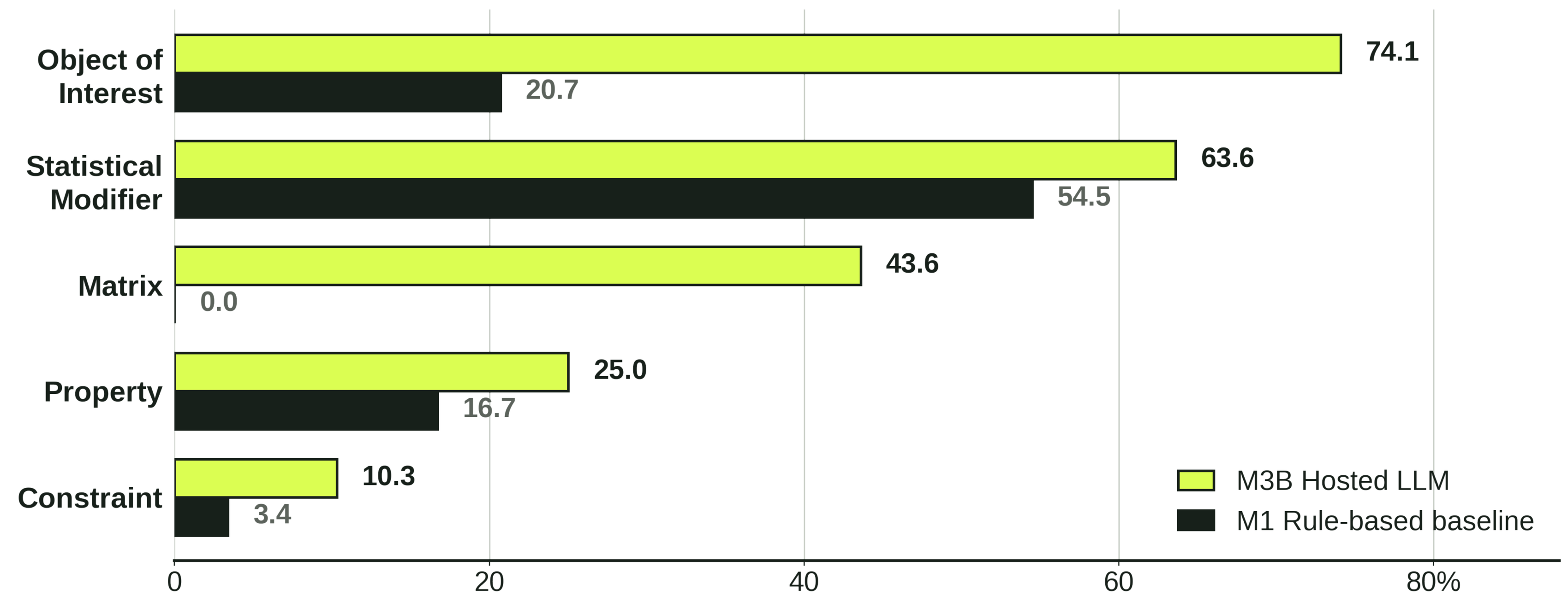
More answers did not mean better answers



Among the reference components covered by the hosted LLM, 44.7% had the wrong exact value. Vocabulary matching covered only 5.6%, but nothing eligible was wrong.

07 WHICH ROLES WERE HARD

Entities were easier than exact roles



Naming the entity observed was easiest. Placing the qualifying condition in the Constraint role was hardest.

08 AND WHICH SOURCES

Every complete strict row came from one source family. The chemistry-heavy sources were markedly harder.

Method references. Magagna et al. (2022) · Coussot et al. (2024) · Wilkinson et al. (2016) · Reimers & Gurevych (2019) · Song et al. (2020).
Data sources. Auchencorth: Centre for Ecology and Hydrology et al. (2017) · ECN chemistry: Rennie et al. (2017a-c; 2020) · COSMOS-UK: Smith et al. (2024) · Siksik: Tetzlaff et al. (2016a) · Gannet eggs: Crosse et al. (2012) · Otter/environmental chemistry: Brand et al. (2019).

*The evidence supports a bounded, provenance-aware, **human-in-the-loop assistance workflow**: rule and vocabulary evidence first, contextual candidates for what remains, and expert review before publication.*

With thanks to Dr Helen Rawsthorne (UKCEH), Dr Matthew Coole (UKCEH) and Dr Alistair Baron (Lancaster University).

Frozen TEST001-TEST060 · 197 non-blank reference components · strict I-ADOPT comparison policy. 43.1% is present-target component exact match, not general accuracy. Expert-reviewed evaluation reference, not a universal gold standard.