

Georeferencing Non-Gazetteered Place Names Using Biological Specimen Records

Aneesha Fernando 

School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Surangika Ranathunga 

School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Kristin Stock 

School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

Raj Prasanna 

Joint Centre for Disaster Research, Massey University, Wellington, New Zealand

Christopher B. Jones 

School of Computer Science and Informatics, Cardiff University, UK

Abstract

Biological specimen records collected by natural history institutions constitute a rich source of temporal geographic knowledge, capturing biodiversity information about regional landscapes as they were recorded at different times. Using digitised data from the Allan Herbarium (New Zealand), this study identifies place names in these specimen locality descriptions that are absent from current gazetteers; we refer to these as non-gazetteer place names (NGPs). These place names are typically historical, vernacular, or colloquial and were used as landmarks to describe a specimen's location at the time of collection. We then investigate the problem of georeferencing the NGPs using only the limited information available in the specimen records. To resolve this, we leverage repeated occurrences of the same place name across specimen records with different specimen locations and spatial relation terms, extracting and inverting these relations to derive constraints on NGP locations. This approach is instantiated within deterministic, probabilistic, and LLM-based methods, enabling a comparative analysis of their strengths and limitations for text-based spatial inference. On a pseudo-NGP benchmark, probabilistic inference achieves the highest accuracy (median error 1.43 km; A@1 km 36%), while the LLM yields competitive but less precise estimates (median error 1.80 km; A@1 km 31%), indicating that, despite advances in LLMs, traditional modelling remains advantageous when high spatial precision is required.

2012 ACM Subject Classification Information systems → Geographic information systems; Computing methodologies → Natural language processing; Mathematics of computing → Probabilistic inference problems

Keywords and phrases Non-gazetteered Place Names, Gazetteers, Toponyms, Georeferencing, Spatial Relation Modelling, Large Language Models

Digital Object Identifier 10.4230/LIPIcs.COSIT.2026.11

Supplementary Material *Dataset:* <https://doi.org/10.6084/m9.figshare.32397048>

Funding This work was supported by the Ministry of Business Innovation and Employment Smart Ideas Fund (grant number MAUX2104), New Zealand.

1 Introduction

The content and metadata associated with biological specimen collection records constitute a rich and underutilised source of geographic information. These records usually include locality descriptions that are routinely recorded digitally by natural history institutions, including museums and herbaria [6, 28]. A typical locality description specifies the location where a



© Aneesha Fernando, Surangika Ranathunga, Kristin Stock, Raj Prasanna, and Christopher B. Jones; licensed under Creative Commons License CC-BY 4.0

17th International Conference on Spatial Information Theory (COSIT 2026).

Editors: Sabine Timpf, Gabriele Filomena, Armand Kapaj, Rui Zhu, Nicholas Giudice, and Ed Manley;

Article No. 11; pp. 11:1–11:22



Leibniz International Proceedings in Informatics
LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

biological specimen was collected and usually contains at least one place name, often several, along with spatial relation terms that describe the specimen’s position relative to those places. Such records span more than a century and capture geographic information as it was recorded at the time of collection. Consequently, some of the place names recorded in these descriptions are not represented in current gazetteers and can be classified as non-gazetteer place names (NGPs). We define NGPs as place names that are entirely absent from existing gazetteers, including alternative names and other variants, and therefore have no associated spatial footprint in those gazetteers. Their absence may stem from temporal changes in toponymy, the use of informal or locally recognised names, or incomplete coverage of certain regions and feature types in existing gazetteers [14]. Given that millions of such records are available worldwide, these descriptions represent a significant and largely untapped source of geographic knowledge. Incorporating these place names, together with their inferred spatial footprints, into digital gazetteers has the potential to improve spatial reference coverage and support more robust place name resolution, not only for specimen georeferencing but also for broader geographical information retrieval applications and analyses.

In this research, we focus on georeferencing NGPs using information available within biological specimen records. The information contained in biological specimen records includes free-text locality descriptions indicating where specimens were collected, coordinates of collection locations, habitat descriptions, and higher-level administrative units such as states or provinces.¹ These elements are used jointly for both georeferencing and disambiguation of place names in locality descriptions, with specimen coordinates serving as important spatial anchors that constrain the interpretation of relative spatial relations and reduce uncertainty in the inferred locations of unknown place names.

Biological specimen collection practices further provide a source of spatial redundancy that can prove valuable. During fieldwork, collectors typically traverse a specific route and record multiple specimens collected within a limited geographic area. As a result, many specimens share common geographic features and place names across their locality descriptions. In modern collection workflows, specimen coordinates are often recorded in the field and subsequently validated by georeferencers. In historical datasets, however, coordinates may also have been assigned retrospectively from locality descriptions, and as a result, some uncertainty may be present in the coordinates. Moreover, a substantial proportion of records, particularly those collected prior to the widespread adoption of GPS, remain ungeoreferenced, and existing georeferences vary in quality [21]. To address this gap, a range of georeferencing methods has been developed, from time-intensive manual interpretation using GIS tools and maps [28] to automated approaches based on rule-based techniques [26], machine learning [22], and modern Large Language Models (LLMs) [10]. Building on this context, our methodology leverages digitised specimen records and exploits the repeated occurrence of the same NPG across multiple locality descriptions as a source of spatial evidence. When an unknown place name appears in several descriptions, the spatial constraints expressed in each record are combined to infer its most plausible location.

Table 1 illustrates this pattern of repeated place name occurrence across multiple locality descriptions. The NPG *Cabstand*² appears in four independent locality descriptions, each linked to a distinct specimen and to the coordinates of the corresponding specimen location.

¹ Example Allan Herbarium specimen record: <https://scd.landcareresearch.co.nz/Specimen/CHR%20508209>.

² The publication “Banks Peninsula Conservation Walks” [9], issued by the Department of Conservation, New Zealand, refers to “Cabstand” as the name of a junction on Summit Road in Banks Peninsula.

Each description employs relative spatial relation terms such as “north of,” “south of,” or “southeast of” to describe the specimen’s location relative to “Cabstand,” thereby imposing qualitative constraints on its location. Taken together, these multiple references impose complementary spatial constraints that can be exploited to infer the location of the NGP by inverting (reversing) the spatial relations. The associated specimen coordinates act as spatial anchors, bounding the plausible region in which Cabstand must lie. By integrating reversed relative direction terms with the spatial distribution of specimen locations, it becomes possible to approximate the location of “Cabstand” even in the absence of an explicit gazetteer entry.

To operationalise this idea, we investigate three complementary georeferencing methodologies. The first approach is a deterministic method that draws on the formal search space model of Chen et al. [7] and constructs an Approximate Location Region (ALR) for an NGP by modelling the spatial relationship constraints expressed in locality descriptions relative to specimen coordinates. When the same place name appears across multiple descriptions, the intersection of their respective constraint regions progressively reduces the ALR, yielding increasingly precise candidate locations. Second, we adopt a probabilistic framework in which spatial relations are represented as likelihood functions conditioned on specimen coordinates and referenced place names. Multiple occurrences are aggregated to estimate the most likely location from the resulting spatial probability distribution, replacing crisp geometric boundaries with a continuous representation. Third, we explore an LLM-based approach by providing locality descriptions and specimen coordinates as inputs in the prompt and tasking the LLM with inferring NGP locations from combined textual and numeric cues. This allows us to assess the extent to which LLMs can integrate spatial language and coordinate information for georeferencing.

Among these three approaches, the probabilistic method yields the most precise results, although the other approaches also perform competitively. To our knowledge, this study represents the first systematic investigation of using biological specimen records as a primary data source for extracting and georeferencing NGPs. The methodologies and findings presented here demonstrate the feasibility and value of exploiting specimen-derived spatial information for enriching digital gazetteers.

The contributions of this work are as follows:

- (1) We demonstrate that biological specimen records contain non-gazetteered place names, highlighting these records as an underutilised source of geographic information over time.
- (2) We propose a data-driven strategy to georeference non-gazetteered place names using biological specimen records by exploiting multiple occurrences of the same place name across different specimen locality descriptions to derive spatial constraints on unknown locations.
- (3) We provide a systematic comparison of deterministic, probabilistic, and LLM-based approaches to text-based spatial modelling and inference of the locations of non-gazetteered place names, evaluating their strengths and limitations.
- (4) We construct a new benchmark dataset for evaluating methods for georeferencing place names derived from biological specimen records, which contain rich spatial knowledge.

The remainder of this paper is organised as follows. Section 2 reviews related work, Section 3 details the methodology. Section 4 presents the results, and Section 5 concludes the paper with a discussion of limitations and future research directions.

■ **Table 1** Example illustrating how multiple specimen locality descriptions refer to the same non-gazetteered place name (“Cabstand”), each accompanied by distinct coordinates and expressed through varying spatial relation terms (e.g., near, north of, south of, southeast of) that describe the NGP’s position relative to the specimen location.

Place name	Locality description	Latitude	Longitude
Cabstand	Banks Peninsula, Brocherries Road near Cabstand	-43.8030	173.0210
	Banks Peninsula, northeast of Akaroa, north of Cabstand, Summit Road	-43.7949	173.0210
	Banks Peninsula, south of Cabstand	-43.8093	173.0172
	Banks Peninsula, Long Bay Road southeast of Cabstand	-43.8039	173.0259

2 Related Work

This section reviews research relevant to the georeferencing of place names absent from gazetteers. It first outlines general approaches to georeferencing place names from textual descriptions, which provide the foundation for the research problem, and then examines studies addressing non-gazetteered place names, clarifying how this problem differs from generic georeferencing and highlighting the existing gap in the literature.

Automated georeferencing links textual descriptions or other location-referencing information to geographic coordinates. While some methods focus on georeferencing entire textual documents, such as Wikipedia articles, news reports, or social media posts [17, 13], others focus specifically on georeferencing the place names mentioned within text descriptions. The latter typically involves two main steps: identifying place names (toponym recognition) and determining their correct locations (toponym resolution or geocoding) [19, 27]. In many applications, both steps are closely tied to gazetteers: recognition benefits from name lists, and resolution selects among candidate gazetteer entries.

Early georeferencing approaches treated toponyms as gazetteer lookups with heuristic disambiguation, favouring populous or administratively prominent places [3, 4]. Unsupervised models improved this by selecting spatially consistent candidate sets using only gazetteer data [16]. However, a fundamental limitation of gazetteer-dependent methods is their incomplete coverage of vernacular, colloquial, and fine-grained place names [14, 1, 24]. As Wang et al. [27] observe, many geoparsing corpora restrict annotations to city-level and above, excluding fine-grained toponyms such as street names or the names of parks and monuments. These methods also struggle to resolve historical or outdated names that are no longer present in modern geographic databases [12]. Nevertheless, such names frequently persist in literary, archival, and other domain-specific textual sources with a historical dimension [3]. To overcome this, gazetteer-independent approaches such as TopoCluster learn spatial word distributions from large georeferenced corpora (GeoWiki) and resolve toponyms based on contextual geographic overlap [8].

Deep learning approaches replaced hand-crafted heuristics with learned contextual representations. For toponym recognition, neural sequence models (e.g., BiLSTM-CRF and CNN-RNN hybrids) capture informal spellings and noisy text, outperforming rule-based and gazetteer-only systems [2, 31, 20]. For toponym resolution, neural models either predict continuous coordinates from contextual embeddings, treating the earth as a regression surface or frame the task as tile classification or shared embedding learning, enabling similarity-based

grounding beyond strict gazetteer lookup [5, 11]. However, these methods rely on large-scale training corpora and still presuppose that targets resemble places represented in their training distributions.

LLMs are beginning to reshape both toponym recognition and resolution by injecting stronger, more explicit geographic knowledge into language representations. Geospatially grounded transformers such as GeoLM [18] and GeoReasoner [29] augment standard pretraining with coordinate-based spatial embeddings and contrastive alignment between textual and geospatial contexts, then fine-tune (or zero-shot apply) these models to sequence-tagging for toponym recognition and toponym linking against gazetteers, achieving state-of-the-art or near state-of-the-art performance on multiple benchmarks. For large-scale resolution, lightweight open-source LLMs (e.g., fine-tuned Mistral, Llama-2, Baichuan2) can be trained to infer an unambiguous type-level referent (city/state/country) from context and then delegate to geocoders such as GeoNames or Nominatim, substantially outperforming previous neural and voting-based resolvers at the global scale [15].

In our study, we target a stricter case of place names that are not found in gazetteers at all, not even as alternate spellings, variants, or official forms. In this setting, resolution cannot fall back on candidate resolution from GeoNames, OpenStreetMap, or similar resources and must infer location from other evidence. To date, relatively few studies have directly targeted the resolution of such truly non-gazetteered place names. A closely related line of work is the Bhugol framework proposed by Sharma et al. [23], which addresses non-gazetteered place names by exploiting external regularities rather than relying on an explicit gazetteer. Bhugol estimates coordinates by clustering either gazetteer place names sharing frequent morphological affixes (via HDBSCAN) or gazetteer place names that co-occur with the target name in a news corpus (via DBSCAN), and assigning the centroid of the most representative spatial cluster as the inferred location; an integrated variant combines both signals. Among existing approaches, this methodology is the closest to our work in its explicit focus on non-gazetteered place names. However, it is strongly region-dependent and relies on recurring naming conventions and extensive corpus evidence, which limits its portability across geographies and languages. Moreover, even its best-performing variant, Bhugol-LS, reports mean distance errors often exceeding 100 km, which may be acceptable for coarse-resolution tasks but falls short in applications requiring fine-grained spatial accuracy.

Chen et al. [7] propose a graph-based approach to georeferencing place names absent from gazetteers by exploiting spatial relationships extracted from natural language, a method that has informed our work. Their framework uses place graphs from collective manually created descriptions, with nodes representing locations and edges encoding qualitative spatial relation terms, enabling inference of a target location by linking a non-gazetteered place name (locatum) to surrounding gazetteered landmarks (relata). However, their approach assumes that the relatum can be reliably identified and disambiguated, and that the locatum corresponds to a formal or alternative gazetteer entry. In contrast, the locality descriptions in our data refer to locations that lack any representation, formal or variant in existing gazetteers or VGI resources. Moreover, our records provide no external contextual cues regarding the prominence or granularity of either locatum or relatum, making it infeasible to construct the probabilistic search space required by their pipeline. Relatedly Yousaf and Wolter [30] propose a reasoning-based model that interprets natural language place descriptions by representing named and unnamed spatial entities as relational constraints and resolving them through spatio-ontological inference. While their approach aligns with our emphasis on spatial relation terms, it ultimately grounds interpretations by matching inferred entities to OpenStreetMap features, whereas our work targets place names that lack any gazetteer representation.

Taken together, the literature suggests a gap in both (i) using specimen records as a source for identifying place names that are absent from gazetteers and (ii) developing methods to georeference place names that are neither gazetteered nor supported by rich corpus-derived evidence. This motivates our effort to extract and georeference non-gazetteered place names using biological specimen records as the source.

3 Methodology

3.1 Locality description corpus

For our experiments, we use biological specimen records from the Allan Herbarium (CHR)³ dataset of Manaaki Whenua – Landcare Research (MWLR), New Zealand. We retain only records that include both specimen coordinates and a free-text locality description. As described earlier, specimen datasets often contain repeated locality descriptions as multiple specimens may be collected at the same site. Because our objective is to use distinct locality descriptions to constrain an NGP’s position, we removed duplicate locality descriptions from the original MWLR extract. We also noted the presence of records that share identical coordinate pairs but differ slightly in wording (e.g., “*Campbell Island, near **footpath towards Met Service hostel***” vs. “*Campbell Island, near **concrete pad on Met Service base***”). To avoid redundant spatial evidence, we removed records with duplicate coordinate pairs, keeping only the first. While this filtering reduces redundancy, it may also discard useful linguistic variation, as records with the same coordinates can contain complementary contextual clues.

Next, we extracted place names from locality descriptions using the *en_core_web_trf* Named Entity Recognition (NER) model in spaCy⁴, fine-tuned on a subset of 50 manually annotated locality descriptions from our data. We then retained only records in which at least one extracted place name participates in an explicit spatial relationship (e.g., “18 km west of Ashburton”). Descriptions that merely list place names without explicitly expressing a spatial relation (e.g., “Green Hill, Upper Poulter Valley, Waimakariri Region”) were excluded, although in some cases such listings may implicitly encode hierarchical relationships.

The initial list of relation terms was developed from common spatial expressions observed during manual inspection of locality descriptions and was subsequently refined iteratively by checking unmatched and incorrectly matched examples. The matching patterns followed a general structure of the form `<spatial relation term> <place name>`. When a locality description contained multiple such expressions, all matching relations were extracted. A detailed discussion of the spatial relation terms identified in the dataset is provided in Section 3.3.1.

After processing, the dataset comprised 13,746 locality descriptions, grouped by extracted place names, together with their associated spatial relations and the coordinates of each locality description.

3.2 Identifying non-gazetteered place names

To demonstrate the presence of NGPs in biological specimen records, we used a two-stage identification procedure: (i) automated screening to construct a candidate set, and (ii) targeted manual verification to assess its reliability. This procedure was applied to the place names extracted from the above-mentioned dataset of 13,746 locality descriptions. Each

³ <https://www.landcareresearch.co.nz/tools-and-resources/collections/allan-herbarium>

⁴ <https://spacy.io/>

extracted place name was checked against the New Zealand Geographic Board (NZGB)⁵ Gazetteer, GeoNames⁶, and OpenStreetMap (OSM)⁷ to determine whether it is already gazetteered.

In the automated stage, we applied sequential filtering using these three gazetteers. First, each candidate was compared against the downloaded NZGB Gazetteer export and matching names were removed. Second, the remaining candidates were cross-checked against the GeoNames export for New Zealand to identify additional gazetteered names not present in the NZGB dataset. Third, unresolved candidates were queried against the OSM Nominatim API, and any strings that could be matched to an OSM feature were excluded. After these steps, 520 candidates were not found in any of the three gazetteers and were retained as candidate NGPs. To reduce sensitivity to minor orthographic variation in the free-text locality field (e.g., misspellings, spacing differences, or abbreviations), we employed fuzzy string matching rather than exact matching, using a similarity threshold of 80% with the RapidFuzz library⁸.

Despite this multi-source filtering, the resulting set still required manual verification because specimen locality descriptions are authored as free text and are not standardised. To quantify the extent of these issues, we manually inspected a random sample of 50 of the 520 automatically identified candidates. Of these, 30 were confirmed as true NGPs.

The observed false positives largely fell into five categories:

- **Alternative or informal naming** (a real place, but referenced using a non-standard variant): e.g., Bowscale Lake for Bowscale Tarn; Hutt Road for Hutt Motorway.
- **Generic feature terms misidentified as place names by spaCy** (common nouns treated as named entities): e.g., Airport in “Kaitaia, Quarry Road, near Airport”; BNZ bank in “Waikanae, marae grounds behind BNZ bank”.
- **Misspellings** (orthographic noise that persisted despite fuzzy matching): e.g., Mt Rockfort for Mt Rochfort; Buffalo beach for Buffalo Beach.
- **Abbreviations** (shortened forms not captured by gazetteer matching): e.g., big r. for Big River; Cobb V. for Cobb Valley.
- **Compound or delimited forms** (multi-name strings that complicate recognition and matching): e.g., “along the Slaughter/Long Burn”; “near Belltown/Manunui hut”.

These error modes highlight that identifying truly non-gazetteered names is not simply a matter of gazetteer lookup. Developing a comprehensive recognition and normalisation pipeline is an important problem in its own right; however, our aim here is to establish that specimen records themselves are a credible source of non-gazetteered names, and the findings in our analysis support this.

Based on the manual audit, we estimated that approximately 60% of the automatically extracted candidates were true NGPs. Applying this precision estimate to the 520 candidates yields an expected ~ 312 true NGP mentions, corresponding to roughly 2% of the 13,746 locality descriptions in our sample (i.e., $520 \times 0.60 \approx 312$; $312/13,746 \approx 0.023$). Although this proportion appears small, it is operationally significant at the collection scale: biodiversity infrastructures aggregate millions of specimen records, and even a 2% rate translates into tens of thousands of locality descriptions containing place references that cannot be resolved

⁵ <https://gazetteer.linz.govt.nz>

⁶ <https://www.geonames.org>

⁷ <https://www.openstreetmap.org>

⁸ <https://pypi.org/project/RapidFuzz>

through conventional gazetteer-based workflows. Notably, this proportion is observed in New Zealand, a relatively well-mapped country with a mature national gazetteer and rich VGI sources, suggesting that the prevalence of NGPs is likely to be higher in regions with less complete mapping coverage.

Table 2 presents a representative subset of NGPs from our dataset, categorised according to the most commonly observed (non-exhaustive) NGP types. Many of these NGPs correspond to historical place names that are absent from current gazetteers, yet retain substantial historical value.

■ **Table 2** Examples of manually verified NGPs identified in the dataset, categorised by the most commonly observed (non-exhaustive) types, together with the locality descriptions in which they appear and the corresponding record dates.

Type	Place name	Example occurring locality description	Recorded date
Farm stations	Beltana station	North Canterbury, hills east of Parnassus, above Beltana Station, near radio station	6/01/1969
	Ngatapa Sunworth Station	Gisborne, northwest of Ngatapa Sunworth Station.	24/02/1990
Trig stations	Oropuke trig	Chatham Island, nearest major locality Oropuke trig, Southern Tableland, east of Oropuke trig, at top end of “The Slump”, near hut	26/02/1985
	Driscoll Trig	Wairarapa, East Wairarapa Ecological District, Te Wharau near Driscoll Trig	7/08/1992
Infrastructure facilities	Huia Filter Station	Titirangi, Woodlands Road, at Huia Filter Station	16/03/1995
	P. & T. Microwave Station	Vernon Hills, Marlborough, near P. & T. microwave station	24/11/1970
Community facilities	Bryant Home	South shore of Raglan Harbour, near Bryant Home	23/10/1967
	Manaroa School	Forest remnant beside Manaroa school near shore, Pelorus Sound, Marlborough	12/11/1971
Recreation places	Christian Youth Camp	Whakamaru, near Christian Youth Camp	25/02/1983
	Catleys (camp site)	Kauaeranga Valley, Thames opposite Catleys	26/11/1973
Hospitality venues	Cave Rock Tearooms	Sumner near Christchurch, Canterbury, near Cave Rock Tearooms	28/10/1971
	Delta Lodge	Near Delta Lodge, Orongorongo Vly.	25/06/1974

3.3 Georeferencing non-gazetteered place names

We present three methods for georeferencing NGPs using biological specimen records: deterministic, probabilistic, and LLM-based. Each method has distinct strengths and weaknesses. The deterministic and probabilistic methods use structured relation inputs derived from locality descriptions, whereas the LLM-based method operates directly on the unprocessed locality descriptions and associated specimen coordinates. However, as noted in the Introduction, all three are based on the same underlying strategy of exploiting repeated occurrences of the same NGP across multiple locality descriptions as sources of spatial evidence to infer its location.

For example, the locality description “*Main Divide, Coromandel Ra, 1 mile SE of Maumaupaki Trig*” indicates that a specimen was collected one mile southeast of Maumaupaki Trig, a trig station in the Maumaupaki area of New Zealand that is absent from current gazetteers. Because the specimen record includes geographic coordinates (latitude: -36.97612 , longitude: 175.5862), this relation can be inverted to infer that the NGP lies one mile northwest of the specimen location.

When multiple descriptions refer to the same NGP using different spatial relations, as illustrated in Table 1, the intersection of these constraints enables a more precise estimate of its location. The following subsections describe how this strategy is operationalised in our study.

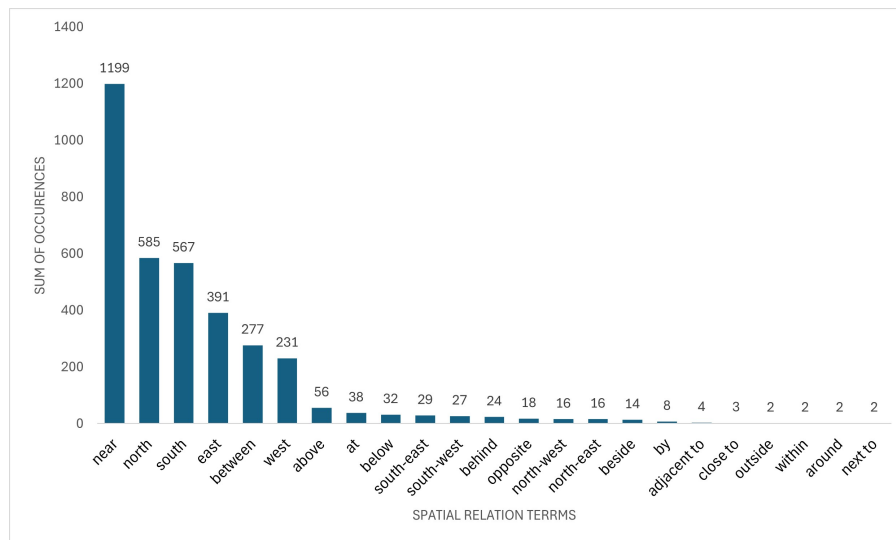
3.3.1 Interpreting spatial relations

Spatial relations in specimen locality descriptions are represented as triplets comprising a relatum (the referenced place), a spatial relation, and a locatum (the target location). For example, in the locality description “*18 km west of Ashburton, on Maronan Valetta Road*”, *Ashburton* is the relatum, *18 km west of* specifies the spatial relation, and the locatum is the specimen’s collection location, which is implicit in the text and represented by the associated coordinates found in the specimen record.

Assuming that the place name to be georeferenced is *Ashburton*, the spatial relation can be inverted, treating the known specimen location as the relatum and the place name as the locatum. Locality descriptions may include additional place names; similar to the example above, “*on Maronan Valetta Road*” indicates that the specimen was collected on Maronan Valetta Road and does not express a direct spatial relationship between Maronan Valetta Road and Ashburton. Accordingly, our methodology considers only direct spatial relationships between the target place name and the specimen location.

We also acknowledge an asymmetry between relatum and locatum, as described in Talmy’s work [25], which highlights differences between the reference object and the object whose location is being described. Reference objects are typically larger, more permanently located, more geometrically complex, and more independent than the objects they help locate. For example, “*the post box is by the post office*” is a more natural expression than “*the post office is by the post box*” because the post office is larger and more permanent, making it a more plausible reference object. However, this asymmetry does not pose a problem for our georeferencing methods, as in our data the locatum is a fixed point observation (the specimen location), and multiple such observations are aggregated to derive spatial constraints.

In our dataset, we observed a wide range of spatial relation terms (Figure 1). The distribution is highly skewed: the proximity relation *near* occurs substantially more frequently than any other term. Cardinal direction relations (*north*, *south*, *east*, and *west*) also appear frequently, whereas the remaining terms occur far less often. In addition, cardinal and



■ **Figure 1** Frequency distribution of spatial relation terms extracted from specimen locality descriptions.

inter-cardinal relations (e.g., south-west, north-east) are sometimes accompanied by an explicit distance, e.g., “*2 km north of*”, which strengthens the spatial constraint beyond direction alone.

Examining the relation terms in our dataset, we attempted to categorise them. However, a key challenge in adopting categorisations from prior studies is that our use case involves inverting the locatum and relatum implied by the locality description; consequently, any categorisation must remain meaningfully interpretable under inversion. While the inverses of cardinal and inter-cardinal relation terms are straightforward, many other relations appear to be symmetrical, such that their interpretation remains unchanged or similar when inverted.

For these non-directional expressions, we therefore treat them as proximity-like constraints even though the perceived degree of proximity could vary. Terms such as *near*, *by*, *close to*, *next to*, *adjacent to*, and *beside* are clearly proximity-related and effectively symmetric (e.g., if *X* is near *Y*, then *Y* is near *X*). Containment-oriented expressions such as *at*, *within*, and *outside* instead constrain the specimen location relative to a region associated with the referenced feature. However, when the relation is inverted, they do not specify a unique direction or operator for locating the target place name; operationally, they still primarily imply geographic closeness to the relatum. Because specimen records represent the specimen location as a point (latitude/longitude), statements such as “the specimen is within/outside *X*” (*X* is a place name) are best interpreted as indicating proximity to *X*, with the direction left unspecified. Orientation-based terms such as *above*, *below*, *behind*, and *opposite* often depend on a local frame of reference, and thus are not consistently invertible from text alone. Across these cases, modelling non-cardinal relations as proximity-like constraints appears to preserve their role as locality restrictions while avoiding misinterpretation in the inverted setting.

3.3.2 Evaluation dataset design and representation

Pseudo-NGP dataset for evaluation

To evaluate our methods for georeferencing, we constructed a subset from the preprocessed dataset of 13,746 locality descriptions mentioned in Section 3.1. To enable evaluation, we required place names with known coordinates as ground-truth data. Therefore, we selected locality descriptions that contain spatial relations referring to place names recorded in the NZGB Gazetteer. Place names were disambiguated and resolved against the NZGB Gazetteer using the *state/province* information provided in the corresponding biological specimen records. As this study serves as a proof of concept, only place names represented as point geometries were selected in order to simplify the evaluation setting. These gazetteered place names were treated as pseudo-NGPs by withholding their coordinates during inference, allowing us to quantitatively assess the accuracy of the proposed methods. This dataset thus provides a benchmark for systematic evaluation.

Among the resolved place names, we further selected those referenced in at least three locality descriptions, as three constraints constitute the minimum required to define a bounded area. Under this criterion, the final evaluation dataset consists of 2,403 locality descriptions referring to 365 place names. Each locality description is associated with specimen coordinates, and each place name has corresponding resolved ground-truth coordinates.

Input data representation

Both proposed deterministic and probabilistic methods operate over a structured JSON representation derived from the dataset. For each target place name, we extract spatial relation mentions from locality descriptions and group them into a single place-centric entry. Each entry follows the schema *place_name* $\rightarrow$ *place_name_coordinates* $\rightarrow$ *relations*:

- **place_name**: identifier for the target place.
- **place_name_coordinates**: ground-truth latitude/longitude of the *place_name* (retained only for evaluation and not used to form constraints).
- **relations**: a list of spatial relations associated with the *place_name*, extracted from specimen locality descriptions as described in Section 3.1. Each relation entity includes the extracted spatial relation term, an associated distance value when provided, and the coordinates of the specimen record from which the relation was extracted. Relation terms are normalised to a canonical form to reduce surface variation (e.g., collapsing whitespace, standardising hyphenation for multiword expressions such as *next to* $\rightarrow$ *next-to*, and mapping direction variants such as *NE*, *north east* to a single label). Cardinal and inter-cardinal relations are represented in their inverted form, consistent with our interpretation that the specimen coordinate is the relatum and the target place name is the locatum.

We have used the place name *Kauangaroa*⁹ as the running example to illustrate this process. Table 3 lists three locality descriptions that reference *Kauangaroa*, along with the specimen coordinates associated with each description and the extracted (and, where applicable, inverted) spatial relation terms. Listing 1 shows the corresponding JSON entry produced from the data in Table 3. We generated one such JSON entry per place name and combined them into a single JSON file covering all 365 places in our dataset as the input.

⁹ *Kauangaroa* is a locality in Wellington, New Zealand.

■ **Listing 1** Example JSON entry used to generate the ALR for the target place name *Kauangaroa*.

```
{
  "place_data": [{
    "place_name": "kauangaroa",
    "place_name_coordinates": {
      "latitude": "-39.928234", "longitude": "175.279616"
    },
    "relations": [
      {
        "relation": "near",
        "distance": "",
        "latitude": "-39.93045", "longitude": "175.2665"
      },
      {
        "relation": "west",
        "distance": "1km",
        "latitude": "-39.92546", "longitude": "175.2909"
      },
      {
        "relation": "east",
        "distance": "",
        "latitude": "-39.91161", "longitude": "175.2635"
      }
    ]
  }]
}
```

Building on this representation, we now describe the proposed methods for georeferencing NGPs.

3.3.3 Deterministic georeferencing using spatial relations

In this approach, we framed the georeferencing of NGPs as an inference problem over a *place graph*. The *place graph* provided a compact representation that connects each target place name to all specimen locations whose locality descriptions mention that place name. We then aggregated the resulting spatial constraints to derive an *Approximate Location Region* (ALR) for each target place *p*. The key steps of this method are as follows:

- **Relatum point set:** For each relation entity in the JSON file generated, we read the specimen coordinates (latitude/longitude) and represented them as relatum points. These points served as the geometric anchors from which constraint regions were generated.

■ **Table 3** Locality descriptions from specimen records that refer to *Kauangaroa* using relative spatial expressions. The latitude and longitude values correspond to the coordinates of the specimen locations described in each locality description.

Place Name	Locality Description	Latitude	Longitude	Extracted Spatial Relation	Inversed Spatial Relation
Kauangaroa	Near Fordell at Kauangaroa	-39.93045	175.2665	near	near (R1)
	1 km east of Kauangaroa, Whangaehu River	-39.92546	175.2909	1 km east	1 km west (R2)
	Whangaehu River catchment west of Kauangaroa	-39.91161	175.2635	west	east (R3)

- **Distance parsing:** When a distance value was present, it was parsed and converted to kilometres (from meters or miles). If no distance was provided, it was treated as unspecified and handled according to the relation-specific defaults described below.
- **Spatial context region:** For each target place p , we derived a bounded spatial context to constrain subsequent geometric operations. This context was defined as the bounding box of the relatum point set, expanded by a buffer determined from the median nearest-neighbour distance among the relatum points. This statistic also provided an estimate of the local spread of observations associated with p .
- **Constraint region construction:** Each relation entity was converted into a feasible region within the spatial context:
 - For cardinal and inter-cardinal relations, we generated a directional region by intersecting the spatial context with the corresponding half-plane(s). If a distance was given, the region was additionally intersected with a distance buffer around the relatum point.
 - For all remaining relations, we generated a proximity-like buffer around the relatum point. When an explicit distance was not provided, we set the buffer radius adaptively based on how tightly the specimen (relatum) points for that place name cluster, using their median nearest-neighbour spacing as a proxy for local spread. The resulting default radius was then clipped to global bounds of 0.2 km and 30 km. The lower bound of 0.2 km prevents unrealistically small buffers that would frequently produce empty intersections given coordinate uncertainty and vague locality descriptions, while the upper bound of 30 km prevents weak proximity relations from dominating the spatial context with overly broad regions.
- **ALR computation:** The ALR for p is computed as the intersection of all retained constraint regions.

Figure 2a illustrates the spatial constraint regions generated for *Kauangaroa* from the JSON schema in Listing 1. The figure shows the feasible region associated with each relation (ordered as in the schema) and the resulting ALR obtained as the intersection of these constraint regions. In this formulation, the ALR provides a feasible region rather than a uniquely determined coordinate; accordingly, for evaluation, we used the centroid of the ALR as a point estimate and interpreted the ALR as an uncertainty region. However, an ALR was not obtained for all places: in some cases, the strict intersection became empty due to incompatible or overly restrictive constraints across locality descriptions, particularly when proximity-like relations were imprecise or when extracted relations contain noise. This limitation motivated our second approach, which replaced hard region intersection with a probabilistic formulation to produce a model-derived point estimate.

3.3.4 Probabilistic georeferencing using spatial relations

In this method, we replaced region intersection with a likelihood-based formulation that evaluated candidate coordinates against all extracted spatial relations for a given place name. Each relation contributed a likelihood (soft preference) around its specimen anchor, and the location that maximised the aggregated likelihood was selected as the final coordinate estimate. The key steps of this method are as follows:

- **Relatum point set:** As before, each relation entity provided a specimen coordinate that was treated as the relatum point anchoring the constraint. Distances, when present, were parsed and converted to kilometres.
- **Candidate search space:** For each place name p , we defined a bounded candidate region based on the specimen locations that reference p .

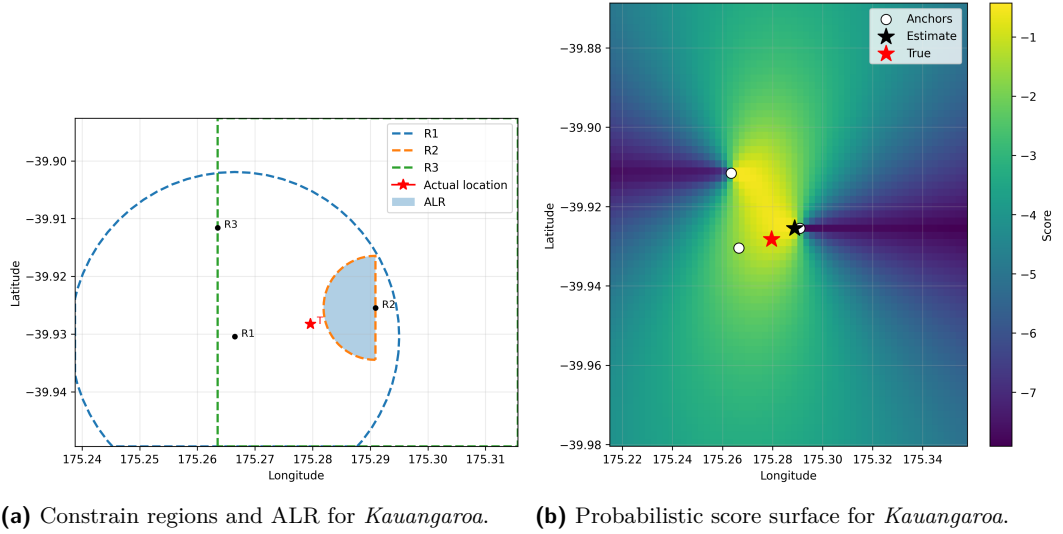


Figure 2 Deterministic (a) and probabilistic (b) georeferencing results for *Kauangaroa*. The prediction errors between the actual and estimated locations are 0.71 km and 0.83 km, respectively.

- We computed a robust spatial centre as the median latitude and median longitude of all referencing specimen coordinates.
- We defined a fixed-radius spatial bound (60 km) around this centre to limit the candidate region and reduce unlikely solutions far from the evidence. This radius is approximately four times the estimated proximity uncertainty parameter (s_{near}) described below, beyond which likelihood contributions become negligible. This fixed-radius search assumes that the specimen anchors for a target place form a spatially coherent set. It may be less appropriate for records containing strong outliers or multiple spatial clusters, where an adaptive or cluster-aware search region would be preferable.
- **Likelihood construction:** For each relation, we define a log-likelihood $\ell(x)$ at candidate location x relative to anchor a , where $d(x, a)$ denotes the great-circle (haversine) distance and $\Delta\theta(x, a)$ the angular deviation from the expected bearing; d_0 is the stated distance when provided. Proximity is modelled via distance decay, while directional and distance deviations are modelled using Gaussian (normal) distributions.
 - *Proximity:* $\ell_{\text{prox}}(x) = -d(x, a)/s_{\text{near}}$.
 - *Directional:* $\ell_{\text{dir}}(x) = -\Delta\theta(x, a)^2/(2\sigma_\theta^2)$.
 - *Directional + distance:* $\ell_{\text{dir+dist}}(x) = -\Delta\theta(x, a)^2/(2\sigma_\theta^2) - (d(x, a) - d_0)^2/(2\sigma_d^2)$.
 Here s_{near} (km) controls proximity decay, σ_θ (degrees) directional uncertainty, and σ_d (km) distance uncertainty.
- **Uncertainty parameter estimation:** As described above, the log-likelihood construction depended on three uncertainty parameters: s_{near} , σ_θ , and σ_d . To estimate these values, we used an auxiliary subset of our preprocessed dataset consisting of place names that appeared in only one or two locality descriptions. These place names are listed in the NZGB Gazetteer but were excluded from the evaluation dataset, which requires at least three descriptions per place name. The auxiliary subset contained 1,485 locality descriptions referring to 1,204 resolved place names. Using the ground-truth coordinates of these place names, we computed: (i) anchor-to-ground-truth distances for proximity-like relations, (ii) angular deviations between observed bearings and expected directions

for directional relations, and (iii) differences between observed and stated distances for distance-qualified relations. Because the distance-related distributions included extreme values, we used the 90th percentile of the distance and absolute distance-difference distributions to determine s_{near} and σ_d , thereby representing the majority of cases. For directional relations, we used the median angular deviation to determine σ_θ . The resulting parameter values were $s_{\text{near}} = 14$ km, $\sigma_\theta = 47^\circ$, and $\sigma_d = 4$ km.

- **Combining evidence:** For a given candidate location, log-likelihoods were summed across all relations to obtain a single score.
- **Coarse-to-fine MAP search:** The bounded region was discretised into an initial coarse grid with 3 km spacing, enabling an efficient first-pass search while preserving sufficient spatial coverage. Each grid point was evaluated by summing relation log-likelihoods, and the best-scoring point was retained. We then iteratively refined the estimate by defining a smaller square window centred on the current best point. The window extent was determined relative to the previous grid spacing, covering a local neighbourhood of candidate cells defined as a square region spanning a fixed multiple of the current grid resolution in each direction. Within this window, progressively finer grid spacings (1.2 km, 0.5 km, and 0.2 km) were applied. At each level, candidates were re-scored, and the maximum was retained. The final latitude/longitude estimate was the highest-scoring point at the finest resolution.

Overall, the method produces a single coordinate for each target place name by selecting the candidate that maximises the aggregated likelihood across all extracted relations. Figure 2b visualises the resulting probabilistic surface for *Kauangaroa*. Warmer colours indicate candidate locations that are more consistent with the extracted spatial relations when evaluated jointly against all specimen anchor points (white markers), and the peak of this surface (black star) corresponds to the model’s preferred coordinate estimate.

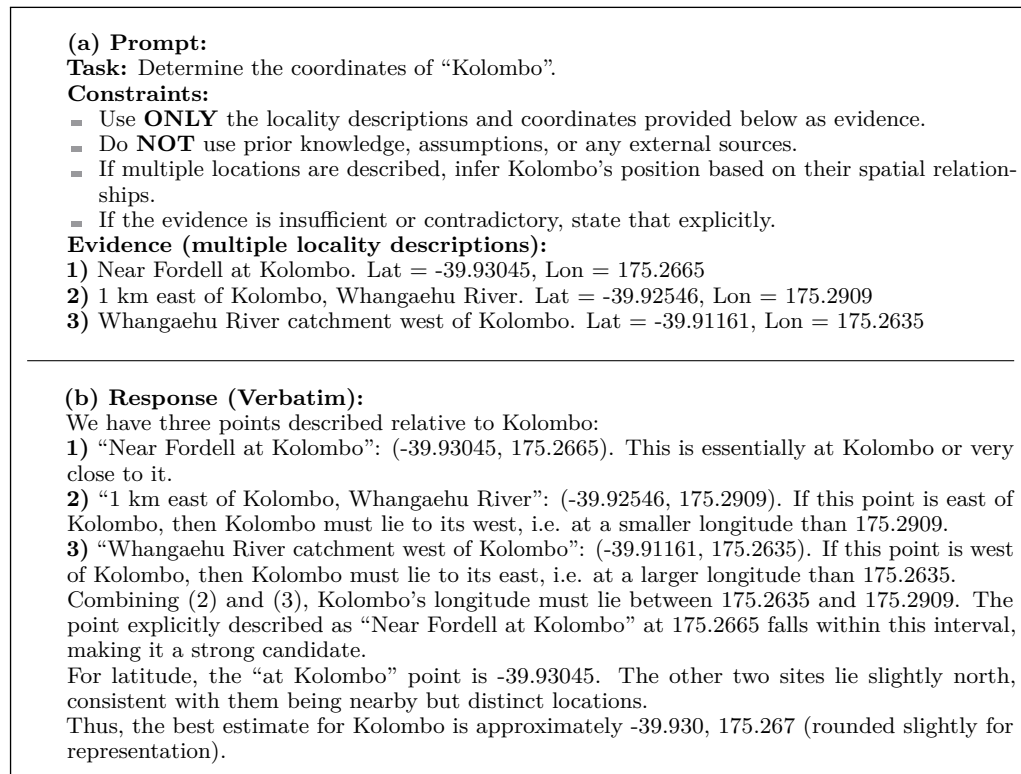
3.3.5 LLM-based georeferencing

As the third approach, we explored how an LLM performs on this spatial modelling style georeferencing task, using the `gpt-5.1-2025-11-13` model via the OpenAI Batch API¹⁰ with a simple zero-shot prompt. Figure 3a presents the inference prompt; only the locality descriptions and coordinates differ across target places. No system prompt was used, and we did not modify the default model settings.

Because our pseudo-NGP dataset consists of real gazetteered toponyms with coordinates withheld for evaluation, we replaced each target name in the prompt with a placeholder (*Kolombo*) to prevent the model from relying on prior knowledge of the name. The model was explicitly instructed to use only the provided locality descriptions and coordinates as evidence, allowing us to evaluate the ability of a general-purpose LLM to perform spatial modelling from textual spatial information. In a true-NGP setting, however, an LLM may still draw on latent knowledge from historical documents, web pages, maps, or other textual sources seen during training. Such knowledge could improve georeferencing when accurate, but it may also introduce unsupported assumptions. Thus, masking provides a stricter evaluation of inference from specimen-based evidence alone, while true-NGP georeferencing would require further evaluation of how LLMs combine provided evidence with their inherent geographic knowledge.

¹⁰<https://developers.openai.com/api/docs/guides/batch/>

The LLM was prompted to return a single coordinate estimate for each target place name using the provided locality descriptions and specimen coordinates. To improve transparency, the model was also asked to return its reasoning alongside the predicted coordinates. We found that including this reasoning step also improved the overall accuracy of the LLM-based method, as reported in Table 4. Figure 3b presents the verbatim response from the LLM, including the reasoning behind the inferred coordinates.



■ **Figure 3** (a) Prompt used to georeference *Kauangaroa* with GPT-5.1. (b) Verbatim response showing the detailed reasoning. The prediction error is 0.8 km.

4 Results and Discussion

To evaluate performance, we report the mean and median error distance (km) between predicted and true coordinates. We also compute A@161 and A@1, measuring the percentage of place names correctly georeferenced within 161 km (100 miles) and 1 km, respectively. A@161 is widely used in toponym resolution and geocoding [8, 12, 23] to capture coarse regional accuracy, while A@1 provides a stricter measure of spatial accuracy.

Table 4 compares the georeferencing performance of the deterministic, probabilistic, and LLM-based methods across these metrics for predicting coordinate estimates for all 365 pseudo-NGP place names in the evaluation dataset. The deterministic model shows the weakest overall performance. It fails to produce estimates for 108 place names, reflecting its reliance on strict rule satisfaction and its inability to handle ambiguous or conflicting spatial constraints. Even when estimates are produced, it records the highest mean error (9.19 km) and the lowest A@1 and A@161 scores.

Both the probabilistic and LLM-based methods successfully generate location estimates for all place names and achieve identical A@161 values, with only two shared outliers exceeding the 161 km threshold. Between these two approaches, the probabilistic model performs best overall, achieving the lowest median error (1.43 km) and the highest A@1 score, indicating more precise fine-grained georeferencing. The two GPT-5.1 prompt variants used for the LLM-based method show broadly comparable performance. Overall, the reasoning prompt performs slightly better in terms of distance-based error, reducing both the mean and the median errors, although the direct prompt achieves a higher A@1 score. The LLM-based variants achieve the lowest mean errors overall, suggesting fewer extreme failures; however, their higher median errors and lower A@1 scores indicate reduced accuracy at finer spatial scales compared to the probabilistic approach.

We examined the two place names for which both the probabilistic and LLM-based methods produced errors exceeding 161 km: Fog Peak and Rockdale. In both cases, the specimen coordinates lie far from the corresponding NZGB gazetted locations. Fog Peak is an off-track mountain peak, which may contribute to ambiguity in locality descriptions, while the gazetted location of Rockdale is distant from the specimen coordinates, despite descriptions stating the specimens are “near Rockdale”. These discrepancies likely reflect issues in the original data rather than failures of the georeferencing methods.

■ **Table 4** Comparison of the georeferencing performance of the three methods. Mean and median denote error distance in kilometres. ↓ indicates lower is better; ↑ indicates higher is better.

Model	Mean error↓	Median error↓	A@1↑	A@161↑	No estimate↓
Deterministic model	9.19	1.72	20.3%	69.6%	108
Probabilistic model	8.95	1.43	36.0%	99.5%	0
GPT-5.1	6.23	1.85	30.9%	99.5%	0
GPT-5.1 with reasoning	5.89	1.80	28.1%	99.5%	0

4.1 Performance based on the spatial relations

To further investigate the results of the three models, we analysed model performance with respect to the spatial relations present in locality descriptions associated with each place name. Spatial relation terms were categorised into three groups: direction, distance with direction, and proximity, which provided additional insight into the types of relations each model handled more effectively. Table 5 summarises model performance across different combinations of spatial relations. In this table, the LLM-based method refers to the GPT-5.1 reasoning-prompt setting.

The results indicate that although the probabilistic model performs well overall, it is less effective when only directional relations are available, underperforming both the deterministic and LLM-based approaches in this setting. In contrast, when only proximity relations are present, the probabilistic model performs comparatively well. When locality descriptions include all three types of spatial constraints, direction, distance, and proximity, both the probabilistic and LLM-based methods achieve strong performance, with the probabilistic model yielding the highest overall accuracy. Notably, the LLM-based method exhibits relatively consistent performance across different types of relational constraints.

These differences can be partly explained by how each method interprets and integrates different types of spatial relations. Direction-only relations constrain bearing but not distance, leaving the location of the likelihood peak along the directional axis underdetermined. As

■ **Table 5** Comparison of georeferencing performance of the three methods by spatial relation type.

Spatial relations found	Model	Mean error (km)↓	Median error (km)↓
Direction only	Deterministic model	10.02	2.30
Direction only	Probabilistic model	34.58	5.25
Direction only	LLM-based model	6.46	1.41
Proximity only	Deterministic model	6.87	1.56
Proximity only	Probabilistic model	7.33	1.09
Proximity only	LLM-based model	7.48	1.32
Direction + Distance	Deterministic model	4.02	2.39
Direction + Distance	Probabilistic model	7.60	0.74
Direction + Distance	LLM-based model	2.59	2.02
Direction + Proximity	Deterministic model	35.01	2.25
Direction + Proximity	Probabilistic model	8.85	1.58
Direction + Proximity	LLM-based model	7.98	1.94
Direction + Distance + Proximity	Deterministic model	33.71	1.57
Direction + Distance + Proximity	Probabilistic model	2.73	1.47
Direction + Distance + Proximity	LLM-based model	2.81	1.82

a result, the probability surface remains diffuse, limiting spatial precision in the absence of additional constraints. Proximity relations, by contrast, align well with probabilistic representations of spatial uncertainty, allowing nearby candidate locations to be weighted without precise orientation. Explicit distance constraints act as strong anchors for all methods, particularly when combined with directional information.

4.2 Key observations on LLM reasoning

We examined the reasoning returned by GPT-5.1 to understand the behaviours that enabled it to georeference effectively across the dataset. Focusing on cases where the LLM outperformed the probabilistic and deterministic methods, two patterns stand out.

First, as shown in Figure 3b, GPT-5.1 is effective at extracting spatial information directly from free text and synthesising it into a coherent interpretation without manual preprocessing. It correctly identifies spatial relations in the text and incorporates associated distance information without relying on an explicit parser. Notably, it independently adopted our approach: inverting the spatial relation so that it is interpreted from the specimen location toward the referenced place name. We also observe that GPT-5.1 accurately performs geographic calculations by combining distance and direction cues, for example, converting stated offsets into rough $\Delta\text{lat}/\Delta\text{lon}$ adjustments and performing basic unit conversions (e.g., miles to kilometres).

Second, the model does not treat each reference equally. Its inference strategy is to find a best-fit solution while ignoring noise, instead of trying to satisfy every relational mention. Rather than weighting all relations equally, the LLM emphasises a subset of mutually consistent (or otherwise more reliable) constraints when deriving an estimate. This behaviour can be observed in the reasoning shown in Figure 3b, in which the LLM primarily uses the “at” coordinates to infer the final coordinates, while the other mentions are used mainly to validate plausibility. This substantially benefits the LLM approach, contributing to the lowest mean error.

4.3 Summary of results

Table 6 provides a summary of the strengths and limitations of each model.

Table 6 Qualitative comparison of georeferencing approaches based on observed behaviour in the experiments.

Method	Strengths	Limitations
Deterministic	Simple and interpretable; Explicitly models an uncertainty region using ALR; Effective when inputs are precise; Exact constraint satisfaction ensures reproducibility.	Requires structured inputs and explicit relation modelling; Sensitive to minor inconsistencies, resulting in the lowest performance.
Probabilistic	Highest overall precision, with the best median error and Acc@1; Performs well when proximity constraints (the most common) are present; Generates estimates for all cases; Reproducible under fixed parameters.	Requires structured inputs and explicit relation modelling; Performs less well with direction-only relations; Affected by inconsistent or imprecise relations.
LLM-based	Operates directly on unprocessed locality descriptions; Produces acceptable estimates for most cases (lowest mean error); Flexible in handling a wide variety of spatial relations (not limited to a pre-defined set).	Reproducibility is not guaranteed; Less precise than the probabilistic model; Costly.

5 Conclusion

This study shows that biological specimen records are a practical source for both identifying and georeferencing non-gazetteered place names within a region. We establish the presence of NGPs in specimen collection records and provide an empirical estimate of their prevalence in the collection. We also compare traditional deterministic and probabilistic georeferencing approaches with a modern LLM-based approach to assess whether LLMs can reliably support a largely spatial modelling-based task. The results indicate that, while the LLM-based method yields broadly acceptable estimates, traditional spatial modelling, particularly probabilistic inference, remains more precise when fine-grained localisation is required.

Limitations. The proposed methods are currently implemented only for NGPs that can be represented as points, limiting applicability to linear or areal features. Performance depends on the availability, spatial distribution, and accuracy of georeferenced specimen anchors, as well as on the diversity of spatial relations available for each place name. The performance of the methods may therefore be biased toward better-constrained place names. In addition, the deterministic and probabilistic methods rely on predefined spatial relation terms and predefined interpretations of those relations. These choices may not fully capture the variability, semantic differences, and context dependence of spatial language. Place name recognition and relation extraction from locality descriptions also remain challenging for deterministic and probabilistic methods due to vague phrasing and inconsistent recording practices, and would benefit from more robust extraction methods.

Future work. Future extensions will support non-point geometries (lines and polygons) and richer representations of uncertainty. We also aim to improve spatial relation parsing to capture both explicit and implicit relations. The LLM-based method could also be further improved through more systematic prompt design, including in-context learning with few-shot examples, and exploration of uncertainty estimation from model outputs. We further aim to develop a hybrid pipeline that combines the strengths of LLMs with probabilistic inference to georeference NGPs. Finally, we will expand the analysis to additional regions, particularly those with limited historical coverage in existing gazetteers, alongside improved methods for NGP recognition.

References

- 1 Elise Acheson, Stef De Sabbata, and Ross S. Purves. A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, 64:309–320, 2017. doi:10.1016/j.compenvurbsys.2017.03.007.
- 2 Edwin Aldana-Bobadilla, Alejandro Molina-Villegas, Ivan López-Arévalo, Shanel Reyes-Palacios, Victor Muñiz-Sanchez, and Jean Arreola-Trapala. Adaptive geoparsing method for toponym recognition and resolution in unstructured text. *Remote Sensing*, 12(18):3041, 2020. doi:10.3390/rs12183041.
- 3 Beatrice Alex, Claire Grover, Richard Tobin, and Jon Oberlander. Geoparsing historical and contemporary literary text set in the city of edinburgh. *Language Resources and Evaluation*, 53(4):651–675, 2019. doi:10.1007/s10579-019-09443-x.
- 4 Beatrice Alex, Clare Llewellyn, Claire Grover, Jon Oberlander, and Richard Tobin. Homing in on twitter users: Evaluating an enhanced geoparser for user profile locations. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*, pages 3936–3944, 2016. URL: <http://www.lrec-conf.org/proceedings/lrec2016/summaries/129.html>.
- 5 Ana Bárbara Cardoso, Bruno Martins, and Jacinto Estima. A novel deep learning approach using contextual embeddings for toponym resolution. *ISPRS Int. J. Geo Inf.*, 11(1):28, 2022. doi:10.3390/ijgi11010028.
- 6 Arthur Chapman and John Wieczorek. Georeferencing best practices, 2020. doi:10.15468/D0C-GG7H-S853.
- 7 Hao Chen, Maria Vasardani, and Stephan Winter. Georeferencing places from collective human descriptions using place graphs. *Journal of Spatial Information Science*, 17(1):31–62, 2018. doi:10.5311/JOSIS.2018.17.417.
- 8 Grant DeLozier, Jason Baldridge, and Loretta London. Gazetteer-independent toponym resolution using geographic word profiles. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1):2382–2388, February 2015. doi:10.1609/aaai.v29i1.9531.
- 9 Department of Conservation. Banks peninsula conservation walks, 2011. URL: <https://www.doc.govt.nz/globalassets/documents/parks-and-recreation/tracks-and-walks/canterbury/banks-peninsula-conservation-walks.pdf>.
- 10 Aneesha Fernando, Surangika Ranathunga, Kristin Stock, Raj Prasanna, and Christopher B. Jones. Georeferencing complex relative locality descriptions with large language models. *International Journal of Geographical Information Science*, 0(0):1–33, 2026. doi:10.1080/13658816.2026.2613291.
- 11 Jacques Fize, Ludovic Moncla, and Bruno Martins. Deep learning for toponym resolution: Geocoding based on pairs of toponyms. *ISPRS Int. J. Geo Inf.*, 10(12):818, 2021. doi:10.3390/ijgi10120818.
- 12 Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. What’s missing in geographical parsing? *Language Resources and Evaluation*, 52(2):603–623, 2018. doi:10.1007/s10579-017-9385-8.

- 13 Bo Han, Paul Cook, and Timothy Baldwin. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*, 49:451–500, 2014. doi:10.1613/jair.4200.
- 14 Linda L. Hill. *Georeferencing - The Geographic Associations of Information*. Digital libraries and electronic publishing. Mit Press, 2009.
- 15 Xuke Hu, Jens Kersten, Friederike Klan, and Sheikh Mastura Farzana. Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge. *International Journal of Geographical Information Science*, 40(3):670–697, 2026. doi:10.1080/13658816.2024.2405182.
- 16 Ehsan Kamalloo and Davood Rafiei. A coherent unsupervised model for toponym resolution. In *Proceedings of the 2018 World Wide Web Conference*, pages 1287–1296. ACM, 2018. doi:10.1145/3178876.3186027.
- 17 Olivier Van Laere, Steven Schockaert, Vlad Tanasescu, Bart Dhoedt, and Christopher B. Jones. Georeferencing wikipedia documents using data from social media sources. *ACM Trans. Inf. Syst.*, 32(3):12:1–12:32, July 2014. doi:10.1145/2629685.
- 18 Zekun Li, Wenxuan Zhou, Yao-Yi Chiang, and Muhao Chen. GeoLM: Empowering language models for geospatially grounded language understanding. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5227–5240, Singapore, December 2023. Association for Computational Linguistics. doi:10.18653/v1/2023.emnlp-main.317.
- 19 Zilong Liu, Krzysztof Janowicz, Ling Cai, Rui Zhu, Gengchen Mai, and Meilin Shi. Geoparsing: Solved or biased? an evaluation of geographic biases in geoparsing. In *Proceedings of the 25th AGILE Conference on Geographic Information Science, Vilnius, Lithuania, July 18-22, 2022*, volume 3, page 9. Copernicus Publications Göttingen, Germany, 2022. doi:10.5194/agile-giss-3-9-2022.
- 20 Kai Ma, Yongjian Tan, Zhong Xie, Qinjun Qiu, and Siqiong Chen. Chinese toponym recognition with variant neural structures from social media messages based on BERT methods. *Journal of Geographical Systems*, 24(2):143–169, 2022. doi:10.1007/s10109-022-00375-9.
- 21 Arnald Marcer, Elspeth Haston, Quentin Groom, Arturo H Ariño, Arthur D Chapman, Torkild Bakken, Paul Braun, Mathias Dillen, Marcus Ernst, Agustí Escobar, et al. Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. *Diversity and Distributions*, 27(3):564–567, 2021.
- 22 Jamie Scott, Kristin Stock, Fraser Morgan, Brandon Whitehead, and David Medyckyj-Scott. Automated georeferencing of antarctic species. In *11th International Conference on Geographic Information Science, GIScience 2021, Poznań, Poland (Virtual Conference), September 27-30, 2021 - Part II*, volume 208 of *LIPICs*, pages 13:1–13:16. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2021. doi:10.4230/LIPICs.GIScience.2021.II.13.
- 23 Praval Sharma, Ashok Samal, Leen-Kiat Soh, and Deepti Joshi. A spatially-aware data-driven approach to automatically geocoding non-gazetteer place names. *ACM Transactions on Spatial Algorithms and Systems*, 10(1):1–34, 2024. doi:10.1145/3627987.
- 24 Philip David Smart, Christopher B. Jones, and Florian A. Twaroch. Multi-source toponym data integration and mediation for a meta-gazetteer service. In *Geographic Information Science, 6th International Conference, GIScience 2010, Zurich, Switzerland, September 14-17, 2010. Proceedings*, volume 6292 of *Lecture Notes in Computer Science*, pages 234–248. Springer, Springer, 2010. doi:10.1007/978-3-642-15300-6_17.
- 25 Leonard Talmy. *Toward a cognitive semantics*, vol. 1 (concept structuring systems), 2000.
- 26 Marieke van Erp, Robert Hensel, Davide Ceolin, and Marian van der Meij. Georeferencing animal specimen datasets. *Transactions in GIS*, 19(4):563–581, 2015. doi:10.1111/tgis.12110.
- 27 Jimin Wang and Yingjie Hu. Are we there yet?: evaluating state-of-the-art neural network based geoparsers using EUPEG as a benchmarking platform. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities, GeoHumanities@SIGSPATIAL 2019, Chicago, IL, USA, November 5, 2019*, pages 2:1–2:6. Association for Computing Machinery, 2019. doi:10.1145/3356991.3365470.

- 28 John Wieczorek, Qinghua Guo, and Robert J. Hijmans. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International journal of geographical information science*, 18(8):745–767, 2004. doi:10.1080/13658810412331280211.
- 29 Yibo Yan and Joey Lee. Georeasoner: Reasoning on geospatially grounded context for natural language understanding. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024, Boise, ID, USA, October 21-25, 2024*, pages 4163–4167. ACM, 2024. doi:10.1145/3627673.3679934.
- 30 Madiha Yousaf and Diedrich Wolter. A reasoning model for geo-referencing named and unnamed spatial entities in natural language place descriptions. *Spatial Cognition & Computation*, 22(3-4):328–366, 2022. doi:10.1080/13875868.2021.2002872.
- 31 Bing Zhou, Lei Zou, Yingjie Hu, Yi Qiang, and Daniel W. Goldberg. Topobert: a plug and play toponym recognition module harnessing fine-tuned BERT. *International Journal of Digital Earth*, 16(1):3045–3064, 2023. doi:10.1080/17538947.2023.2239794.