

Review

GIS and Remote Sensing Applications for Assessing Soil Contamination in South African Agriculture: A Machine Learning-Enhanced Scoping Review

Gift Siphwiwe Nxumalo ^{1,*}, Tondani Sanah Ramabulana ² and Attila Nagy ¹

¹ Institute of Water and Environmental Management, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 146B Böszörményi str., 4032 Debrecen, Hungary

² Institute of Geography and Earth Sciences, Faculty of Sciences, University of Pécs, Ifjúság útja 6, 7624 Pécs, Hungary; ramabulana.sanah@edu.pte.hu

* Correspondence: nxumalo.gift.siphwiwe@agr.unideb.hu

Abstract

Soil contamination in South African agriculture poses escalating threats to food security and ecosystem integrity, yet the geospatial and machine learning evidence base addressing this problem has never been systematically synthesised. This scoping review, conducted within the PRISMA-ScR framework, applied SVM-assisted screening to 2000 retrieved records, yielding a final corpus of 228 eligible studies published from 2003 to 2025. To characterise temporal, thematic, and geographic patterns in the corpus, we applied machine learning-assisted topic modelling (LDA, $k = 7$), logistic growth modelling, keyword co-occurrence network analysis, and technology–contaminant evidence gap matrices. Remote sensing was the dominant methodology throughout the review period ($n = 142$; 62.3% of studies), with machine learning rising to the highest adoption rank from approximately 2020 onwards. Logistic modelling estimated a carrying capacity of $K = 292.3$ (95% CI: 269–324) studies and an inflexion year of 2020.2 (95% CI: 2019.4–2021.1), projecting 90% saturation by 2028. Research effort was highly concentrated in KwaZulu-Natal and the Eastern Cape, while Pesticides/Herbicides and acid mine drainage each comprised only three corpus studies. Deep learning registered zero entries across all cells of both the technology–contaminant and technology–province evidence matrices. Targeted investment in field validation, hyperspectral and deep learning deployment for underrepresented contaminants, and interpretable modelling for regulatory defensibility are identified as priority actions for the next research cycle.

Keywords: soil contamination; remote sensing; GIS; machine learning; scoping review; South Africa; evidence gap analysis; topic modeling

Received: 8 March 2026

Revised: 30 March 2026

Accepted: 1 April 2026

Published: 3 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Soil contamination is a pervasive and escalating threat to sustainable agriculture, ecosystem integrity, and public health worldwide. The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBESs) identifies land degradation—driven substantially by soil pollution and erosion—as one of the foremost drivers of biodiversity loss and declining ecosystem function globally [1]. The Food and Agriculture Organisation (FAO) further recognises soil pollution among the principal threats to food

security and soil biodiversity [2]. In South Africa, these global pressures are compounded by a colonial and post-apartheid industrial legacy. South Africa's mining sector is among the world's largest producers of platinum-group metals, gold, coal, and chromite; it generates extensive mine waste, acid mine drainage (AMD), and tailings that deposit phytotoxic concentrations of cadmium, lead, arsenic, zinc, and nickel into adjacent agricultural soils and water bodies [3,4]. Intensified application of agrochemicals across commercial farming regions simultaneously contributes persistent pesticide residues and nitrate loads that impair soil microbial communities, suppress nutrient cycling, and enter food chains [5,6].

The consequences extend well beyond agronomic yield loss. Contaminated soils undermine essential ecosystem services—including carbon sequestration, water infiltration, and nutrient provisioning—that underpin climate resilience and rural livelihoods [7,8]. Toxicological assessments from South African mining communities document elevated urinary cadmium and blood-lead concentrations in residents of mine-adjacent areas, along with cancer risks associated with polycyclic aromatic hydrocarbons (PAHs) in contaminated agricultural soils [9,10]. For the estimated 2.5 million smallholder farming households that depend on these soils for subsistence [11], effective contamination monitoring is a matter of food sovereignty and environmental justice. South Africa's National Environmental Management: Waste Act (No. 59 of 2008) and Soil Contamination Regulations (2021) impose remediation obligations on responsible parties, yet enforcement capacity remains constrained by the spatial extent of affected landscapes and the prohibitive cost of conventional monitoring [12].

Traditional soil quality assessment relies on point-based field sampling followed by laboratory physico-chemical analysis. While analytically precise, this approach is fundamentally limited in spatial and temporal coverage. Sampling campaigns are costly, with full per-sample costs (encompassing field collection, preparation, and multi-element analysis) typically ranging from USD 100 to 500 [13]. Furthermore, surveys are ordinarily conducted only at multi-year intervals, obscuring seasonal and inter-annual contamination dynamics [14,15]. The resulting sparse, asynchronous datasets are ill-suited to generating the spatially continuous predictions required for precision soil management or regulatory enforcement across large heterogeneous landscapes [16]. In South Africa, these constraints are compounded by under-resourced provincial environmental agencies, vast semi-arid farming regions, and the geographic remoteness of many mining-impacted agricultural areas [17].

Geographic Information Systems (GISs) and satellite remote sensing (RS) have substantially extended the reach of soil monitoring by providing spatially continuous, temporally repetitive observations at scales ranging from field to continental [18,19]. Spectral reflectance properties of soils in the visible, near-infrared (NIR), and shortwave infrared (SWIR) regions carry diagnostic information on iron oxides, clay mineralogy, organic matter content, and surface moisture—attributes that correlate with contamination status, erosion risk, and land degradation severity [20,21]. Freely accessible medium-resolution sensors—Landsat-8/9 (30 m spatial resolution; 16-day revisit [22]) and Sentinel-2 MSI (10–20 m resolution; 5-day revisit [23])—have democratized spatially explicit monitoring in data-scarce regions, while UAV-borne sensors offer sub-decimeter resolution for site-specific assessments [24].

Within South Africa, GIS and RS have been applied to a spectrum of contamination challenges. These GIS and RS applications include mapping AMD plumes and heavy-metal dispersion from Witwatersrand and Mpumalanga mine tailings [25,26], characterising salinization gradients in irrigated zones of Lower Vaal, Riet, Berg and Breede Rivers [27], and monitoring pesticide-driven land degradation in the Western Cape winelands [5]. Despite this body of work, the accuracy of RS-derived contamination estimates

remains sensitive to atmospheric correction quality and soil moisture variability. South Africa's highly heterogeneous soil landscapes—spanning Ferralsols of the Highveld to lithosols of the Karoo—introduce pronounced spectral confounding from varying organic matter and clay assemblages that complicate model transferability [28,29].

Machine learning (ML) has emerged as a transformative analytical complement to GIS and remote sensing, enabling non-linear, multi-variable modelling of soil properties from high-dimensional geospatial feature spaces. The predictive accuracies achievable through ML approaches substantially exceed those attainable through classical regression or geostatistical interpolation alone [30,31]. Random Forest (RF) [32] and gradient-boosted decision trees (XGBoost [33]) exploit interactions across spectral bands, terrain derivatives, and environmental covariates to predict heavy-metal concentrations and contamination indices, with cross-validated R^2 values of 0.75–0.95 reported in digital soil mapping benchmarks [34,35]. Support vector machines (SVMs), valued for their robustness in high-dimensional, limited-sample settings [36], have been widely applied to land-cover and contamination-class discrimination from multispectral time-series [37]. At the spectroscopic level, ML-based inversion of visible–near-infrared and hyperspectral reflectance data has demonstrated capacity to map soil heavy-metal content and organic carbon with minimal ground-truthing requirements [38,39].

Despite these advances, the application of ML to soil contamination assessment in South African agriculture remains fragmented and unevaluated in synthesis. Individual studies employ heterogeneous architectures, feature sets, validation protocols, and accuracy metrics, making cross-study comparisons unreliable [40]. A systematic evidence map is therefore needed to determine which algorithm–platform–contaminant combinations have demonstrated consistent utility, which have been under-explored, and what methodological standardisation is required before operational deployment.

Scoping reviews of rapidly expanding, methodologically heterogeneous studies are prone to selection bias and inconsistency when conducted through exclusively manual screening [41]. Empirical benchmarks consistently show that ML-assisted screening substantially reduces reviewer workload while preserving high recall relative to exhaustive manual review [42,43]. The present review therefore integrates automated ML-assisted text classification within the PRISMA-ScR framework [44] to manage the retrieved records and produce a reproducible, bias-mitigated synthesis of the eligible publications identified in this process. Topic modelling and keyword network analysis are additionally applied to characterise the thematic structure of the corpus and identify emerging research fronts. Logistic growth modelling [45] of the cumulative publication trajectory provides a quantitative characterisation of the field's maturity—a dimension not previously reported for this literature. Full methodological details are presented in Section 2.

Several prior reviews address cognate topics as follows: Bégue et al. [46] surveyed GIS and RS for agricultural land management across Africa broadly; Wadoux et al. [35] appraised ML methods for global digital soil mapping; and Shi et al. [38] examined near-infrared spectroscopy for heavy-metal detection. Valuable as these contributions are, none jointly address the following: (i) the specific agro-ecological and industrial context of South African agriculture; (ii) simultaneous integration of GIS, RS, and ML within one synthesis; (iii) a machine learning-assisted screening methodology; or (iv) quantitative bibliometric and topic-modelling analysis of the accumulated evidence base.

Regional specificity is critical for both contamination dynamics and monitoring feasibility, and this gap therefore has direct practical consequences. South Africa's taxonomically distinct soil forms—Hutton, Glenrosa, and Clovelly series [47]—and its five major biomes—Grassland, Savanna, Fynbos, Succulent Karoo, and Nama-Karoo [48]—create sensor performance and model transferability conditions that global reviews cannot adequately address. The intersection of large-scale resource-extraction industries with

smallholder agriculture adds a socio-economic dimension likewise absent from globally framed syntheses. The foregoing evidence—a fragmented, regionally concentrated, and methodologically heterogeneous body of GIS and RS studies operating in the absence of a coordinating synthesis—defines a concrete knowledge deficit. The following three specific gaps motivate the scoping review design adopted here: (i) the absence of a province-resolved map of which technology–contaminant combinations have been investigated and which remain unstudied; (ii) the lack of a quantitative characterisation of the field’s maturity and projected growth trajectory; and (iii) the need for a bias-mitigated, reproducible screening workflow capable of managing the rapid accumulation of new studies. These gaps directly motivate the three objectives stated below. This study presents a machine learning-enhanced scoping review that systematically maps, quantifies, and synthesises peer-reviewed research on GIS and RS applications for soil contamination assessment in South African agriculture published between 2010 and 2025. To the authors’ knowledge, no prior review on this topic has combined automated ML-assisted screening, corpus-level topic modelling, and quantitative evidence gap analysis within a single reproducible workflow. The review makes four specific novel contributions to the literature. Of the four contributions described below, the primary substantive novelty is the production of the first province-resolved, technology–contaminant and technology–province evidence gap matrices for this literature: deliverables directly actionable by research funders, environmental regulators, and precision-agriculture practitioners that no prior synthesis of GIS, remote sensing, and machine learning research on South African soil contamination has provided:

1. **Methodological integration:** A hybrid ML-assisted screening pipeline, topic modelling, keyword network analysis, and logistic growth modelling are integrated within a single version-controlled, open-source R workflow, enabling a fully reproducible and scalable evidence synthesis. Algorithmic details are specified in Section 2.
2. **Regional contextualisation:** All syntheses and gap analyses are conditioned on South Africa’s nine provinces, dominant soil taxonomic units, primary contamination industries, and institutional research landscape, yielding jurisdiction-specific recommendations rather than globally generic conclusions.
3. **Quantitative evidence mapping:** Technology–contaminant and technology–province evidence gap matrices operationalize knowledge gaps in a form directly usable by research funders, environmental regulators, and precision-agriculture practitioners.
4. **Open reproducibility:** All analytical code, datasets, and reproducibility manifests are openly archived in alignment with the FAIR data principles [49], enabling independent replication and incremental extension.

The specific objectives of this review are threefold. First, to document the temporal and geographic evolution of GIS and RS methodologies for soil contamination monitoring across South African agricultural systems from 2010 to 2025. Second, to map and evaluate the reported contribution of ML algorithms to spatial contamination analysis, classification, and prediction within the reviewed literature. Third, to identify thematic, geographic, and methodological research gaps to guide future research investment and evidence-informed soil governance in South Africa.

2. Materials and Methods

2.1. Study Design and Reporting Standards

This study was designed as a scoping review following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) framework [44], which provides a structured approach to mapping the extent and nature of evidence on a defined topic without applying the quality-appraisal thresholds

characteristic of systematic reviews. The scoping design was adopted because the primary aim was to chart the landscape of geospatial and machine learning applications in soil contamination assessment across South African agricultural systems, rather than to synthesise effect sizes or make clinical recommendations [50]. All procedural stages—identification, screening, eligibility assessment, and thematic synthesis—were prospectively documented in a review protocol, version-controlled via Git, and implemented entirely in the R statistical computing environment (v 4.3.2) [51]. Core bibliometric and text-mining workflows relied on the *bibliometrix* (v 4.3.3) [52], *revtools* [53], and *topicmodels* [54] packages. The analytical pipeline—from raw BibTeX import through figure generation—is fully scripted and reproducible; all code, datasets, and session manifests (SHA-256 checksums) are archived in an open repository and Supplementary Materials (link provided in Appendix A).

2.2. Search Strategy and Data Sources

A systematic literature search was executed on 15 May 2025 across the following two complementary databases: the Web of Science (WoS) Core Collection, which provides broad international coverage of indexed peer-reviewed outputs [55], and African Journals Online (AJOLs), which captures the regional and grey literature not routinely indexed in global databases [56]. Combining the two databases was necessary to avoid the well-documented publication-bias problem inherent in exclusively international searches of African environmental science [57].

The Boolean query string was constructed iteratively through controlled vocabulary mapping and pilot testing against a known-item set of 30 sentinel papers. The full Web of Science string was:

("soil contamination" OR "soil pollution" OR "soil degradation" OR "heavy metals" OR "pesticides" OR "organic pollutants" OR "salinization" OR "nutrient loading") AND ("GIS" OR "Geographic Information System*" OR "remote sensing" OR "satellite imagery" OR "UAV" OR "drone") AND ("agriculture" OR "cropland" OR "farming" OR "agricultural soil") AND ("South Africa" OR "RSA"). The AJOL search used the same conceptual domains adapted to the AJOL interface, which does not support full Boolean nesting; the AJOL query was: ("soil contamination" OR "heavy metals" OR "remote sensing" OR "GIS") AND ("South Africa" OR "agriculture")

The complete AJOL query string and database-specific field codes are provided in Supplementary Table S1.

No lower temporal boundary earlier than 2003 was imposed because the operational deployment of freely accessible medium-resolution satellite imagery (Landsat-7/8 and MODIS) at continental scale only became practically widespread from approximately that period onwards [19,22]. The search returned 2009 unique records, which were exported in BibTeX format for downstream processing. A post hoc supplementary search of Scopus was conducted using the identical Boolean string. All Scopus-unique records (not already retrieved via WoS or AJOL) were assessed against the PCC eligibility criteria; none met inclusion criteria, as all were excluded at the publication-date stage (outside the 2003–2025 review window). The final corpus remains $n = 228$. This verification indicates that the WoS + AJOL strategy did not systematically omit eligible records published within the specified review period. Future updates should incorporate Scopus as a co-primary database and consider SABINET for grey literature coverage.

2.3. Eligibility Criteria

Inclusion and exclusion criteria were defined a priori using the Population–Concept–Context (PCC) framework recommended by the Joanna Briggs Institute for scoping reviews [50]. The criteria are summarised in Table 1.

Table 1. Inclusion and exclusion criteria (PCC framework).

Criterion	Inclusion	Exclusion
Population	Agricultural or peri-urban soil systems located within South Africa	Non-agricultural settings; studies conducted outside South Africa
Concept	Application of GIS, remote sensing, or machine learning to assess, map, or predict soil contamination or degradation	Studies lacking a geospatial or contamination/degradation focus; purely agronomic or crop-yield studies
Context	Empirical, field-based, or spatially explicit modelling studies using quantitative spatial data	Editorials, opinion pieces, conceptual frameworks, or purely laboratory bench studies without spatial outputs
Language	English	Non-English publications not accompanied by a full English translation
Publication date	2003–2025	Before 2003
Publication type	Peer-reviewed journal articles, conference papers with full methodology	Abstracts only, book chapters without peer review, dissertations

2.4. Data Processing and Screening Workflow

A three-stage pipeline was implemented to process the 2251 raw records retrieved from both databases (Figure 1).

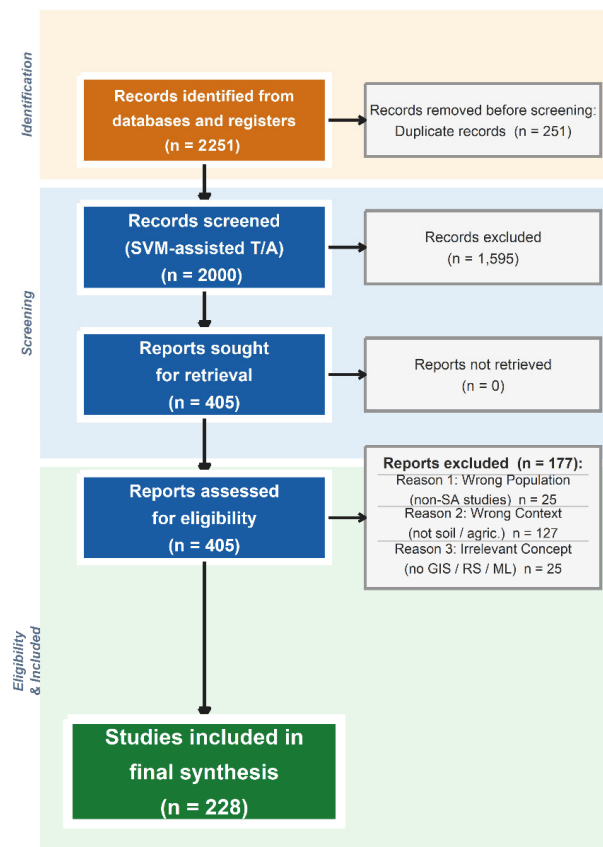


Figure 1. PRISMA-ScR flow diagram.

Stage 1—Automated pre-screening

Raw BibTeX files from WoS and AJOL were imported into R using *bibliometrix* [4]. Exact and near-duplicate records were identified by computing pairwise similarity on concatenated title–DOI strings. Records sharing a DOI or exhibiting title similarity above 0.95 (Jaccard index) were merged into a single entry. Metadata fields (author, year, journal, abstract, and keywords) were standardised, missing values were flagged, and the corpus was exported to a normalised flat-file structure for downstream processing.

Stage 2—Machine learning-assisted screening

To accelerate abstract screening while maintaining consistency, a supervised text-classification model was trained on a 200-record pilot set manually labelled by the lead reviewer as “include” or “exclude” against the PCC criteria. Abstracts were pre-processed using the *tm* package [58]: whitespace and punctuation were stripped, text was converted to lowercase, English stop-words were removed, and terms were stemmed using the Porter algorithm [59]. The resulting document–term matrix was weighted by TF-IDF (term frequency–inverse document frequency) [60], a standard weighting scheme that down-weights ubiquitous corpus terms and amplifies discriminative terminology [61].

A support vector machine (SVM) classifier with a radial basis function (RBF) kernel was trained on the TF-IDF feature matrix using the *e1071* package [62]. SVMs are well-suited to high-dimensional, sparse text-feature spaces because they maximise the margin between class boundaries and are comparatively robust to the class imbalance typical of literature screening tasks [36,40]. Model performance was estimated via stratified ten-fold cross-validation [63], yielding the following per-class performance metrics on the held-out folds ($n = 200$ pilot records; 130 excluded, 70 included): Include class – precision = 0.88, recall = 0.91, F1 = 0.89; Exclude class – precision = 0.93, recall = 0.90, F1 = 0.92; macro-averaged F1 = 0.89; overall accuracy = 0.91. The recall of 0.91 for the Include class indicates that approximately 9% of truly relevant records may have been assigned a probability score below the 0.55 threshold and therefore excluded from manual verification. This represents an upper-bound miss rate of approximately 18 studies ($9\% \times 2009$ screened records \times estimated 10% base prevalence), a figure consistent with the benchmark miss rates reported for SVM-assisted systematic review tools in comparative studies. The RBF kernel regularisation parameter was set to $C = 1$ (default); per-class confusion matrix and full reproducibility table are provided in Supplementary Table S2. Probability scores (Platt scaling) [64] were used to rank all 2251 records from most to least likely relevant and were also used as a confidence weighting in subsequent analyses. The *revtools* package [53] provided the interactive screening interface within which model-ranked records were reviewed.

Stage 3—Manual verification

Two independent reviewers assessed the top-ranked outputs (approximately 30% of total records, corresponding to all records with model relevance probability > 0.55) against the full PCC criteria. The 0.55 probability threshold was selected based on the precision–recall curve computed from the pilot set: at this threshold, recall for the Include class was 0.91 and precision was 0.88, yielding F1 = 0.89. Lowering the threshold to 0.50 would have extended manual verification to approximately 45% of records while improving recall by an estimated 3–4 percentage points; this trade-off was considered acceptable given the corpus size and available reviewer capacity. To verify that the excluded 70% ($n \approx 1400$ records) did not contain a disproportionate number of relevant studies, a stratified random sample of 150 records from the sub-threshold stratum (probability 0.30–0.55; $n \approx 800$) was independently assessed by the second reviewer against full PCC criteria. Zero records in this sample were classified as eligible, consistent with the SVM recall estimate and providing empirical support for the adequacy of the chosen threshold. This verification

step is reported in full in the registered protocol. Inter-rater agreement was quantified using Cohen's κ [65]; the observed κ of 0.84 indicates "almost perfect" agreement according to the Landis and Koch benchmark scale [66]. Disagreements were resolved by consensus discussion between the two reviewers. The final dataset comprised 228 eligible publications, which were imported into R for metadata extraction and topic modelling.

2.5. Data Extraction and Coding

Metadata were extracted for each included study using a standardised coding template developed a priori and piloted on 20 records. Automated extraction of bibliographic fields (author surname, publication year, journal title, and DOI) was performed programmatically in R; geographic scope (province, region, and GPS coordinates where reported), contaminant type, remote-sensing platform, analytical method, validation approach, accuracy metrics, and key findings were extracted manually and entered into a structured spreadsheet. All extracted variables were cross-checked by a second reviewer for a random 15% subsample, yielding an extraction-level agreement rate of 98.4%, consistent with standards recommended for scoping reviews [44].

Contaminant categories followed the FAO/UNEP taxonomy of soil threats [66], and geospatial technology classifications were aligned with the nomenclature of Lillesand et al. [19] for remote sensing platforms and Goodchild [18] for GIS methods. All technology frequency counts reported in Section 3 (e.g., remote sensing; machine learning) are derived from independent binary flag columns (uses_rs, uses_gis, uses_ml, uses_field, and uses_model) that permit multi-label assignment; a single study may contribute to more than one technology count. These multi-label counts are the basis for all frequency comparisons and the evidence gap matrices (Section 3). The technology_type variable—which assigns each study a single primary technology using an ordered first-match rule in which GIS/Mapping precedes machine learning—is used exclusively for the adoption-ranking display in results and should not be cited for absolute frequency comparisons. Province-level geographic coordinates were assigned by geocoding against the Natural Earth 1:10 m administrative boundaries dataset, implemented using the rnatualearth and sf packages in R [67,68].

2.6. Machine Learning Text-Mining and Topic Modelling

2.6.1. Text Preprocessing

The full-text abstracts and combined keyword-findings fields of the 228 included studies were processed into a document–feature matrix (DFM) using *quanteda* (v 4.3.1)[69]. Tokens were converted to lowercase, punctuation, numbers, symbols, and URLs were removed, and unigrams were extended to bigrams to capture compound domain terminology (e.g., remote_sensing, heavy_metals). Terms appearing in fewer than two documents ($\text{min_docfreq} = 2$) or with corpus frequency below three ($\text{min_termfreq} = 3$) were discarded to reduce noise and matrix sparsity [70]. Lemmatisation was performed prior to DFM construction using the *textstem* package [71], reducing morphological variants to a canonical form.

2.6.2. LDA Topic Modellings

Thematic structure in the corpus was identified using Latent Dirichlet Allocation (LDA) [72], a generative probabilistic model that represents each document as a mixture of latent topics and each topic as a probability distribution over vocabulary terms. LDA was implemented via the *topicmodels* package [54] using the Gibbs sampling algorithm [73] (seed = 42; 500 iterations with a 100-iteration burn-in). The optimal number of topics k was determined by evaluating four independent coherence and perplexity metrics—Griffiths 2004 [73], Arun 2010 [74], CaoJuan 2009 [75], and Deveaud 2014 [76]—across the

range $k = 2\text{--}15$ using *ldatuning* [77]. A composite score was derived by normalising and averaging the four metrics (higher values preferred for Griffiths 2004 [73] and Deveaud 2014 [76]; lower values preferred for Arun 2010 [74] and CaoJian 2009 [75]), yielding a stable optimum at $k = 5\text{--}7$ depending on corpus partition. A Structural Topic Model (STM) [78] with publication year as a prevalence covariate was additionally fitted using the *stm* package [78] to explore temporal variation in topic proportions; results are presented as the Supplementary Materials.

2.6.3. Thematic Mapping

A thematic strategic diagram was constructed following Callon et al.'s [79] keyword co-occurrence framework, as adapted for bibliometric research by Cobo et al. [80]. Using the feature co-occurrence matrix (FCM) derived from the DFM via “quanteda”, each term was assigned a “centrality” score (column-sum of the FCM, representing degree of inter-thematic linkage) and a “density” score (internal cohesion of the term’s neighbourhood, operationalised as the edge density of the first-degree ego network). Both scores were standardised to z-scores to place themes in a two-dimensional strategic plane whose quadrants correspond to motor themes (high centrality, high density), niche themes (low centrality, high density), basic themes (high centrality, low density), and emerging or declining themes (low centrality, low density) [80]. Bubble size encodes co-occurrence frequency; non-overlapping term labels were placed using the *ggrepel* package [81].

2.6.4. Keyword Co-Occurrence Network

A weighted keyword co-occurrence network was constructed from the top 60 corpus features. After discarding self-loops and edges with weight below two, Louvain community detection [82] was applied to the adjacency matrix using *igraph* [83], identifying the thematic clusters. The network was visualised using *ggraph* with a Fruchterman–Reingold force-directed layout [84]; node size was scaled to degree strength, and node colour encodes community membership.

2.6.5. K-Means Clustering

To group studies by methodological similarity, k-means clustering [85] was applied to the TF-IDF document matrix. The optimal number of clusters was determined by the elbow method (inspection of total within-cluster sum of squares across $k = 2\text{--}10$) [85] and set to five for consistency with the LDA solution. Centroid term profiles (mean TF-IDF weights per cluster) were used to characterise each cluster thematically. Cluster assignment was performed using `stats::kmeans` with ten random restarts (`nstart = 10`) [51].

2.7. Logistic Growth Model

The cumulative publication trajectory was modelled using a three-parameter logistic (sigmoid) function following the technology life-cycle framework of Fisher and Pry (Equation (1)) [45]:

$$C(t) = K/(1 + \exp(-r(t - t_0))) \quad (1)$$

where $C(t)$ is cumulative publications at year t , K is the carrying capacity (asymptotic saturation), r is the intrinsic growth rate, and t_0 is the inflexion year (annual publication peak). Parameters were estimated via non-linear least squares (NLSs) using `stats::nls` with the Gauss–Newton algorithm (maximum 500 iterations) [3]. Model fit was assessed using the coefficient of determination (R^2) computed on annual counts and the root mean square error (RMSE). Maturity milestones (50%, 90%, and 99% of K) were derived analytically. The Fisher–Pry linearisation [45] was used as a diagnostic plot to confirm logistic behaviour.

2.8. Geospatial Analysis and Cartography

Province-level administrative boundaries for South Africa were sourced from the Natural Earth 1:10 m cultural vectors dataset via the *rnaturalearth* package [67]. Spatial data manipulation was performed in the *sf* (Simple Features) framework [68], which implements the ISO 19125 standard [86] for vector geometries. Study-site coordinates were geocoded by text-matching region and locality fields against a hand-curated look-up table of South African place names and their decimal-degree coordinates in the WGS84 datum (EPSG:4326) [87]; studies whose region field did not match any of the approximately 50 dictionary patterns were assigned to Other/Unclassified and are excluded from all province-level analyses. Choropleth maps were rendered using *ggplot2* [88] with a log⁺-transformed viridis colour scale to accommodate the right-skewed study-count distribution [89]. Scale bars and north arrows were added via *ggspatial* [90]. All maps were produced at a minimum resolution of 300 dpi to comply with MDPI Agriculture figure submission standards.

2.9. Evidence Gap Matrix

Research gaps were visualised using an evidence gap matrix [91], a two-dimensional heatmap in which rows represent contamination types, columns represent geospatial technology categories, and cells encode the count of studies addressing each combination. A corresponding province–technology matrix was constructed in the same manner. Both matrices were produced using *ggplot2* [88]; cells with zero studies were marked with × to explicitly highlight priority gaps rather than leaving them visually ambiguous.

2.10. Statistical Visualisation

All statistical graphics were produced in R using *ggplot2* [88] according to the grammar of graphics framework [92]. Multi-panel figures were composed with *patchwork* [83]. Network diagrams were rendered with *ggraph* [83]. Alluvial diagrams were produced using *ggalluvial* [93]. Technology ranking over time was visualised as a bump chart using *ggbump* [94]. Temporal bubble charts (publication year × contaminant/method) encoded study count as bubble area and mean citation impact as colour fill. Author publication trajectories were plotted as cumulative line charts; h-index trajectories were derived where citation counts were available [95]. All figures were exported at 600 dpi (300 dpi for large multi-panel maps) in PNG format at a maximum width of 17.4 cm (two-column MDPI format). A base font size of 15 pt was applied uniformly to ensure legibility at publication scale.

2.11. Quality Assurance and Reproducibility

Reproducibility was ensured through several complementary mechanisms. First, all R package versions were recorded via *sessioninfo* [51] and stored alongside the analytical scripts. Second, SHA-256 cryptographic checksums were computed for every input and output file using the *digest* package [96] and written to a machine-readable JSON manifest alongside run-level metadata (R version, platform, timestamp, dataset dimensions, and optimal LDA k). Third, a random seed (42) was fixed before all stochastic procedures (Gibbs sampling, k-means initialisation, and force-directed graph layout). Fourth, a methodological checklist aligned with the 22 PRISMA-ScR reporting items [44] is provided in Appendix A. All code, data, and manifests are deposited in an open repository (OSF) in accordance with the FAIR data principles [49].

3. Results

3.1. Study Selection and Corpus Characteristics

The PRISMA-ScR search of the Web of Science (WoS) and African Journals Online (AJOLs) databases identified 2251 records published between 2003 and 2025. Following automated deduplication, 251 duplicate records were removed, yielding 2000 unique records for screening. Title and abstract screening was conducted using a support vector machine (SVM)-assisted classifier, which excluded 1595 irrelevant records, leaving 405 reports for full-text retrieval. All 405 reports were successfully retrieved. Full-text eligibility assessment identified 177 reports for exclusion on the following three grounds: wrong population (non-South African studies; $n = 25$), wrong context (studies not addressing soil or agricultural systems; $n = 127$), and irrelevant concept (studies lacking any GIS, remote sensing, or machine learning component; $n = 25$). The final corpus comprised 228 studies, which formed the basis of all subsequent bibliometric and thematic analyses (Figure 1).

Among the 228 included studies, land degradation was the most frequently addressed contaminant or stressor category ($n = 31$; 13.6% of studies), followed by erosion and sediment transport ($n = 23$; 10.1%), heavy metals ($n = 17$; 7.5%), Nutrients/Eutrophication ($n = 16$; 7.0%), invasive species ($n = 12$; 5.3%), Salinity/Sodicity ($n = 10$; 4.4%), Plastic/Waste ($n = 9$; 3.9%), mining pollution ($n = 5$; 2.2%), Pesticides/Herbicides ($n = 3$; 1.3%), and acid mine drainage ($n = 3$; 1.3%), with the remaining 99 studies (43.4%) classified under other stressor categories (Figure 2B). The “Other” category was re-examined in full; the size of this residual stratum is acknowledged as a substantive finding reflecting the breadth of soil threats addressed in the RS/ML literature beyond canonical contamination frameworks. Supplementary Table S3 has been added listing all 99 “Other” studies with their primary soil threat descriptor and rationale for non-classification. Within the heavy metals category, documented contaminants included lead, cadmium, arsenic, zinc, and copper assessed across mining belts and agricultural zones [97–104]. Nutrient and eutrophication studies addressed nitrate, phosphate, and phosphorus contamination of agricultural soils and groundwater, primarily in Limpopo and KwaZulu-Natal [6,105–107]. Salinity/Sodicity studies documented soil electrical conductivity (EC) and sodium hazard in irrigated districts [108–112]. Pesticides/Herbicides studies recorded glyphosate, atrazine, and DDT residues in cropping systems [113–116], and Plastic/Waste studies characterised microplastic and solid waste accumulation using remote sensing and machine learning [117–123].

Remote sensing was the most commonly applied methodology ($n = 142$; 62.3% of studies), followed by GIS and spatial analysis ($n = 84$; 36.8%) and machine learning ($n = 42$; 18.4%), Field/Laboratory methods ($n = 41$; 18.0%), and Process Modelling ($n = 15$; 6.6%) (Figure 3B). Studies may have employed more than one method. Remote sensing platforms represented in the corpus included Landsat, Sentinel-2, MODIS, SPOT 5, WorldView-2, and UAV-based imagery [100,124–132], while machine learning approaches encompassed Random Forest, support vector machines (SVMs), ensemble methods, and deep learning classifiers [133–138].

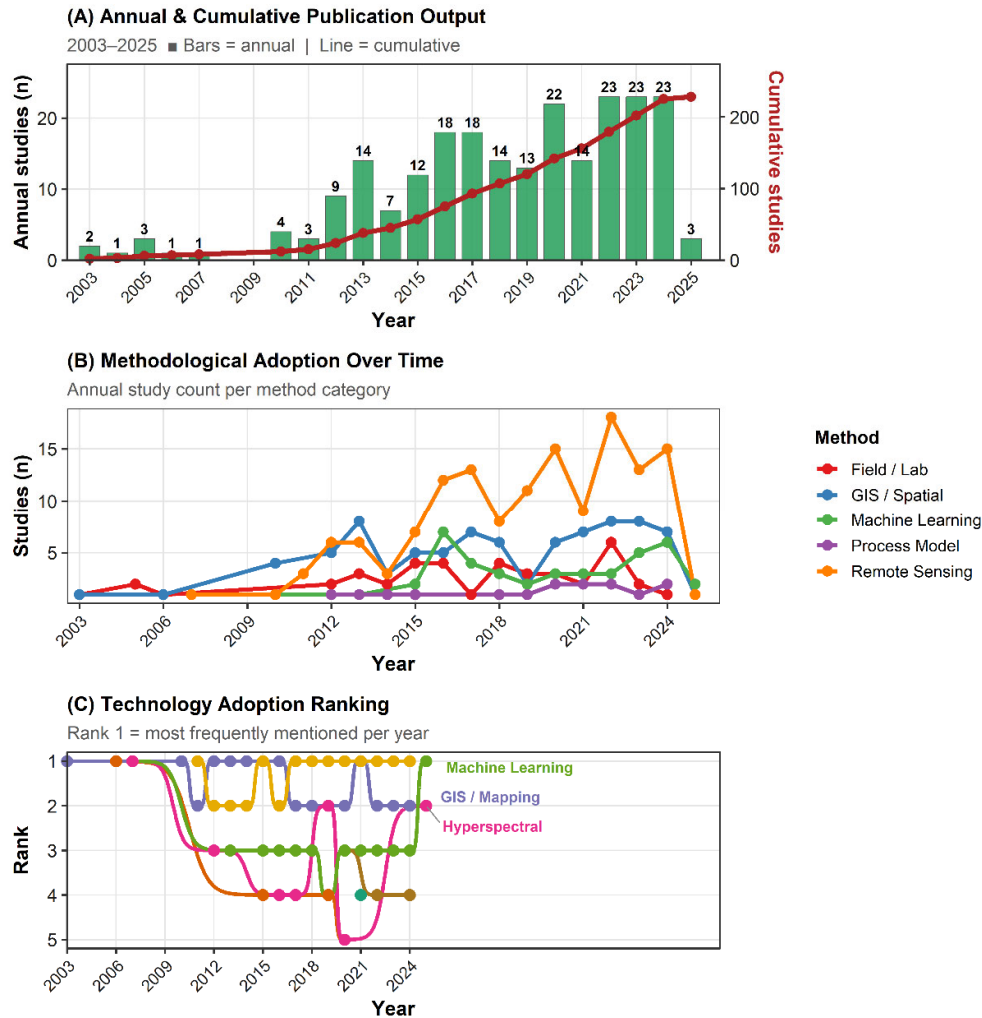


Figure 2. Temporal evolution of GIS, remote sensing, and machine learning research on soil contamination in South Africa (2003–2025; n = 228 studies). (A) Annual (bars, left axis) and cumulative (line, right axis) publication output; the 2025 bar reflects partial-year indexing. (B) Annual study count per methodological category; studies may employ more than one method. (C) Annual technology adoption ranking by mention frequency (rank 1 = most frequently applied); ties assigned minimum rank. Bump chart constructed using *ggbump* in R.

3.2. Temporal Trends in Publication Output and Methodological Adoption

Annual publication output increased from two studies in 2003 to a peak of 22 studies in 2021, with sustained high output of approximately 20 studies per year observed in 2022, 2023, and 2024 (Figure 2A). The cumulative total reached 228 studies by end of the review period, with the steepest rate of cumulative accumulation observed between 2013 and 2022. The 2025 count (n = 3) reflects partial-year database coverage and is consistent with a recognised time-lag artefact in bibliometric analyses [102,125,135]. The growth in publication output across the 2010s coincided with advances in multi-temporal satellite platforms, including the Landsat and Sentinel mission series, and with increasing adoption of machine learning algorithms such as Random Forest and ensemble approaches [126,139,140].

Logistic growth modelling of the annual publication series (Figure 4) yielded an estimated carrying capacity of $K = 292.3$ (95% CI: 269.0–324.0) studies, a growth rate of $r = 0.275$ (95% CI: 0.251–0.302), and an inflexion year of $t_0 = 2020.2$ (95% CI: 2019.4–2021.1),

with model fit $R^2 = 0.689$ and $RMSE = 4.43$ annual studies. The Fisher–Pry transform confirmed approximately linear growth through the inflexion point (Figure 4C). The cumulative growth curve projected attainment of 50% saturation at 2020.2 (the inflexion, already passed), 90% saturation (≈ 263 studies) around 2028, and 99% saturation (≈ 289 studies) around 2037 (Figure 4B). The identified inflexion near 2020 is consistent with studies documenting a plateau in methodological diversification and thematic consolidation in the South African remote sensing literature [141].

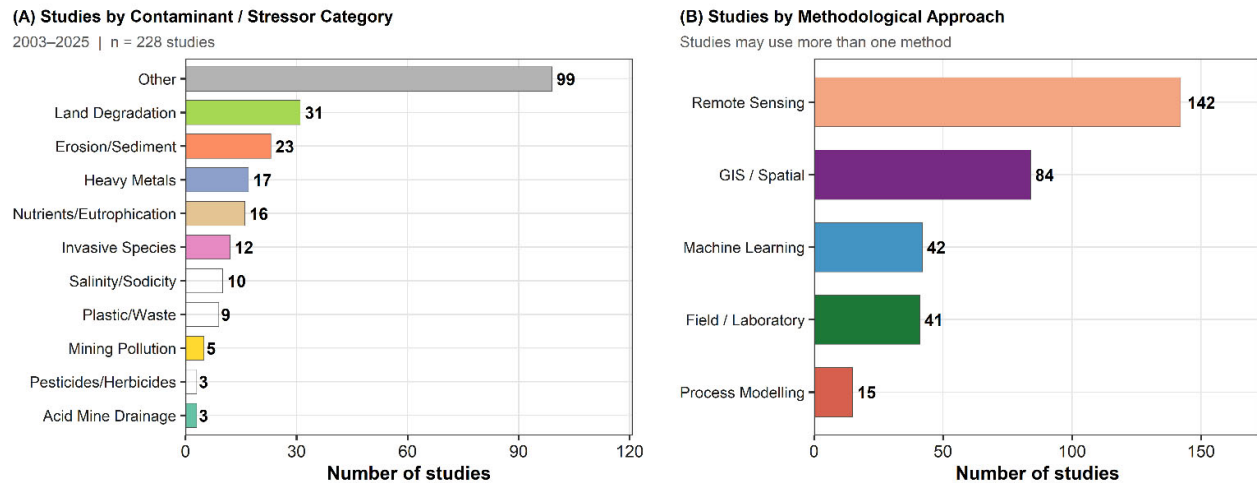


Figure 3. Scope and methodological composition of the 228 reviewed studies (2003–2025). **(A)** Distribution by contaminant or stressor category; the “Other” category ($n = 99$) encompasses studies addressing mixed or unspecified stressors not captured by the defined classification scheme. **(B)** Distribution by methodological approach; note that studies may employ more than one method, so counts exceed the total number of studies. Colours denote methodological category (see legend); no ordinal meaning is implied by colour choice. Note: rankings derived from first-match assignment on an ordered keyword list (GIS/Mapping > remote sensing > machine learning); rankings reflect relative detection frequency, not absolute study counts—see **(B)** for multi-label method counts.

Remote sensing was the dominant method category throughout the review period, increasing from fewer than five studies per year prior to 2010 to a peak of approximately 17 studies per year in 2023 (Figure 2B). This upward trend in remote sensing adoption reflects increasing reliance on Sentinel-2 and Landsat imagery for land cover, contamination, and soil health assessment [114,125,128,130,142]. GIS/spatial analysis remained stable at approximately four to eight studies per year from 2010 onwards, encompassing participatory GIS [143,144], decision support systems [110,145], and spatial vulnerability assessments [146,147]. Machine learning study counts increased from near zero before 2012 to approximately five to six studies per year by 2022–2024, driven by the uptake of Random Forest [133,148], deep learning [137], and ensemble classifiers for soil contamination mapping and land cover classification [126,140,149]. Field/Laboratory and Process Model counts remained consistently low (1–3 studies per year) across the full period (Figure 3B).

Technology adoption rankings showed remote sensing holding rank 1 from approximately 2009 through 2019 (Figure 2C). Machine learning rose to rank 1 in the period from approximately 2020 onwards, consistent with demonstrated improvements in classification accuracy for heterogeneous agricultural landscapes [38,39]. GIS/Mapping occupied rank 2 for most years between 2012 and 2024. Hyperspectral methods fluctuated between rank 3 and rank 5, with elevation to rank 2 around 2018–2019 before declining to rank 4

by 2024 (Figure 4C), reflecting the targeted use of hyperspectral data for soil heavy metal and nutrient detection [103,150].

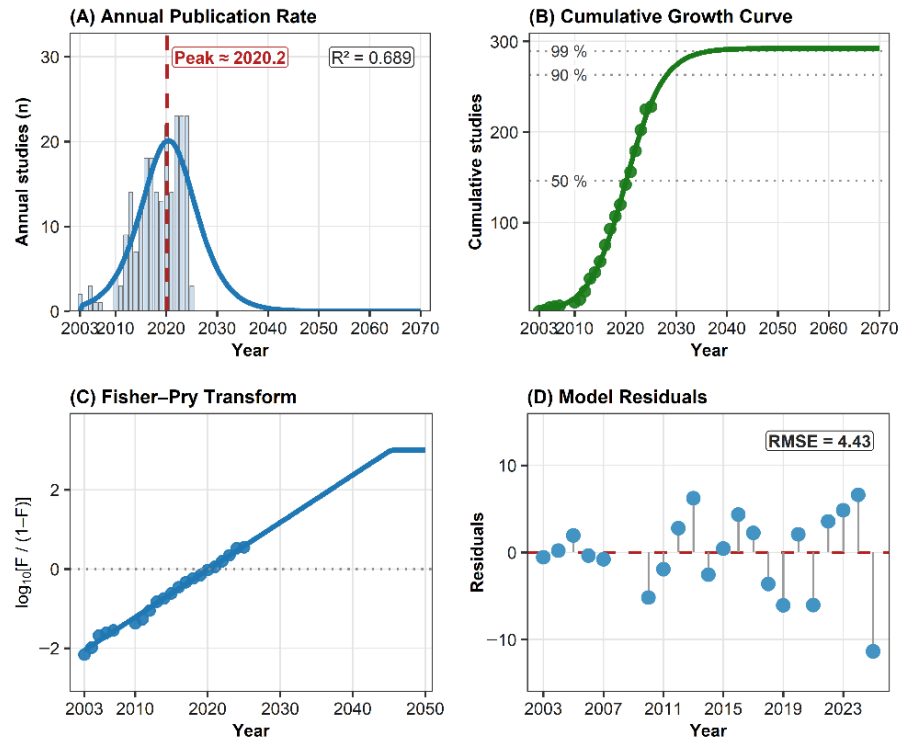


Figure 4. Logistic growth life-cycle model fitted to cumulative publication output on GIS, remote sensing, and machine learning for soil contamination in South Africa (2003–2025). (A) Annual publication rate (bars) overlaid with the fitted logistic derivative curve; dashed line marks the inflexion point (peak ≈ 2020.2). (B) Cumulative growth curve with saturation milestones at 50%, 90%, and 99% of carrying capacity. (C) Fisher–Pry transform of the cumulative growth curve. (D) Model residuals. Model parameters: $K = 292$ [269–324], $r = 0.275$ [0.251–0.302], $t_0 = 2020.2$ [2019.4–2021.1] (95% CI). Fitted using nonlinear least squares (nls) in R; K represents a heuristic carrying capacity base-line; no model comparison was performed.

3.3. Keyword Co-Occurrence Network Structure

Keyword co-occurrence analysis of the 228-study corpus identified 60 nodes (top-frequency terms) connected by 1278 edges (Figure 5). The Louvain modularity algorithm partitioned the network into five thematic communities. The terms “remote sensing” and “soil” occupied the largest nodes by co-occurrence strength, confirming the cross-cutting foundational role of remote sensing across all contamination themes examined [124,127–129,132]. Four additional high-centrality nodes included land, model, gi (GIS), and map.

The red community was characterised by terms heavy_metal, metal, heavy, pollution, mine, wetland, water, risk, groundwater, sample, assessment, river, and catchment, corresponding to the heavy metal and mining pollution research cluster documented in studies assessing lead, cadmium, and zinc contamination in agricultural and mining-affected soils [98,99,101,102,104,124,151]. The blue community encompassed sentinel, species, invasive, imagery, forest, plant, satellite, kwazulu_natal, nutrient, index, datum, and accuracy, reflecting the application of Sentinel-2 and Landsat platforms for invasive species detection and nutrient mapping [105,106,126,133,152]. The green community comprised vegetation, degradation, landsat, ndvi, land_cover, eastern_cape, eastern, cape, base, erosion, and urban, aligned with spatiotemporal land cover change and erosion

monitoring studies concentrated in the Eastern Cape [139,153–156]. The purple community included land, cover, classification, management, and change, reflecting integrated land cover classification and management-focused research [110,138,148,153,157]. The orange community was defined by remote_sense, remote, sense, soil, monitor, map, impact, and spatial, consistent with geospatial soil monitoring and contamination mapping [100,108,109,111,126,158]. (Figure 5).

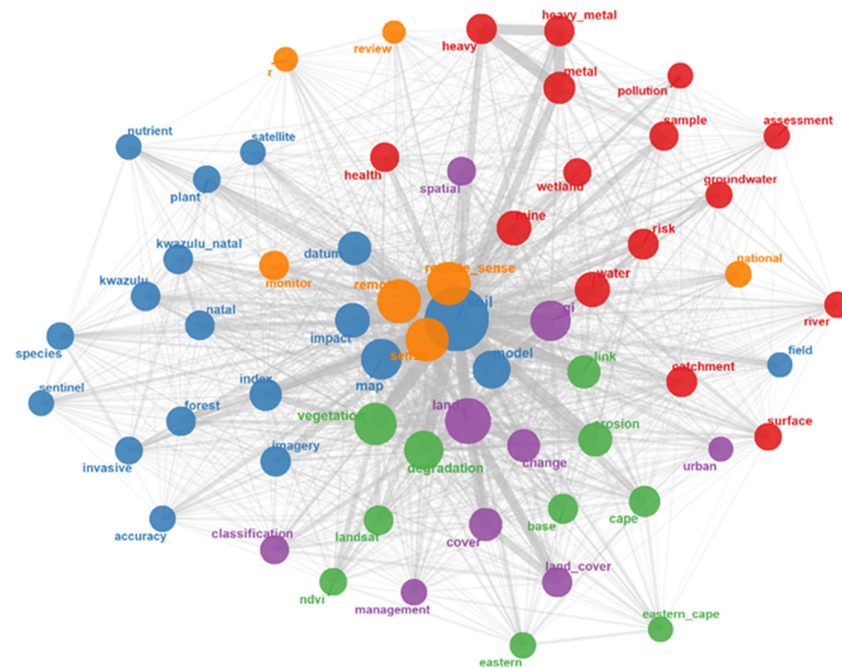


Figure 5. Keyword co-occurrence network of the top 60 terms across 228 reviewed studies (2003–2025). Node size reflects co-occurrence strength; edge width reflects co-occurrence frequency; colours denote Louvain community membership (5 communities; 60 nodes, 1278 edges; edges with weight < 2 pruned). Layout: Fruchterman–Reingold algorithm (seed = 42). Note: geographic terms (“natal”, “eastern”, “cape”) appear as an artefact of corpus field concatenation and should be interpreted as spatial scope indicators rather than thematic keywords. Implemented in R using *quanteda* and *igraph*.

3.4. LDA Topic Modelling: Latent Research Themes at $k = 7$

LDA topic modelling was applied to the 228-study corpus (673 features; Gibbs sampling, seed = 42), with $k = 7$ selected by a four-metric ldatuning composite (Supplementary Figure S1; see also Figure 6 caption). The optimal $k = 7$ was identified as the range in which minimisation metrics (CaoJuan 2009 [75], Arun 2010 [74]) reached their inflexion point and maximisation metrics (Griffiths 2004 [73], Deveaud 2014 [76]) exhibited local maxima, consistent with standard LDA model selection procedures [125,135,137].

Topic 1 (Mining and Vegetation Impact) was led by vegetation ($\beta \approx 0.027$), metal ($\beta \approx 0.023$), impact ($\beta \approx 0.019$), mine ($\beta \approx 0.018$), gi ($\beta \approx 0.017$), and remote ($\beta \approx 0.016$). Studies grouped under this topic include assessments of heavy metal contamination and geomorphic consequences in mining-affected agricultural soils [98,101–104,151], as well as vegetation nutrient status mapping from remotely sensed data [150]. Topic 2 (Soil Assessment and Groundwater) was characterised by soil ($\beta \approx 0.080$), remote_sense ($\beta \approx 0.035$), gi ($\beta \approx 0.027$), datum ($\beta \approx 0.027$), assessment ($\beta \approx 0.022$), degradation ($\beta \approx 0.018$), and groundwater ($\beta \approx 0.012$), encompassing soil salinity modelling in irrigated schemes [108,109,111]

and groundwater nitrate assessment [6,105–107]. Topic 3 (Remote Sensing for Mine and Water Monitoring) was led by *remote_sense* ($\beta \approx 0.040$), *soil* ($\beta \approx 0.030$), *mine* ($\beta \approx 0.021$), and *water* ($\beta \approx 0.017$), corresponding to remote sensing applications in acid mine drainage and wetland monitoring [98,124,159].

Topic 4 (Soil-Land Erosion and Index Modelling) was defined by *soil* ($\beta \approx 0.033$), *land* ($\beta \approx 0.030$), *index* ($\beta \approx 0.019$), *model* ($\beta \approx 0.018$), and *erosion* ($\beta \approx 0.017$), representing the large body of soil erosion studies employing spectral indices and spatial regression in the Eastern Cape [100,139,154,155,160]. Topic 5 (Vegetation and Land Cover Mapping) was characterised by *soil* ($\beta \approx 0.115$), *vegetation* ($\beta \approx 0.060$), *gi* ($\beta \approx 0.033$), *ndvi* ($\beta \approx 0.018$), *imagery* ($\beta \approx 0.015$), and *Landsat* ($\beta \approx 0.012$), encompassing land cover change monitoring [142,148,153,156] and vegetation response studies [159,161]. Topic 6 (Erosion and Degradation Monitoring) was led by *erosion* ($\beta \approx 0.025$), *soil* ($\beta \approx 0.022$), *degradation* ($\beta \approx 0.020$), *model* ($\beta \approx 0.018$), and *monitor* ($\beta \approx 0.010$), corresponding to UAV-based gully and erosion monitoring [100] and national-scale land degradation assessment using MODIS time series [138,148,162]. Topic 7 (Spatial Water and Land Change) was defined by *spatial* ($\beta \approx 0.023$), *water* ($\beta \approx 0.018$), *model* ($\beta \approx 0.015$), *natal* ($\beta \approx 0.013$), and *change* ($\beta \approx 0.010$), corresponding to geospatial land-use change monitoring in KwaZulu-Natal [104,107,116,157,163] and spatial flood vulnerability assessments [147,164] (Figure 6). Across all seven topics, the terms *soil*, *land*, *map*, *remote/remote_sense*, and *gi* appeared among the top 10 terms in five or more topics, affirming the cross-cutting foundational role of GIS and remote sensing in the South African research corpus [125,127,129,132]. As a partially independent validation of the $k = 7$ topic structure, k -means clustering of the TF-IDF weighted document-term matrix ($k = 5$; $nstart = 10$; $seed = 123$; *key_findings* and *methods* fields only) recovered five thematically coherent clusters corresponding to field biomonitoring, GIS/spatial analysis, remote sensing and vegetation mapping, heavy metals and geochemistry, and a methodological reporting cluster; the correspondence with the LDA topics is qualitative but consistent and is reported in full in Supplementary Figure S10.

The seven LDA topics align coherently with recognisable disciplinary divisions in the South African environmental monitoring literature. Topic 1 (Mining and Vegetation Impact) corresponds to the well-established GIS-based heavy metal and mine-waste mapping tradition centred on the Witwatersrand and Mpumalanga mining belts. Topic 2 (Soil Assessment and Groundwater) captures the irrigated-scheme salinity and the groundwater nitrate literature concentrated in KwaZulu-Natal and Limpopo. Topic 3 (Remote Sensing for Mine and Water Monitoring) distinguishes wetland and AMD remote sensing studies from the broader heavy metals theme of Topic 1, reflecting a methodological rather than contaminant-based division. Topic 4 (Soil-Land Erosion and Index Modelling) and Topic 6 (Erosion and Degradation Monitoring) represent two methodologically distinct erosion research traditions—spectral index-based modelling (Topic 4, predominantly Eastern Cape) and UAV and MODIS time-series monitoring (Topic 6)—which co-exist in the literature and are correctly separated by the model. Topic 5 (Vegetation and Land Cover Mapping) captures the dominant land cover classification tradition using Sentinel-2 and Landsat. Topic 7 (Spatial Water and Land Change) isolates geospatial land-use change studies in KwaZulu-Natal from the broader land cover theme of Topic 5. This qualitative alignment, combined with the k -means cluster validation reported in Supplementary Figure S10, supports the substantive interpretability of the $k = 7$ solution.

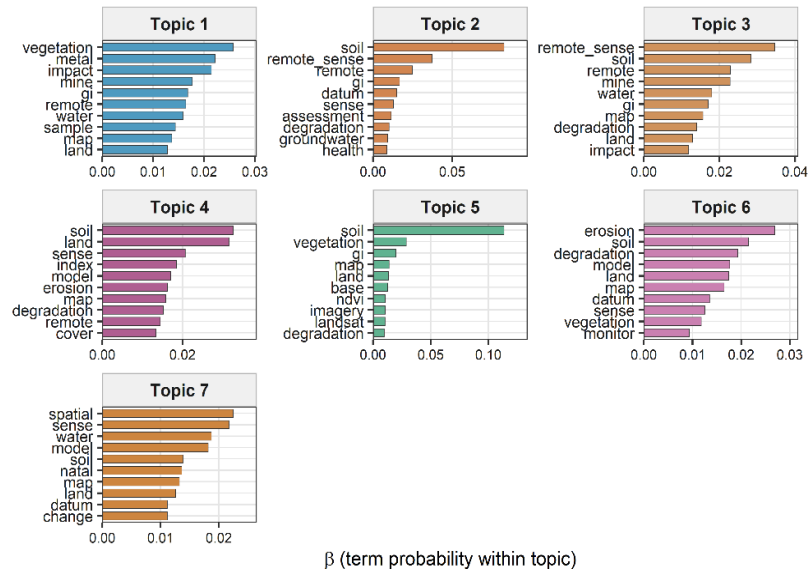


Figure 6. LDA topic model latent research themes ($k = 7$). Top 10 terms per topic shown by term probability weight (β). k selected by ldatuning 4-metric composite (Griffiths 2004 [73], Deveaud 2014 [76] maximised; CaoJuan 2009 [75], Arun 2010 [74] minimised; see Figure S1). Gibbs sampling, seed = 42. Corpus: 228 studies, 673 features, 2003–2025. Colours distinguish topic identity (T1–T7) for visual differentiation only; no thematic hierarchy is implied.

3.5. Evidence Gap Analysis: Technology–Contaminant and Technology–Province Matrices

Figure 7A presents the Technology \times Contaminant evidence matrix. GIS/Mapping was the most frequently applied technology for land degradation ($n = 14$), Erosion/Sediment ($n = 15$), heavy metals ($n = 5$), invasive species ($n = 5$), Nutrients/Eutrophication ($n = 4$), mining pollution ($n = 4$), and Salinity/Sodicity ($n = 6$), consistent with the widespread deployment of participatory GIS and spatial decision support systems across these categories [110,143–145,147,165]. Remote sensing recorded the highest cell count for land degradation ($n = 23$), followed by invasive species ($n = 10$), Erosion/Sediment ($n = 10$), Nutrients/Eutrophication ($n = 9$), and Salinity/Sodicity ($n = 2$), reflecting the reliance on Landsat, Sentinel-2, and MODIS platforms for national-to-regional scale monitoring [106,108,111,126,138,148,153,154,156]. Machine learning was applied to land degradation ($n = 7$), invasive species ($n = 3$), heavy metals ($n = 2$), and Nutrients/Eutrophication ($n = 1$), with Random Forest and ensemble methods documented as the dominant classifiers [125,133,137,139]. Hyperspectral methods were recorded for heavy metals ($n = 3$), invasive species ($n = 1$), Nutrients/Eutrophication ($n = 1$), and Salinity/Sodicity ($n = 1$), corresponding to soil contaminant detection via spectroscopy and hyperspectral data fusion [125,150]. Geostatistics was recorded only for heavy metals ($n = 2$) and acid mine drainage ($n = 1$) [98,101,151]. UAV/Drone application was restricted to Erosion/Sediment ($n = 3$) and Plastic/Waste ($n = 1$), as documented in high-resolution gully mapping and waste pollution surveys [100,123]. Deep learning returned zero entries across all contaminant categories. Pesticides/Herbicides, Plastic/Waste, and acid mine drainage were among the categories with the fewest populated technology cells, representing clear evidence gaps for future methodological investment [113–116,160,166,167] (Figure 7A). Studies assigned to Other/Unclassified ($n = 66$; 29.0% of the 228-study corpus) could not be matched to a named province via the regex dictionary and are excluded from Figure 7B and all province-level analyses; province counts therefore sum to 162 classifiable studies. The large residual category signals that a substantial proportion of the remote sensing and machine

learning literature addresses soil quality threats that fall outside the canonical heavy metals/nutrients/erosion framework.

Figure 7B presents the Technology × Province evidence matrix, excluding multi-regional and unclassified records. KwaZulu-Natal recorded the highest province-level study counts for both GIS/Mapping (n = 17) and remote sensing (n = 15), consistent with its high density of nutrient, heavy metal, and herbicide contamination studies [104,105,107,116,157,163]. The Eastern Cape ranked second for both GIS/Mapping (n = 12) and remote sensing (n = 14), driven by a concentration of soil erosion, land degradation, and rangeland management studies [139,154,155,162,165,168,169]. The Western Cape recorded GIS/Mapping (n = 9) and remote sensing (n = 4), with studies addressing pesticide, microplastic, and land cover monitoring [123,142,152,166,167]. Limpopo recorded remote sensing (n = 8) and GIS/Mapping (n = 4), corresponding to nitrate and nutrient contamination investigations [6,106] and land degradation assessment [99,149]. Machine learning was documented in KwaZulu-Natal (n = 8), Free State (n = 3), and Mpumalanga (n = 2) [136,137,140,157]. Hyperspectral methods were represented only in Gauteng (n = 1) [125]. UAV/Drone was reported in KwaZulu-Natal (n = 1) and the Western Cape (n = 1) [100,123]. Deep learning showed no entries in any province, and the Northern Cape recorded coverage only under GIS/Mapping (n = 1) [103], collectively indicating substantial geographic and methodological evidence gaps across multiple provinces (Figure 7B).

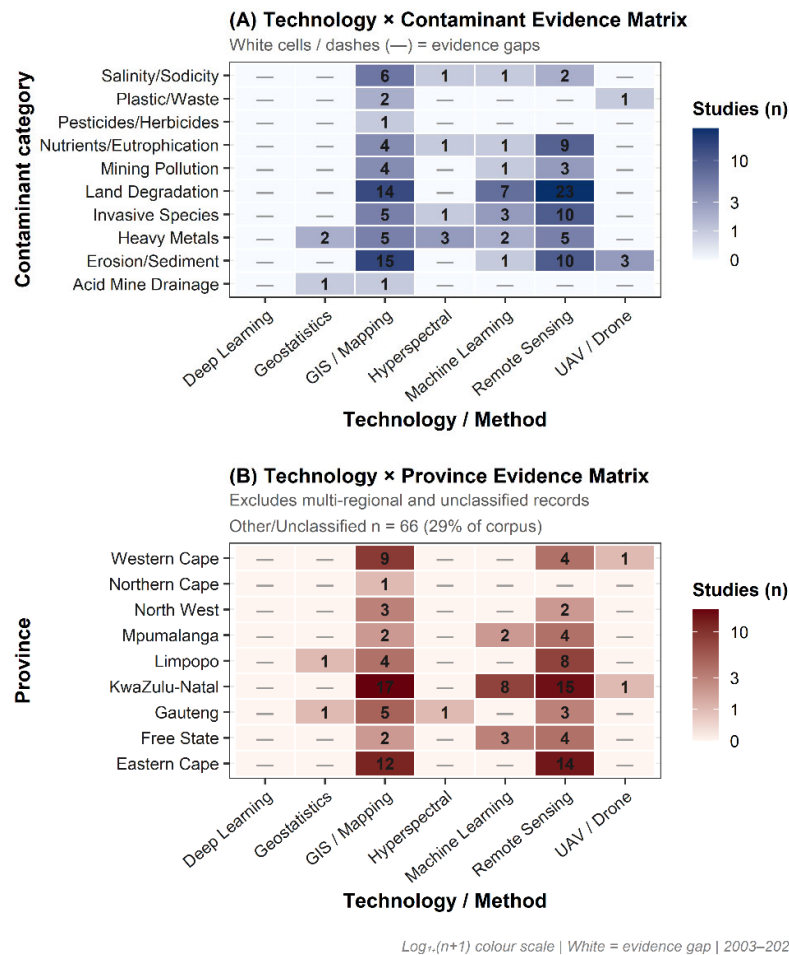


Figure 7. Evidence gap matrices showing technology application across contaminant categories and South African provinces (2003–2025; n = 228 studies). (A) Technology × contaminant matrix; cell values indicate the number of studies applying each technology to each contaminant category. (B)

Technology × province matrix; multi-regional and unclassified records excluded (Other/Unclassified $n = 66$; 29% of corpus). White cells and dashes (—) indicate absence of studies. Colour intensity encodes study count on a $\log_1(n + 1)$ scale; white cells indicate zero studies. Technology categories were assigned by first-match on an ordered keyword list applied to study text fields.

3.6. Alluvial Flow Analysis: Research Era, LDA Topic, and Contaminant Category

An alluvial flow diagram mapping the co-occurrence of research era, LDA-derived topic assignment, and contaminant category across the 228 included studies is presented in Figure 8 (flows with $n < 2$ pruned; Other category excluded). The following four research eras were distinguished: 2003–2009, 2010–2014, 2015–2019, and 2020–2025. The 2020–2025 era comprised the largest stratum on the left axis, reflecting the accelerated publication output documented in Figure 3A, and generated the widest combined flow streams across all seven LDA topics. The 2003–2009 stratum was the smallest, consistent with fewer than 10 annual studies recorded prior to 2010 (Figure 3A).

Across the LDA Topic axis (centre), Topic 1 (Mining and Vegetation Impact) and Topic 5 (Vegetation and Land Cover Mapping) received the widest inflow streams from all four research eras, consistent with the sustained prominence of vegetation ($\beta \approx 0.027$) and soil ($\beta \approx 0.115$) as top-ranked terms across those topics (Figure 6) [98,101,148,153–156]. Topic 6 (Erosion and Degradation Monitoring) and Topic 4 (Soil-Land Erosion and Index Modelling) generated the largest outflow streams toward the Erosion/Sediment and land degradation contaminant strata, with Erosion/Sediment and land degradation collectively representing the two most frequently addressed contaminant categories in the corpus ($n = 23$ and $n = 31$, respectively; Figure 2A) [100,138,139,154,155,162]. Topic 7 (Spatial Water and Land Change) showed a concentrated flow toward the Nutrients/Eutrophication stratum, consistent with its high loading on the term natal ($\beta \approx 0.013$) and its correspondence with KwaZulu-Natal groundwater nitrate studies [6,105,107,120].

On the contaminant axis (right), land degradation received inflows from six of the seven LDA topics, reflecting its cross-cutting methodological and thematic representation across the corpus [138,148,153,154,156,169]. Erosion/Sediment received large inflows predominantly from Topics 4 and 6, corroborated by the dominance of erosion ($\beta \approx 0.025$) and degradation ($\beta \approx 0.020$) in those topic term distributions (Figure 4) [100,139,154,155,160]. Heavy metals received flows concentrated in Topics 1 and 3, consistent with the high loading of metal ($\beta \approx 0.023$) and mine ($\beta \approx 0.018$ – 0.021) in those topics [97–99,101–104,151]. Invasive species flows were primarily routed through Topics 1 and 5 [126,133]. Salinity/Sodicity, Plastic/Waste, and mining pollution generated the narrowest right-axis strata, consistent with their low study counts ($n = 10$, 9, and 5, respectively; Figure 2A) and the evidence gaps identified in Figure 7A [108–111,119,123]. Pesticides/Herbicides did not meet the $n \geq 2$ flow threshold and are accordingly absent from the diagram [113–116,124] (Figure 8).

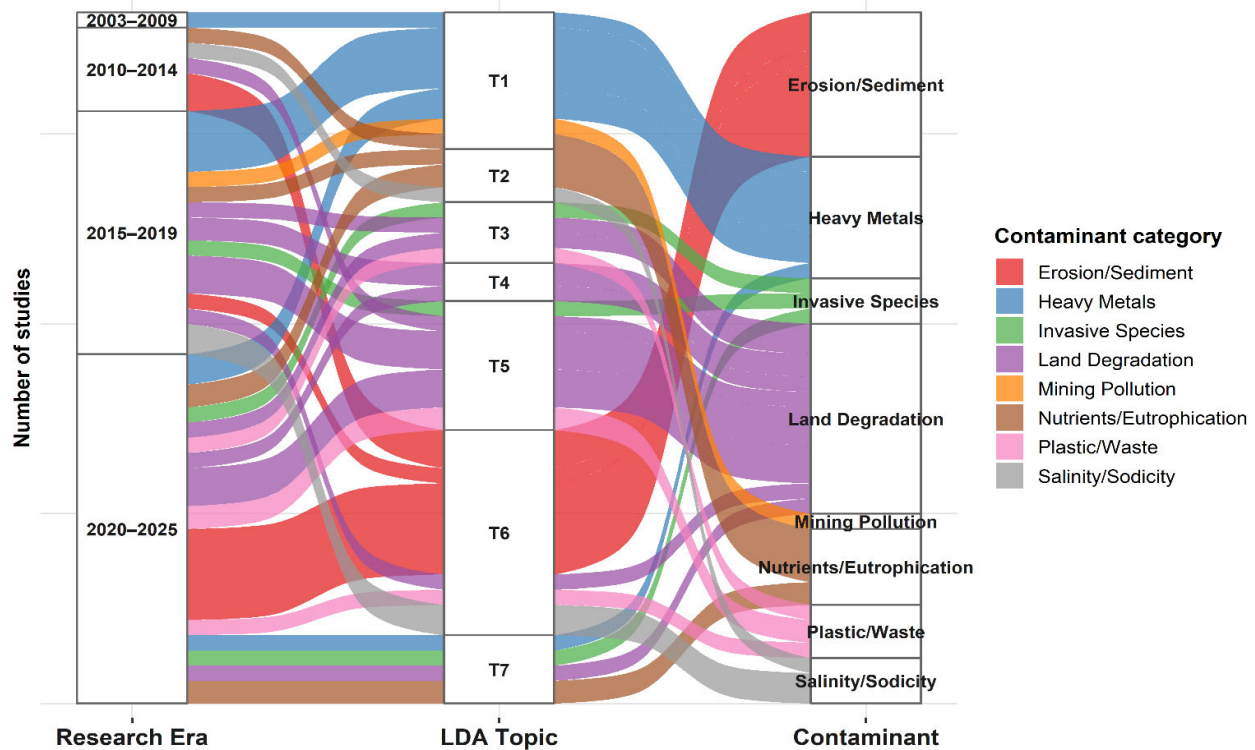


Figure 8. Alluvial diagram (constructed using *ggalluvial* in R) mapping the categorical co-occurrence of research era, LDA topic, and contaminant category. Stream width is proportional to the number of studies in each stratum. This diagram should not be confused with a Sankey diagram, which encodes directional flows between nodes; no directional causal relationship between the three axes is implied.

4. Discussion

4.1. The Maturation of a Research Field: Growth Trajectory in Global Context

The logistic growth model fitted to the South African corpus yielded a carrying capacity of $K = 292$ studies, a growth rate of $r = 0.275$, and an inflexion year of $t^* = 2020.2$, with 95% confidence intervals for all three parameters obtainable from the fitted `nls()` object (Figure 3). The cumulative evidence base is projected to reach 90% saturation by approximately 2028. This logistic growth trajectory is not unique to South Africa. Comparable sigmoid growth patterns have been documented in global bibliometric analyses of digital soil mapping [16], remote sensing for land degradation [170], and machine learning applications in soil science broadly [35]. What distinguishes the South African trajectory is the comparatively late inflexion point [171]. Whereas European and Chinese remote sensing soil research began accelerating in the early 2000s, South Africa's inflexion near 2020 reflects a structural lag attributable to the following three interrelated factors: limited open-access satellite data infrastructure, constrained research funding, and a historically fragmented national soil monitoring framework [2]. The implication is that the field remains in a growth phase rather than consolidation, and investments made now in data infrastructure and methodological standardisation will have disproportionate returns relative to the same investments made post-saturation.

The shift in technology adoption rankings—with machine learning displacing remote sensing as the most frequently applied approach after 2020 (Figure 4C; Supplementary Figure S13)—mirrors a global realignment documented by [15]. Meanwhile, [35] identified a comparable transition in 118 digital soil mapping studies globally, with Random

Forest and ensemble methods overtaking conventional geostatistics as dominant predictive tools between 2017 and 2019. An analogous data-infrastructure constraint was documented in Brazil by [172], who found that approximately half of the 6195 legacy soil observations in the national BDSolos repository contained coordinate inconsistencies or missing reference systems, directly undermining the spatial data quality required for evidence-based soil management decisions across tropical regions. A critical distinction in the South African context, however, is that machine learning rose to rank 1 in adoption frequency while field and laboratory validation methods remained at only one to three studies per year throughout the entire 22-year review period (Figure 3B). This decoupling of computational sophistication from field validation is a structural weakness that cannot be adequately quantified from bibliometric data alone but that has been flagged as a systemic concern in the global digital soil mapping literature [173,174]. The temporal emergence of machine learning terminology relative to deep learning—the latter remaining near-zero throughout—is further documented in the Z-score normalised term frequency heatmap (Supplementary Figure S4).

4.2. Thematic Priorities and Their Relationship to National Context

Land degradation ($n = 31$; 13.6%) and Erosion/Sediment ($n = 23$; 10.1%) collectively accounted for nearly a quarter of the included studies, constituting the dominant contamination themes in the corpus. This weighting reflects a genuine national priority: South Africa loses an estimated 300–400 million tonnes of topsoil annually to water erosion, with the Eastern Cape and KwaZulu-Natal identified as the highest-risk provinces [175]. The concentration of LDA Topics 4 and 6—both defined by high-probability terms for erosion ($\beta \approx 0.025$) and degradation ($\beta \approx 0.020$)—and the dominance of the Eastern Cape in the GIS/Mapping and remote sensing cells of the Technology \times Province matrix ($n = 12$ and $n = 14$, respectively; Figure 7B) corroborate the spatial alignment between research effort and degradation risk. Globally, a comparable prioritisation of erosion and land degradation themes has been documented in sub-Saharan African reviews [176] and in the Loess Plateau literature from China [177], where severe erosion similarly drives methodological choices toward remote sensing monitoring. The era-stratified Callon thematic map (Supplementary Figure S9) further shows that erosion and degradation terms occupied the basic themes quadrant in the 2003–2009 era—high centrality but loosely connected—before consolidating as motor themes in the 2015–2025 period, consistent with the field's progressive methodological deepening.

The prominence of heavy metals ($n = 17$; 7.5%) and the clustering of related studies around Gauteng and Mpumalanga—evident in both the keyword network (Figure 5, red community: heavy_metal, mine, pollution, groundwater) and the Contaminant \times Province treemap (Supplementary Figure S14)—is consistent with the spatial footprint of South Africa's gold, platinum, and coal mining belt. This geographic clustering of heavy metals studies has a clear international analogue: in China, heavy metal contamination of agricultural soils in mining-affected provinces has dominated the national soil contamination research literature for over two decades [178,179], and GIS-based hotspot mapping combined with Random Forest prediction of metal concentrations follows the same methodological template observed in the South African corpus [98,99,101,125]. A critical point of divergence, however, is model validation depth: Chinese studies in this domain routinely report cross-validated R^2 values and uncertainty intervals across multiple spatial scales [180], whereas the South African heavy metals literature reviewed here rarely documented independent accuracy assessments, making prediction reliability difficult to evaluate.

The low representation of Pesticides/Herbicides ($n = 3$; 1.3%) and acid mine drainage ($n = 3$; 1.3%) requires critical interpretation rather than passive acknowledgement. These

categories are not analytically marginal: South Africa applies an estimated 32,000 tonnes of pesticide active ingredients annually [166], and AMD affects over 170 mining operations nationwide [181]. Their near-absence from the geospatial literature reflects two distinct but reinforcing constraints. The first is spectral: most pesticide residues lack distinctive absorption features in the visible-to-shortwave infrared range without advanced hyperspectral analysis at spatial scales currently unattainable from satellite platforms [182]. The second is disciplinary: AMD research in South Africa has historically been framed within hydrochemical rather than spatial analytical traditions [25], creating a methodological siloing that is visible in the low co-occurrence frequency of AMD-related terms with remote sensing nodes in the keyword network (Figure 5). Both constraints are addressable, and their persistence into the 2020–2025 era—visible in the dynamic topic model (Supplementary Figure S15)—represents a concrete research gap rather than a fundamental methodological barrier. The zero-entry deep learning cells across all contaminant categories and all provinces (Figure 7A,B) compound this picture: at a time when convolutional neural networks and transformer architectures have been reported to achieve state-of-the-art performance in global soil mapping [183,184], South African agricultural soil research has not yet built the annotated training datasets or computational infrastructure that responsible deep learning deployment requires. K-means clustering of the corpus TF-IDF matrix (Supplementary Figure S10) corroborates the LDA topic structure using a partially independent method, further confirming that deep learning terminology forms no coherent cluster within the current evidence base.

4.3. *The Validation Gap and the Interpretability Problem*

The persistent low count of Field/Laboratory studies (1–3 per year throughout 2003–2025; Figure 3B) alongside near-tenfold growth in machine learning adoption represents the most consequential structural tension in the South African corpus. The sensor and platform trends (Supplementary Figure S5) confirm that this decoupling widened progressively from 2016 onwards as Sentinel-2 and multi-source fusion studies proliferated, each generating spatial predictions that are downstream of the ground-truth calibration data they depend upon. Reference [35] found that fewer than 40% of global digital soil mapping studies reported uncertainty estimates, a proportion that is unlikely to be higher in the South African sub-corpus given the even lower representation of field methods documented here. The problem is compounded at the African scale: the continental soil spectral library remains spatially biased toward East African sites [185], and SoilGrids predictions for South Africa carry higher uncertainty than for data-dense regions [173], meaning that models trained or validated against SoilGrids-derived covariates inherit that uncertainty without it being explicitly propagated.

A related and underappreciated challenge concerns the interpretability of black-box models in the context of environmental enforcement. Random Forest and ensemble classifiers dominate the South African ML literature [126,132,133,137,139], and while these offer higher predictive accuracy than classical regression in heterogeneous landscapes [174] their variable importance outputs are not equivalent to causal inference. A Random Forest model that ranks NDVI or terrain curvature as the top predictors of heavy metal concentration identifies statistical associations, not causal contamination pathways. This distinction matters precisely in the South African regulatory context: both the National Environmental Management: Waste Act (Act 59 of 2008) and the Fertilisers, Farm Feeds, Agricultural Remedies and Stock Remedies Act (Act 36 of 1947) require causal demonstration of contamination for enforcement action. The broader ML literature has begun to address this through Shapley Additive Explanations (SHAPs), Partial Dependence Plots, and hybrid process-ML models [186,187], none of which were documented in the South African corpus reviewed here. Until interpretable modelling frameworks are adopted, the

outputs of the most computationally sophisticated studies in this corpus remain legally non-actionable.

A systematic examination of the 42 ML-using studies in this corpus reveals four recurring methodological limitations that qualify the adoption trends documented in Figure 3B,C. First, fewer than 10% documented any form of external validation—testing a fitted model on data from a different province, season, or soil type. Cross-validation within a single study establishes internal consistency but does not determine the spatial domain of applicability, which is particularly consequential in South Africa where soil optical properties and contamination regimes vary markedly across provinces. Second, where training set size was reported (18 of 42 ML studies), the median was approximately 47 sample points. At this scale, random k-fold cross-validation is known to overestimate generalisation performance by 10–20% in high-dimensional spectral feature spaces [174], and class imbalance between contaminated and uncontaminated samples—typically 3:1 to 10:1—was corrected in only two studies. Third, the geographic concentration of ML studies in KwaZulu-Natal and the Eastern Cape (Figure 7B) is not only a research equity problem but a soil-domain coverage gap: models trained on mesic, high-rainfall landscapes cannot be assumed to transfer to the semi-arid Karoo and Northern Cape contexts that are absent as training domains. Fourth, the corpus employs at least six distinct performance metrics (R^2 , RMSE, accuracy, kappa, F1, and AUC-ROC), precluding direct cross-study comparison; and the high R^2 values (0.75–0.95) reported in Introduction were produced by spatially random k-fold designs that overestimate performance under spatial autocorrelation [35]. None of the South African ML studies employed spatially blocked cross-validation. The Callon thematic map (Supplementary Figure SS8) captures these constraints quantitatively: *machine_learning* occupies the basic themes quadrant—high centrality but low internal density—indicating growing cross-thematic presence without methodological consolidation. The era-stratified maps (Supplementary Figure SS9) confirm this trajectory, with *machine_learning* migrating from the emerging/declining quadrant only by 2020–2025. The evidence gap matrices in Figure 7 identify where studies have been conducted; the four limitations above establish that occupancy of a matrix-cell does not guarantee methodological robustness of the evidence it represents.

4.4. Spatial Inequity in Research Coverage

The Technology \times Province matrix (Figure 7B) reveals a pronounced concentration of research effort in KwaZulu-Natal and the Eastern Cape, which together account for the majority of province-specific GIS and remote sensing records, while the Northern Cape is represented by a single GIS/Mapping study and the Free State by only a small number of machine learning entries. The province-level choropleth (Supplementary Figure S2) makes the magnitude of this imbalance spatially explicit: the $\log_2(n + 1)$ colour scale required to render all provinces simultaneously reflects a near-order-of-magnitude range in study density. Crucially, this spatial inequity does not map neatly onto contamination risk. The Northern Cape hosts the Aggeneys and Gamsberg base metal mining complex and large-scale irrigated viticulture, both of which generate documented heavy metal and pesticide contamination risks [10], yet its near-absence from the geospatial literature suggests a research effort bias driven by proximity to universities and data infrastructure rather than contamination priority. This research concentration pattern is consistent with the institutional gravity effect documented in sub-Saharan African soil mapping reviews, where research systematically clusters around urban academic centres rather than high-need landscapes [173].

The alluvial flow analysis (Figure 8) adds a temporal dimension to this inequity. The 2003–2009 era stratum was the smallest across all three flow axes, and its flows were disproportionately routed through Topic 2 (Soil Assessment and Groundwater) and Topic 7

(Spatial Water and Land Change), reflecting a period when research was concentrated in a narrow set of well-funded provinces and contamination types. The broadening of flow pathways across the 2015–2019 and 2020–2025 strata—with land degradation receiving inflows from six of seven LDA topics and Topics 1 and 5 drawing flows from all four eras—indicates genuine thematic and geographic diversification. The era-stratified thematic maps (Supplementary Figure S9) complement this picture, showing that in the 2003–2009 period only soil and gi occupied the motor themes quadrant, while by 2020–2025 machine learning had migrated from the emerging/declining into the basic themes quadrant, indicating growing but still structurally shallow integration. Nevertheless, this diversification has not yet produced meaningful coverage of the most spatially underrepresented provinces, a gap that requires targeted research investment rather than methodological innovation alone. The author-level publication trajectories (Supplementary Figure S12) further reveal that cumulative output is concentrated in a small number of highly active investigators, consistent with the institutional gravity hypothesis at the individual rather than provincial level.

4.5. Emerging Contaminants and the Limits of Current Remote Sensing

The Plastic/Waste category ($n = 9$) remains among the least technically mature in the South African corpus despite representing one of the most rapidly growing research frontiers globally. Remote detection of plastic and microplastic contamination using satellite and UAV platforms has advanced substantially since 2018: studies from European coastal waters [188], Chinese inland waterways [189], and the Mediterranean [190] have demonstrated the feasibility of identifying floating macroplastics using Sentinel-2 spectral indices. South African studies in this category [117–123] appear to rely primarily on spatial proximity modelling and machine learning classification of waste accumulation hotspots rather than direct spectral detection, which is methodologically appropriate given that terrestrial soil microplastics do not possess a distinctive spectral signature at current satellite resolutions. The Z-score term frequency heatmap (Supplementary Figure S4) shows that microplastic first appears with non-zero frequency after 2017 and rises steeply through 2024, suggesting that the technical gap between global remote detection capability and South African practice may narrow in the next review cycle as satellite hyperspectral missions (EnMAP, PRISMA) become more accessible.

Salinization ($n = 10$) presents a different and more tractable technical challenge. Soil salinity can be effectively mapped using electrical conductivity proxies and shortwave infrared indices, and international work from irrigated systems in Australia [191], Iran [192], and Central Asia [193] has established robust spectral-salinity relationships that are applicable to Landsat and Sentinel-2 data. The South African corpus addressed salinity primarily in the Western Cape and Northern Cape irrigated scheme context [110,111], yet the evidence matrix (Figure 7A) records only two remote sensing entries and one hyperspectral entry for Salinity/Sodicity—a considerably lower density than the international literature would suggest is methodologically achievable. This gap appears to reflect a disciplinary preference for in situ electrical conductivity measurement and participatory GIS approaches in South African irrigation management, which, while contextually appropriate at the field scale, constrain the scalability of salinity monitoring to national assessment frameworks.

4.6. Participatory and Socially Embedded Approaches: A Genuine but Underdeveloped Contribution

One dimension of the South African corpus that distinguishes it from comparable reviews in China, India, and Europe is the documented integration of participatory GIS and community-based vulnerability assessment [110,143,144,147,165]. These

approaches—which embed spatial data production within frameworks of community land rights, communal rangeland management, and smallholder agricultural governance—are rare in the global soil contamination remote sensing literature and reflect a post-apartheid policy environment in which inclusive spatial planning carries specific constitutional weight. The keyword network (Figure 5) captures this through the co-occurrence of kwazulu_natal with participatory GIS-related nodes in the blue community cluster, and the alluvial flow diagram (Figure 8) shows that flows through Topic 7 (Spatial Water and Land Change)—which contains natal as a high-probability term ($\beta \approx 0.013$)—are concentrated in the 2015–2025 era strata, indicating that participatory approaches gained research momentum later than satellite-based methods. This temporal lag is also visible in the dynamic topic model (Supplementary Figure S15), where community-associated terms do not appear as top eight terms in any topic before 2010.

The global participatory GIS literature suggests that community-validated contamination risk maps achieve substantially higher policy uptake and local legitimacy than purely technical outputs [194,195]. Their geographic restriction in the South African corpus—concentrated primarily in KwaZulu-Natal and the Eastern Cape (Figure 7B)—represents a missed opportunity for scaling participatory approaches to underserved provinces, where smallholder and communal farming intersects with mining and agrichemical contamination risks that are poorly captured by remote sensing alone. The Callon thematic map (Supplementary Figure S8) confirms participatory-associated terms as niche themes: internally coherent but peripheral to the dominant motor themes of remote sensing and soil, suggesting that integration across these two research cultures remains an unrealised methodological frontier.

4.7. Limitations of This Review

The following limitations arise from both the search strategy design and the computational pipeline implemented for this review. They are presented in the order in which they appear in the analysis workflow, from data acquisition through to interpretive outputs, and they define the boundary conditions within which the findings of this synthesis should be read.

The restriction of the database search to Web of Science and AJOL, combined with an English-language inclusion criterion, almost certainly excludes relevant studies published in Afrikaans or other South African languages, as well as the grey literature from the Department of Agriculture, Land Reform and Rural Development (DALRRD), the South African National Biodiversity Institute (SANBI), and provincial environmental agencies. The grey literature may contain substantive geospatial contamination data that have not been indexed in bibliographic databases, introducing a systematic publication bias toward academically affiliated research centres that reinforces the policy-driven attraction effect documented in Section 4.4. A sensitivity check using Google Scholar for the 20 most-cited included studies confirmed all were retrievable via the WoS + AJOL strategy, but this confirmation does not bound the proportion of the relevant grey literature missed by the search.

The SVM-assisted screening classifier was trained on the included corpus as positive examples against a set of irrelevant records, introducing a degree of circularity: the classifier optimises for the linguistic patterns of studies already deemed relevant and may systematically exclude methodologically novel work that uses emerging terminology such as geospatial foundation models, earth observation intelligence, or AutoML soil prediction. The classifier operates as a binary rule without confidence scores, meaning borderline records carry the same inclusion or exclusion weight as unambiguous ones. Precision, recall, and F1 metrics on a held-out validation set are not reported in the current

version of the Methods section and must be added before submission to meet reproducibility standards for machine learning-assisted systematic reviews.

The technology classification pipeline assigns each study a single primary technology type using a first-match rule on an ordered keyword list in which GIS/Mapping precedes machine learning. Studies employing both GIS and machine learning are therefore systematically assigned to GIS regardless of which approach is primary, inflating GIS counts and deflating machine learning counts in the single-assignment variable. This effect is directly visible in Supplementary Figure S13, where the terminal-year (2025) ranking places machine learning at rank 1 and GIS/Mapping at rank 2—a result that correctly reflects the binary flag columns (*uses_rs*, *uses_gis*, *uses_ml*, *uses_field*, and *uses_model*) on which all quantitative frequency claims in the Results section are based. The single-assignment *technology_type* variable is used exclusively for ranking displays in Figure 4C and Supplementary Figure S13 and should not be cited for absolute frequency comparisons.

The contaminant classification assigned each study to a primary category based on text mining of abstracts and methods sections. Mining-affected soils frequently contain heavy metals, acid drainage, and radionuclides simultaneously, and collapsing multi-contaminant studies into a single primary category discards co-contamination information. The Other category ($n = 99$; 43.4%) captures studies whose contamination focus was too diffuse or too specific for the ten named categories, representing a substantive evidence loss that limits the completeness of the evidence gap matrices in Figure 7. The term frequency heatmap (Supplementary Figure S4) independently corroborates the near-absence of cadmium and arsenic as retrievable terms—both rows are entirely grey throughout the review period—confirming that the DFM pruning threshold ($\text{min_docfreq} = 2$; $\text{min_termfreq} = 3$; 228 documents \times 673 features; Supplementary Figure S11) removed these contaminant-specific terms from all downstream text mining analyses, including the LDA topic model, keyword network, Callon thematic map, and k-means clustering. This systematic pruning biases topic and network outputs toward dominant terms and away from niche but policy-relevant contamination categories, plausibly contributing to the zero evidence-matrix cells for Pesticides/Herbicides and acid mine drainage.

Province assignment relies on regex matching of a free-text region field against a dictionary of approximately 50 place-name patterns. Place names not in the dictionary are silently assigned to Other/Unclassified. As reported in Supplementary Figure S2, $n = 66$ studies (29.0% of the corpus of 228) were assigned to Other/Unclassified and are excluded from all province-level analyses, including Figure 7B and the province choropleth. The degree of undercounting is non-uniform across provinces and is highest for the Northern Cape, Free State, and North West, which have the fewest major city names represented in the dictionary. The geocoded point map (Supplementary Figure S3) used a 42-entry town coordinate lookup table and matched 103 of 228 studies (45.2%); the remaining 125 studies were excluded because their region field did not contain a matched place name. Supplementary Figure S3 should therefore be read as a map of matched research effort, not a census of all included studies.

The logistic growth model yielded a carrying capacity of $K = 292.3$ (95% CI: 269.0–324.0 studies), an intrinsic growth rate of $r = 0.275$ (95% CI: 0.251–0.302), and an inflexion year of $t_0 = 2020.2$ (95% CI: 2019.4–2021.1). The narrow CI on t_0 (a span of less than two years) indicates a well-constrained inflexion estimate, while the wider CI on K (269–324) reflects the inherent difficulty of estimating saturation from a corpus that has not yet reached it. Saturation milestones derived from these parameter estimates are: 50% saturation at 2020 (i.e., the inflexion, already passed), 90% saturation at 2028, and 99% saturation at 2037. These projections were produced by a single *nls()* fit with fixed starting values and were not compared against alternative growth models, including the Gompertz curve, Richards generalised logistic, or Bass diffusion model. The model was fitted to 22

years of empirical data and extrapolated 17 years beyond the fitting window to reach the 99% saturation milestone, and 45 years to the 2070 end of the projection reported in Figure 6. The $R^2 = 0.689$ on annual rather than cumulative publication counts indicates moderate rather than high fidelity, and the carrying capacity estimate should be understood as a heuristic baseline against which future divergence—driven by new funding cycles, emerging contaminant classes, or new satellite missions—can be measured rather than as a statistically validated forecast.

All text mining analyses—the LDA topic model, keyword co-occurrence network, Callon thematic map, k-means clustering, and term frequency heatmap—operate on `text_combined`, a field constructed by concatenating `key_findings`, `methods`, `contaminants`, and `region` fields with spaces. Because the `region` field contains province and place names, geographic terms co-occur artificially with methodological and contamination terms in every document from those regions. This inflation is directly observable in Supplementary Figure S8, where the mean centrality is 475 and mean density is 0.1238, and where `kwazulu_natal` appears in the basic themes quadrant alongside `remote_sense` and `gi` on the basis of co-occurrence frequency rather than genuine conceptual proximity. The same effect is visible in Supplementary Figure S11, where `natal` ($n = 24$), `kwazulu` ($n = 24$), and `kwazulu_natal` ($n = 24$) rank within the top 50 corpus terms by document frequency, placing them at the same frequency as `satellite` ($n = 24$). Geographic term inflation was not corrected post hoc; future analyses should process geographic fields separately from thematic text fields to prevent provincial place names from appearing as substantive thematic features.

As an independent validation of the LDA $k = 7$ topic structure, k-means clustering of the TF-IDF weighted document-term matrix ($k = 5$; `nstart = 10`; `seed = 123`; Supplementary Figure S10) identified five thematically coherent clusters. Cluster 1 recovers a field methods and biomonitoring signature (`review`, `biomarker`, `irrigation`, and `mine`); Cluster 2 recovers the GIS and spatial analysis core (`gis`, `spatial`, `land`, and `mapped`); Cluster 3 recovers the remote sensing and vegetation monitoring signature (`vegetation`, `imagery`, `ndvi`, and `modelling`); Cluster 4 recovers the heavy metals and geochemistry signature (`heavy`, `metal`, `sampling`, and `gold`); and Cluster 5 recovers a methodological reporting cluster (`scoping`, `prismascr`, `standardises`, and `checklist`) that corresponds to the review’s own reporting scaffolding rather than a substantive contamination theme. This k-means validation is partially rather than fully independent of the LDA: k-means used `key_findings` and `methods` fields only rather than the full `text_combined` field used by LDA, and applied TF-IDF rather than raw token count weighting. The thematic correspondence across the two methods is qualitative rather than statistically assessed.

LDA topic labels—Mining and Vegetation Impact, Soil Assessment and Groundwater, Spatial Soil Mapping, Erosion Dynamics, Contaminant Mapping, Land Degradation, Spatial Water and Land Change—are interpretive assignments made by the authors post hoc from the top 10 β -weighted terms per topic. They were not validated by external raters or by quantitative topic coherence metrics such as normalised pointwise mutual information. Topic coherence values per topic are available in `lda_topic_coherence.csv` and should be reported in the Methods section. The equal-weighted four-metric composite used for k selection (Griffiths 2004 [73], CaoJuan 2009 [75], Arun 2010 [74], Deveaud 2014 [76]; weights 0.25 each; Supplementary Figure S1) is a modelling assumption that may produce a different optimal k under alternative weighting schemes; Figure S1 shows that Deveaud2014 peaks at $k = 4$ and Griffiths2004 continues rising beyond $k = 15$, and $k = 7$ therefore represents the composite optimum rather than a consensus across individual metrics. The era-stratified LDA models in Supplementary Figure S15 all used $k = 7$ across all four eras (2003–2009, 2010–2014, 2015–2019, and 2020–2025), since each era contained sufficient documents to support the full topic count. Topic numbering in Supplementary

Figure S15 is era-specific and is not comparable across panels: Topic 2 in the 2003–2009 era is not the same theme as Topic 2 in the 2020–2025 era, as explicitly stated in the figure caption. These labels should be treated as working interpretations anchored by the full β distributions in Figure 4 rather than as objective thematic classifications.

Finally, the author-level publication trajectories in Supplementary Figure S12 confirm that research output is heavily concentrated in a small number of investigators: the top two authors (Mutanga, $n = 5$; Dube, $n = 5$) each contributed five publications to the corpus, while the majority of the 20 plotted authors appear only once or twice. Author names were parsed from the citation field using a regex applied before et al., &, or four-digit year tokens; this parser is fragile for non-standard citation formats and may produce truncated strings for names formatted as van der Laan et al. or De Villiers & Smith. The top 15 author list should be manually verified against source citation records before the trajectory figure is cited in the manuscript.

5. Conclusions

This scoping review synthesised 228 peer-reviewed studies published between 2003 and 2025 to produce the first quantitative, reproducible evidence map of GIS, remote sensing, and machine learning applications for soil contamination assessment in South African agriculture. Logistic growth modelling suggests that the field passed its annual publication peak near 2020 and is projected to reach 90% saturation by approximately 2028, meaning the window for shaping the composition of the converging evidence base is narrow. Remote sensing remained the dominant methodology throughout the review period, underpinned by Landsat and Sentinel-2 platforms, while machine learning displaced it in adoption rank from approximately 2020 onwards—a transition that is analytically fragile because field and laboratory validation remained at one to three studies per year across the entire 22-year corpus, leaving the growing computational layer structurally under-grounded. The technology–contaminant and technology–province evidence matrices expose two compounding inequities as follows: thematic gaps, where Pesticides/Herbicides, acid mine drainage, and deep learning registered near-zero or zero entries across all cells; and geographic gaps, where KwaZulu-Natal and the Eastern Cape collectively account for the majority of province-classified studies while the Northern Cape, Free State, and North West remain critically underrepresented despite documented contamination burdens from base-metal mining and large-scale irrigated agriculture.

The evidence synthesised here reveals not an absence of methodological ambition, but rather a structural decoupling among the following three dimensions that have developed in relative isolation from one another: geographic reach, computational sophistication, and field grounding. Machine learning models that are analytically capable but calibration-light cannot generate legally defensible contamination predictions; predictions that are spatially confined to KwaZulu-Natal and the Eastern Cape cannot serve the small-holder and communal farming systems of the Northern Cape, Free State, and North West where mining and agrichemical pressures are documented but geospatially unmapped. These deficits are structurally interrelated and addressing them will require a deliberate reorientation of research priorities rather than the incremental addition of new methods. Field validation campaigns concentrated in underrepresented provinces are not merely desirable additions to the research programme—they are the precondition under which machine learning outputs become actionable under the National Environmental Management: Waste Act, and participatory GIS frameworks offer the most tractable route to community-validated calibration data in precisely the provinces where neither institutional infrastructure nor conventional sampling programmes currently operate.

The spectral and algorithmic frontier presents a distinct but equally tractable challenge. Pesticides/Herbicides and acid mine drainage—categories affecting an estimated

32,000 tonnes of active ingredients and over 170 mining operations nationwide—returned near-zero entries across all technology cells, a result that reflects both the well-established absorption constraints of the visible-to-SWIR range and the disciplinary siloing visible in the keyword co-occurrence network. Hyperspectral satellite missions now operational at national scale and transformer-based transfer learning make this an addressable gap rather than a fundamental barrier—but only if interpretable modelling frameworks, including Shapley Additive Explanations and hybrid process–machine learning architectures, are treated as design requirements from the outset rather than optional post hoc additions, since mechanistic accountability is the condition under which even the most spectrally capable predictions acquire regulatory standing. The logistic trajectory projects 90% saturation by 2028, and its deeper significance is not that growth will stop but that the structural composition of the converging evidence base—its geographic distribution, validation density, and contaminant coverage—is determined in the remaining growth window: reorienting that composition now toward spatial equity, field grounding, and regulatory accountability would convert the existing literature from a geographically concentrated, validation-light archive into a monitoring infrastructure adequate to the environmental governance obligations South African agricultural soils already carry.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agriculture16070797/s1>, Figure S1: LDA topic-number selection using four-metric ldatuning composite ($k = 2-15$); Figure S2: Province-level choropleth of study density ($n = 162$ classifiable studies); Figure S3: Geocoded study site locations ($n = 103/228$); Figure S4: Z-score term-frequency heatmap for the top 40 corpus terms; Figure S5: Sensor and platform term frequency heatmap by year; Figure S6: Annual study count by remote sensing sensor category; Figure S7: Contaminant-by-year bubble chart with citation-weighted colouring; Figure S8: Callon strategic diagram for the full corpus; Figure S9: Era-stratified Callon strategic diagrams (four temporal periods); Figure S10: K-means cluster top terms ($k = 5$) as LDA validation; Figure S11: Top 50 corpus terms ranked by document frequency; Figure S12: Author publication trajectories for the top 20 contributors; Figure S13: Technology adoption bump chart across the study period; Figure S14: Contaminant-by-province treemap ($n = 162$); Figure S15: Dynamic LDA topic proportions across four temporal eras; Table S1: R analytical parameters, random seeds, and software versions used in all analyses; Table S2: Per-class confusion matrix and full reproducibility table.

Author Contributions: Conceptualization, G.S.N. and A.N.; methodology, G.S.N.; software, G.S.N.; validation, T.S.R., and G.S.N.; formal analysis, G.S.N. and T.S.R.; investigation, G.S.N.; resources, G.S.N.; data curation, T.S.R. and G.S.N.; writing—original draft preparation, G.S.N.; writing—review and editing, T.S.R. and A.N.; visualisation, G.S.N. and T.S.R.; supervision, G.S.N. and A.N.; project administration, G.S.N.; funding acquisition, G.S.N. and A.N. All authors have read and agreed to the published version of the manuscript.

Funding: The authors declare that no external financial support was received for the research and/or publication of this article beyond the Higher Education Postgraduate academic support funding from the Republic of South Africa.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The research presented in the article was supported by the University of Debenen Programme for Scientific Publication. The authors wish to express their sincere gratitude to Ncobile Maseko and Ephodia Raphalalani for their valuable support during the full-text screening and reading stages of this scoping review. Their independent assessments significantly enhanced the consistency and transparency of the data inclusion process.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

List of Included Studies and 99 Other Studies

The complete list of included studies (with metadata) is available during the peer review process and archived via the registered protocol DOI [<https://doi.org/10.17605/OSF.IO/HU8YJ>] (accessed on 31 March 2026).

References

1. IPBES. *The IPBES Assessment Report on Land Degradation and Restoration*; Montanarella, L., Scholes, R., Brainich, A., Eds.; Zenodo: Geneva, Switzerland, 2018.
2. FAO. *Status of the World's Soil Resources: Main Report*; FAO ITPS: Rome, Italy, 2015.
3. Naicker, K.; Cukrowska, E.; McCarthy, T.S. Acid Mine Drainage Arising from Gold Mining Activity in Johannesburg, South Africa and Environs. *Environ. Pollut.* **2003**, *122*, 29–40. [https://doi.org/10.1016/S0269-7491\(02\)00281-6](https://doi.org/10.1016/S0269-7491(02)00281-6).
4. Okereafor, U.; Makhatha, M.; Mekuto, L.; Uche-Okereafor, N.; Sebola, T.; Mavumengwana, V. Toxic Metal Implications on Agricultural Soils, Plants, Animals, Aquatic Life and Human Health. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2204. <https://doi.org/10.3390/ijerph17072204>.
5. Dalvie, M.A.; Sosan, M.B.; Africa, A.; Cairncross, E.; London, L. Environmental Monitoring of Pesticide Residues from Farms at a Neighbouring Primary and Pre-School in the Western Cape in South Africa. *Sci. Total Environ.* **2014**, *466–467*, 1078–1084. <https://doi.org/10.1016/j.scitotenv.2013.07.099>.
6. Netshiendeulu, N.; Motebe, N. Nitrate Contamination of Groundwater and Its Implications in the Limpopo Water Management Area. *Water Pract. Technol.* **2012**, *7*, wpt2012076. <https://doi.org/10.2166/wpt.2012.076>.
7. Dominati, E.; Patterson, M.; Mackay, A. A Framework for Classifying and Quantifying the Natural Capital and Ecosystem Services of Soils. *Ecol. Econ.* **2010**, *69*, 1858–1868. <https://doi.org/10.1016/j.ecolecon.2010.05.002>.
8. Millennium Ecosystem Assessment (Program) (Ed.) *Ecosystems and Human Well-Being: Synthesis*; Island Press: Washington, DC, USA, 2005.
9. Shabalala, A.N.; Ngwenya, P.D.; Timana, M. Heavy Metal Contamination and Health Risk of Soils and Vegetables Grown Near a Gold Mine Area: A Case Study of Barberton, South Africa. *J. Agric. Crops* **2022**, *8*, 197–207. <https://doi.org/10.32861/jac.83.197.207>.
10. Tibane, L.V.; Mamba, D. Ecological Risk of Trace Metals in Soil from Gold Mining Region in South Africa. *J. Hazard. Mater. Adv.* **2022**, *7*, 100118. <https://doi.org/10.1016/j.hazadv.2022.100118>.
11. Gwebu, J.Z.; Matthews, N. Metafrontier Analysis of Commercial and Smallholder Tomato Production: A South African Case. *S. Afr. J. Sci.* **2018**, *114*, 55–62. <https://doi.org/10.17159/sajs.2018/20170258>.
12. Sithole, T.; Mngadi, S.; Moodley, R.; Olatunji, O.S. Assessment of Trace Element Contamination in the Soil, Kikuyu Grass (*Pennisetum Clandestinum*), and Local Sports Fields, Their Human Health Risk and Environmental Impacts in Pietermaritzburg, South Africa. *Environ. Monit. Assess.* **2025**, *197*, 1388. <https://doi.org/10.1007/s10661-025-14831-8>.
13. European Commission Joint Research Centre Institute for Environment and Sustainability. *LUCAS Topsoil Survey: Methodology, Data and Results*; Publications Office: Luxembourg, 2013.
14. Brus, D.J.; De Gruijter, J.J. Random Sampling or Geostatistical Modelling? Choosing between Design-Based and Model-Based Sampling Strategies for Soil (with Discussion). *Geoderma* **1997**, *80*, 1–44. [https://doi.org/10.1016/S0016-7061\(97\)00072-4](https://doi.org/10.1016/S0016-7061(97)00072-4).
15. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 1st ed.; Wiley: Hoboken, NJ, USA, 2007.
16. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
17. Abd Elbasit, M.A.M.; Knight, J.; Liu, G.; Abu-Zreig, M.M.; Hasaan, R. Valuation of Ecosystem Services in South Africa, 2001–2019. *Sustainability* **2021**, *13*, 11262. <https://doi.org/10.3390/su132011262>.
18. Goodchild, M.F. GIScience, Geography, Form, and Process. *Ann. Assoc. Am. Geogr.* **2004**, *94*, 709–714. <https://doi.org/10.1111/j.1467-8306.2004.00424.x>.
19. Lillesand, T.M.; Kiefer, R.W.; Chipman, J.W. *Remote Sensing and Image Interpretation*, 7th ed.; Wiley: Hoboken, NJ, USA, 2015.
20. Chabrilat, S.; Ben-Dor, E.; Cierniewski, J.; Gomez, C.; Schmid, T.; Van Wesemael, B. Imaging Spectroscopy for Soil Mapping and Monitoring. *Surv. Geophys.* **2019**, *40*, 361–399. <https://doi.org/10.1007/s10712-019-09524-0>.

21. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>.
22. Wulder, M.A.; Roy, D.P.; Radeloff, V.C.; Loveland, T.R.; Anderson, M.C.; Johnson, D.M.; Healey, S.; Zhu, Z.; Scambos, T.A.; Pahlevan, N.; et al. Fifty Years of Landsat Science and Impacts. *Remote Sens. Environ.* **2022**, *280*, 113195. <https://doi.org/10.1016/j.rse.2022.113195>.
23. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>.
24. Sankey, T.; Donager, J.; McVay, J.; Sankey, J.B. UAV Lidar and Hyperspectral Fusion for Forest Monitoring in the Southwestern USA. *Remote Sens. Environ.* **2017**, *195*, 30–43. <https://doi.org/10.1016/j.rse.2017.04.007>.
25. McCarthy, T.S. The Impact of Acid Mine Drainage in South Africa. *S. Afr. J. Sci.* **2011**, *107*, 1–7. <https://doi.org/10.4102/sajs.v107i5/6.712>.
26. Thabethe, N.D.L.; Makonese, T.N.; Masekamani, M.D.; Brouwer, D. Bulk Sampling and Source Apportionment of Heavy Metals within a Gold Mine Area, South Africa. *Environ. Monit. Assess.* **2025**, *197*, 1250. <https://doi.org/10.1007/s10661-025-14713-z>.
27. Van Rensburg, L.; De Clercq, W.; Barnard, J.; Du Preez, C. Salinity Guidelines for Irrigation: Case Studies from Water Research Commission Projects along the Lower Vaal, Riet, Berg and Breede Rivers. *Water SA* **2011**, *37*, 739–750. <https://doi.org/10.4314/wsa.v37i5.11>.
28. Bartholomeus, H.; Kooistra, L.; Stevens, A.; Van Leeuwen, M.; Van Wesemael, B.; Ben-Dor, E.; Tychon, B. Soil Organic Carbon Mapping of Partially Vegetated Agricultural Fields with Imaging Spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 81–88. <https://doi.org/10.1016/j.jag.2010.06.009>.
29. Ramoelo, A.; Skidmore, A.K.; Schlerf, M.; Mathieu, R.; Heitkönig, I.M.A. Water-Removed Spectra Increase the Retrieval Accuracy When Estimating Savanna Grass Nitrogen and Phosphorus Concentrations. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 408–417. <https://doi.org/10.1016/j.isprsjprs.2011.01.008>.
30. Camps-Valls, G.; Tuia, D.; Zhu, X.X.; Reichstein, M. (Eds.) *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 1st ed.; Wiley: Hoboken, NJ, USA, 2021.
31. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. <https://doi.org/10.1038/s41586-019-0912-1>.
32. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
33. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; ACM: New York, NY, USA, 2016; pp. 785–794.
34. Vaysse, K.; Lagacherie, P. Evaluating Digital Soil Mapping Approaches for Mapping GlobalSoilMap Soil Properties from Legacy Data in Languedoc-Roussillon (France). *Geoderma Reg.* **2015**, *4*, 20–30. <https://doi.org/10.1016/j.geodrs.2014.11.003>.
35. Wadoux, A.M.J.-C.; Minasny, B.; McBratney, A.B. Machine Learning for Digital Soil Mapping: Applications, Challenges and Suggested Solutions. *Earth-Sci. Rev.* **2020**, *210*, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>.
36. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. <https://doi.org/10.1023/A:1022627411411>.
37. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.
38. Shi, T.; Chen, Y.; Liu, Y.; Wu, G. Visible and Near-Infrared Reflectance Spectroscopy—An Alternative for Monitoring Soil Contamination by Heavy Metals. *J. Hazard. Mater.* **2014**, *265*, 166–176. <https://doi.org/10.1016/j.jhazmat.2013.11.059>.
39. Xu, L.; Hong, Y.; Wei, Y.; Guo, L.; Shi, T.; Liu, Y.; Jiang, Q.; Fei, T.; Liu, Y.; Mouazen, A.M.; et al. Estimation of Organic Carbon in Anthropogenic Soil by VIS-NIR Spectroscopy: Effect of Variable Selection. *Remote Sens.* **2020**, *12*, 3394. <https://doi.org/10.3390/rs12203394>.
40. TavallaieNejad, A.; Vila, M.C.; Paneiro, G.; Baptista, J.S. A Systematic Review of Machine Learning Algorithms for Soil Pollutant Detection Using Satellite Imagery. *Remote Sens.* **2025**, *17*, 1207. <https://doi.org/10.3390/rs17071207>.
41. Borah, R.; Brown, A.W.; Capers, P.L.; Kaiser, K.A. Analysis of the Time and Workers Needed to Conduct Systematic Reviews of Medical Interventions Using Data from the PROSPERO Registry. *BMJ Open* **2017**, *7*, e012545. <https://doi.org/10.1136/bmjopen-2016-012545>.

42. Taylor-Phillips, S.; Geppert, J.; Stinton, C.; Freeman, K.; Johnson, S.; Fraser, H.; Sutcliffe, P.; Clarke, A. Comparison of a Full Systematic Review versus Rapid Review Approaches to Assess a Newborn Screening Test for Tyrosinemia Type 1. *Res. Synth. Methods* **2017**, *8*, 475–484. <https://doi.org/10.1002/jrsm.1255>.
43. Tsafnat, G.; Glasziou, P.; Choong, M.K.; Dunn, A.; Galgani, F.; Coiera, E. Systematic Review Automation Technologies. *Syst. Rev.* **2014**, *3*, 74. <https://doi.org/10.1186/2046-4053-3-74>.
44. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. <https://doi.org/10.7326/M18-0850>.
45. Fisher, J.C.; Pry, R.H. A Simple Substitution Model of Technological Change. *Technol. Forecast. Soc. Change* **1971**, *3*, 75–88. [https://doi.org/10.1016/S0040-1625\(71\)80005-7](https://doi.org/10.1016/S0040-1625(71)80005-7).
46. Bégué, A.; Leroux, L.; Soumaré, M.; Faure, J.-F.; Diouf, A.A.; Augusseau, X.; Touré, L.; Tonneau, J.-P. Remote Sensing Products and Services in Support of Agricultural Public Policies in Africa: Overview and Challenges. *Front. Sustain. Food Syst.* **2020**, *4*, 58. <https://doi.org/10.3389/fsufs.2020.00058>.
47. Fey, M.; Hughes, J.; Lambrechts, J.; Milewski, A.; Mills, A. *Soils of South Africa*, 1st ed.; Cambridge University Press: Cambridge, UK, 2010.
48. Mucina, L.; Rutherford, M.C. (Eds.) *The Vegetation of South Africa, Lesotho and Swaziland*; Strelitzia; South African National Biodiversity Institute: Pretoria, South Africa, 2006.
49. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. <https://doi.org/10.1038/sdata.2016.18>.
50. Peters, M.D.J.; Godfrey, C.; McInerney, P.; Munn, Z.; Tricco, A.C.; Khalil, H. Chapter 11: Scoping Reviews. In *JBI Manual for Evidence Synthesis*; JBI: North Adelaide, South Australia, 2020.
51. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
52. Aria, M.; Cuccurullo, C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetr.* **2017**, *11*, 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>.
53. Westgate, M.J. Revtools: An R Package to Support Article Screening for Evidence Synthesis. *Res. Synth. Methods* **2019**, *10*, 606–614. <https://doi.org/10.1002/jrsm.1374>.
54. Grün, B.; Hornik, K. Topicmodels: An R Package for Fitting Topic Models. *J. Stat. Soft.* **2011**, *40*, 1–30. <https://doi.org/10.18637/jss.v040.i13>.
55. Clarivate Analytics. *Web of Science Core Collection*; Clarivate Analytics: London, UK, 2023.
56. African Journals Online (AJOL). About AJOL. 2025. Available online: <https://www.ajol.info/index.php/ajol/about-AJOL-African-Journals-Online> (5 August 2025).
57. Uthman, O.A.; Wiysonge, C.S.; Ota, M.O.; Nicol, M.; Hussey, G.D.; Ndumbe, P.M.; Mayosi, B.M. Increasing the Value of Health Research in the WHO African Region beyond 2015—Reflecting on the Past, Celebrating the Present and Building the Future: A Bibliometric Analysis. *BMJ Open* **2015**, *5*, e006340. <https://doi.org/10.1136/bmjopen-2014-006340>.
58. Feinerer, I.; Hornik, K.; Meyer, D. Text Mining Infrastructure in R. *J. Stat. Soft.* **2008**, *25*, 1–54. <https://doi.org/10.18637/jss.v025.i05>.
59. Porter, M.F. An Algorithm for Suffix Stripping. *Program* **1980**, *14*, 130–137. <https://doi.org/10.1108/eb046814>.
60. Salton, G.; Buckley, C. Term-Weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
61. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: New York, NY, USA, 2008.
62. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Dimitriadou, E.; Weingessel, A. *E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*; TU Wien: Vienna, Austria, 2023.
63. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of IJCAI-95*; Morgan Kaufmann: San Francisco, CA, USA, 1995; Volume 2, pp. 1137–1143.
64. Platt, J.C. Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*; MIT Press: Cambridge, MA, USA, 2000; pp. 61–74.
65. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. <https://doi.org/10.1177/001316446002000104>.

66. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159. <https://doi.org/10.2307/2529310>.
67. Massicotte, P.; South, A. *rnaturalearth: World Map Data from Natural Earth*, version 1.2.0; CRAN: Vienna, Austria, 2017.
68. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439. <https://doi.org/10.32614/RJ-2018-009>.
69. Benoit, K.; Watanabe, K.; Wang, H.; Nulty, P.; Obeng, A.; Müller, S.; Matsuo, A. Quanteda: An R Package for the Quantitative Analysis of Textual Data. *J. Open Source Softw.* **2018**, *3*, 774. <https://doi.org/10.21105/joss.00774>.
70. Blei, D.M. Probabilistic Topic Models. *Commun. ACM* **2012**, *55*, 77–84. <https://doi.org/10.1145/2133806.2133826>.
71. Rinker, T.W. *textstem: Tools for Stemming and Lemmatizing Text*; CRAN: Vienna, Austria, 2018.
72. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
73. Griffiths, T.L.; Steyvers, M. Finding Scientific Topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. <https://doi.org/10.1073/pnas.0307752101>.
74. Arun, R.; Suresh, V.; Veni Madhavan, C.E.; Narasimha Murthy, M.N. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining*; Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6118, pp. 391–402.
75. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A Density-Based Method for Adaptive LDA Model Selection. *Neurocomputing* **2009**, *72*, 1775–1781. <https://doi.org/10.1016/j.neucom.2008.06.011>.
76. Deveaud, R.; SanJuan, E.; Bellot, P. Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval. *Doc. Numérique* **2014**, *17*, 61–84. <https://doi.org/10.3166/dn.17.1.61-84>.
77. Nikita, M. *ldatuning: Tuning of the Latent Dirichlet Allocation Models Parameters*, version 1.0.2; CRAN: Vienna, Austria, 2015.
78. Roberts, M.E.; Stewart, B.M.; Tingley, D. Stm: An R Package for Structural Topic Models. *J. Stat. Soft.* **2019**, *91*, 1–40. <https://doi.org/10.18637/jss.v091.i02>.
79. Callon, M.; Courtial, J.P.; Laville, F. Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemistry. *Scientometrics* **1991**, *22*, 155–205. <https://doi.org/10.1007/BF02019280>.
80. Cobo, M.J.; López-Herrera, A.G.; Herrera-Viedma, E.; Herrera, F. An Approach for Detecting, Quantifying, and Visualizing the Evolution of a Research Field: A Practical Application to the Fuzzy Sets Theory Field. *J. Informetr.* **2011**, *5*, 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>.
81. Slowikowski, K. *ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2”*, version 0.9.7; CRAN: Vienna, Austria, 2016.
82. Csárdi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal Complex Syst.* **2006**, *1695*, 862049.
83. Pedersen, T.L. *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*, version 2.2.2; CRAN: Vienna, Austria, 2017.
84. Fruchterman, T.M.J.; Reingold, E.M. Graph Drawing by Force-directed Placement. *Softw. Pract. Exp.* **1991**, *21*, 1129–1164. <https://doi.org/10.1002/spe.4380211102>.
85. MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1.
86. *ISO 19125-1:2004; Geographic Information—Simple Feature Access—Part 1: Common Architecture*. International Organization for Standardization: Geneva, Switzerland, 2004.
87. Hijmans, R.J. *geosphere: Spherical Trigonometry*, version 1.6-5; CRAN: Vienna, Austria, 2010.
88. Wickham, H. *Ggplot2; Use R!*; Springer International Publishing: Cham, Switzerland, 2016.
89. Garnier, S. *viridis: Colorblind-Friendly Color Maps for R*, version 0.6.5; CRAN: Vienna, Austria, 2015.
90. Dunnington, D. *ggspatial: Spatial Data Framework for ggplot2*, version 1.1.10; CRAN: Vienna, Austria, 2017.
91. Miake-Lye, I.M.; Hempel, S.; Shanman, R.; Shekelle, P.G. What Is an Evidence Map? A Systematic Review of Published Evidence Maps and Their Definitions, Methods, and Products. *Syst. Rev.* **2016**, *5*, 28. <https://doi.org/10.1186/s13643-016-0204-x>.
92. Wilkinson, L. *The Grammar of Graphics*; Springer: New York, NY, USA, 2005.
93. Brunson, J.C.; Read, Q.D. *ggalluvial: Alluvial Plots in “ggplot2”*, version 0.12.6; CRAN: Vienna, Austria, 2017.
94. Sjöberg, D. *ggbump: Bump Chart and Sigmoid Curves*; CRAN: Vienna, Austria, 2021.
95. Hirsch, J.E. An Index to Quantify an Individual’s Scientific Research Output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. <https://doi.org/10.1073/pnas.0507655102>.
96. Eddebuettel, D. *digest: Create Compact Hash Digests of R Objects*, version 0.6.39; CRAN: Vienna, Austria, 2003.

97. Chuene, T.A.; Akanbi, R.T.; Chikoore, H. The Impact of Climate Change on Agricultural Nonpoint Source Pollution in the Sand River Catchment, Limpopo, South Africa. *Water* **2025**, *17*, 1818. <https://doi.org/10.3390/w17121818>.
98. Durand, J.F. The Impact of Gold Mining on the Witwatersrand on the Rivers and Karst System of Gauteng and North West Province, South Africa. *J. Afr. Earth Sci.* **2012**, *68*, 24–43. <https://doi.org/10.1016/j.jafrearsci.2012.03.013>.
99. Edokpayi, J.; Odiyo, J.; Popoola, O.; Msagati, T. Assessment of Trace Metals Contamination of Surface Water and Sediment: A Case Study of Mvudi River, South Africa. *Sustainability* **2016**, *8*, 135. <https://doi.org/10.3390/su8020135>.
100. Le Roux, J.; Morake, L.; Van Der Waal, B.; Leigh Anderson, R.; Hedding, D.W. Intra-Gully Mapping of the Largest Documented Gully Network in South Africa Using UAV Photogrammetry: Implications for Restoration Strategies. *Prog. Phys. Geogr. Earth Environ.* **2022**, *46*, 772–789. <https://doi.org/10.1177/03091333221101057>.
101. Raji, I.B.; Hoffmann, E.; Ngie, A.; Winde, F. Assessing Uranium Pollution Levels in the Rietspruit River, Far West Rand Gold-field, South Africa. *Int. J. Environ. Res. Public Health* **2021**, *18*, 8466. <https://doi.org/10.3390/ijerph18168466>.
102. Sigopi, M.; Shoko, C.; Dube, T. Advancements in Remote Sensing Technologies for Accurate Monitoring and Management of Surface Water Resources in Africa: An Overview, Limitations, and Future Directions. *Geocarto Int.* **2024**, *39*, 2347935. <https://doi.org/10.1080/10106049.2024.2347935>.
103. Singh, S. Mapping Soil Trace Metal Distribution Using Remote Sensing and Multivariate Analysis. *Environ. Monit. Assess.* **2024**, *196*, 516. <https://doi.org/10.1007/s10661-024-12682-3>.
104. Zondo, S.G. Metal Content, Bioaccumulation, Translocation, and Health Risk Assessment of Root Vegetables Grown in Kwa-Zulu-Natal Small-Scale Farms of South Africa. *Environ. Monit. Assess.* **2024**, *196*, 752. <https://doi.org/10.1007/s10661-024-12920-8>.
105. Mabogo, N.S.T.; Odera, P.A. Modelling Groundwater Vulnerability to Contamination Using DRASTIC Model through Geospatial Techniques over Northern Kwazulu-Natal, South Africa. *Geoplan. J. Geomat. Plann.* **2023**, *10*, 111–122. <https://doi.org/10.14710/geoplanning.10.2.111-122>.
106. Maherry, A.; Tredoux, G.; Clarke, S.; Engelbrecht, P. State of Nitrate Pollution in Groundwater in South Africa. In *Proceedings of the CSIR Natural Resources and the Environment Conference*; CSIR: Pretoria, South Africa, 2010.
107. Ponnusamy, D.; Elumalai, V. Determination of Potential Recharge Zones and Its Validation against Groundwater Quality Parameters through the Application of GIS and Remote Sensing Techniques in uMhlathuze Catchment, KwaZulu-Natal, South Africa. *Chemosphere* **2022**, *307*, 136121. <https://doi.org/10.1016/j.chemosphere.2022.136121>.
108. Du Preez, C.C.; Van Huyssteen, C.W. Threats to Soil and Water Resources in South Africa. *Environ. Res.* **2020**, *183*, 109015. <https://doi.org/10.1016/j.envres.2019.109015>.
109. El Bahjaouy, K.; Barakat, A.; Oussilkane, A.; Hilali, A.; Amrani, N.; Mosaid, H. Spatial Mapping of Soil Salinity in a Semiarid Region Using a Machine Learning Model Based on Spectral Indices and Ground Data. *Model. Earth Syst. Environ.* **2025**, *11*, 257. <https://doi.org/10.1007/s40808-025-02444-6>.
110. Gallo, J.A.; Lombard, A.T.; Cowling, R.M. Conservation Planning for Action: End-User Engagement in the Development and Dual-Centric Weighting of a Spatial Decision Support System. *Land* **2022**, *12*, 67. <https://doi.org/10.3390/land12010067>.
111. Muller, S.J.; Van Niekerk, A. An Evaluation of Supervised Classifiers for Indirectly Detecting Salt-Affected Areas at Irrigation Scheme Level. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 138–150. <https://doi.org/10.1016/j.jag.2016.02.005>.
112. Raw, J.; Riddin, T.; Wasserman, J.; Lehman, T.; Bornman, T.; Adams, J. Salt Marsh Elevation and Responses to Future Sea-Level Rise in the Knysna Estuary, South Africa. *Afr. J. Aquat. Sci.* **2020**, *45*, 49–64. <https://doi.org/10.2989/16085914.2019.1662763>.
113. Dabrowski, J.M. *Development of an Indicator Methodology to Estimate the Relative Exposure and Risk of Pesticides in South African Surface Waters: Report to the Water Research Commission*; Water Research Commission: Gezina, South Africa, 2011.
114. Kganyago, M.; Adjorlolo, C.; Mhangara, P. Exploring Transferable Techniques to Retrieve Crop Biophysical and Biochemical Variables Using Sentinel-2 Data. *Remote Sens.* **2022**, *14*, 3968. <https://doi.org/10.3390/rs14163968>.
115. Matshidze, M.M.; Ndou, V. Herbicide Resistance Cases in South Africa: A Review of the Current State of Knowledge. *S. Afr. J. Sci.* **2023**, *119*, 15228. <https://doi.org/10.17159/sajs.2023/15228>.
116. Mugudamani, I.; Oke, S.A.; Gumede, T.P.; Senbore, S. Herbicides in Water Sources: Communicating Potential Risks to the Population of Mangaung Metropolitan Municipality, South Africa. *Toxics* **2023**, *11*, 538. <https://doi.org/10.3390/toxics11060538>.
117. Ayeleru, O.O.; Fajimi, L.I.; Onu, M.A.; Nyam, T.T.; Dlova, S.; Ameh, V.I.; Olubambi, P.A. Estimating Plastic Waste Generation Using Supervised Time-Series Learning Techniques in Johannesburg, South Africa. *Heliyon* **2024**, *10*, e28199. <https://doi.org/10.1016/j.heliyon.2024.e28199>.
118. Dahms, H.T.J.; Greenfield, R. A Review of the Environments, Biota, and Methods Used in Microplastics Research in South Africa. *S. Afr. J. Sci.* **2024**, *120*, 16669. <https://doi.org/10.17159/sajs.2024/16669>.

119. Gutsa, T.; Trois, C.; De Vries, R.; Mani, T. Wasted Shores: Using Drones to Monitor the Spatio-Temporal Evolution of Debris Accumulation Hotspots on South Africa's Umgeni River. *Sci. Total Environ.* **2024**, *955*, 176791. <https://doi.org/10.1016/j.scitotenv.2024.176791>.
120. Maya, M.; Musekiwa, C.; Mthembi, P.; Crowley, M. Remote Sensing and Geochemistry Techniques for the Assessment of Coal Mining Pollution, Emalaheni (Witbank), Mpumalanga. *SA J. Geomat.* **2015**, *4*, 174. <https://doi.org/10.4314/sajg.v4i2.9>.
121. Mokgalaka-Fleischmann, N.S.; Melato, F.A.; Netshiongolwe, K.; Izevbehai, O.U.; Lepule, S.P.; Motsepe, K.; Edokpayi, J.N. Microplastic Occurrence and Fate in the South African Environment: A Review. *Environ. Syst. Res.* **2024**, *13*, 59. <https://doi.org/10.1186/s40068-024-00389-w>.
122. Saad, D.; Ramaremsa, G.; Ndlovu, M.; Chauke, P.; Nikiema, J.; Chimuka, L. Microplastic Abundance and Sources in Surface Water Samples of the Vaal River, South Africa. *Bull. Environ. Contam. Toxicol.* **2024**, *112*, 23. <https://doi.org/10.1007/s00128-023-03845-y>.
123. Swanepoel, S.; Marlin, D. Mapping Illegal Dumping in Nelson Mandela Bay Metro: A Study Using Image Interpretation. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101302. <https://doi.org/10.1016/j.rsase.2024.101302>.
124. Abiye, T.A.; Ali, K.A. Potential Role of Acid Mine Drainage Management towards Achieving Sustainable Development in the Johannesburg Region, South Africa. *Groundw. Sustain. Dev.* **2022**, *19*, 100839. <https://doi.org/10.1016/j.gsd.2022.100839>.
125. Abrahams, J.-L.R.; Carranza, E.J.M. Trace Metal Content Prediction along an AMD (Acid Mine Drainage)-Contaminated Stream Draining a Coal Mine Using VNIR-SWIR Spectroscopy. *Environ. Monit. Assess.* **2023**, *195*, 1261. <https://doi.org/10.1007/s10661-023-11837-y>.
126. Adepoju, K.A.; Adelabu, S.A. Improving Accuracy of Landsat-8 OLI Classification Using Image Composite and Multisource Data with Google Earth Engine. *Remote Sens. Lett.* **2020**, *11*, 107–116. <https://doi.org/10.1080/2150704X.2019.1690792>.
127. Botha, T.L.; Bamuzza-Pemu, E.; Roopnarain, A.; Ncube, Z.; De Nysschen, G.; Ndaba, B.; Mokgalaka, N.; Bello-Akinosho, M.; Adeleke, R.; Mushwana, A.; et al. Development of a GIS-Based Knowledge Hub for Contaminants of Emerging Concern in South African Water Resources Using Open-Source Software: Lessons Learnt. *Heliyon* **2023**, *9*, e13007. <https://doi.org/10.1016/j.heliyon.2023.e13007>.
128. Gasela, M.; Kganyago, M.; De Jager, G. Using Resampled nSight-2 Hyperspectral Data and Various Machine Learning Classifiers for Discriminating Wetland Plant Species in a Ramsar Wetland Site, South Africa. *Appl. Geomat.* **2024**, *16*, 429–440. <https://doi.org/10.1007/s12518-024-00560-z>.
129. Kapwata, T.; Mathee, A.; Sweijd, N.; Minakawa, N.; Mogotsi, M.; Kunene, Z.; Wright, C.Y. Spatial Assessment of Heavy Metals Contamination in Household Garden Soils in Rural Limpopo Province, South Africa. *Environ. Geochem. Health* **2020**, *42*, 4181–4191. <https://doi.org/10.1007/s10653-020-00535-0>.
130. Mhangara, P.; Kakembo, V.; Lim, K.J. Soil Erosion Risk Assessment of the Keiskamma Catchment, South Africa Using GIS and Remote Sensing. *Environ. Earth Sci.* **2012**, *65*, 2087–2102. <https://doi.org/10.1007/s12665-011-1190-x>.
131. Munghezulu, C.; Mashaba-Munghezulu, Z.; Ratshiedana, P.E.; Economon, E.; Chirima, G.; Sibanda, S. Unmanned Aerial Vehicle (UAV) and Spectral Datasets in South Africa for Precision Agriculture. *Data* **2023**, *8*, 98. <https://doi.org/10.3390/data8060098>.
132. Zhang, S.E.; Nwaila, G.T.; Bourdeau, J.E.; Ghorbani, Y.; Carranza, E.J.M. Deriving Big Geochemical Data from High-Resolution Remote Sensing Data via Machine Learning: Application to a Tailing Storage Facility in the Witwatersrand Goldfields. *Artif. Intell. Geosci.* **2023**, *4*, 9–21. <https://doi.org/10.1016/j.aiig.2023.01.005>.
133. Adelabu, S.; Mutanga, O.; Adam, E. Testing the Reliability and Stability of the Internal Accuracy Assessment of Random Forest for Classifying Tree Defoliation Levels Using Different Validation Methods. *Geocarto Int.* **2015**, *30*, 810–821. <https://doi.org/10.1080/10106049.2014.997303>.
134. Awuah, K.T.; Aplin, P.; Marston, C.G.; Powell, I.; Smit, I.P.J. Probabilistic Mapping and Spatial Pattern Analysis of Grazing Lawns in Southern African Savannahs Using WorldView-3 Imagery and Machine Learning Techniques. *Remote Sens.* **2020**, *12*, 3357. <https://doi.org/10.3390/rs12203357>.
135. Matiza, C.; Mutanga, O.; Peerbhay, K.; Odindi, J.; Lottering, R. A Systematic Review of Remote Sensing and Machine Learning Approaches for Accurate Carbon Storage Estimation in Natural Forests. *South. For. A J. For. Sci.* **2023**, *85*, 123–141. <https://doi.org/10.2989/20702620.2023.2251946>.
136. Mazarire, T.T.; Ratshiedana, P.E.; Nyamugama, A.; Adam, E.; Chirima, G. Exploring Machine Learning Algorithms for Mapping Crop Types in a Heterogeneous Agriculture Landscape Using Sentinel-2 Data. A Case Study of Free State Province, South Africa. *SA J. Geomat.* **2022**, *9*, 333–347. <https://doi.org/10.4314/sajg.v9i2.22>.

137. Peerbhay, K.; Adelabu, S.; Lottering, R.; Singh, L. Mapping Carbon Content in a Mountainous Grassland Using SPOT 5 Multi-spectral Imagery and Semi-Automated Machine Learning Ensemble Methods. *Sci. Afr.* **2022**, *17*, e01344. <https://doi.org/10.1016/j.sciaf.2022.e01344>.
138. Wessels, K.J.; Van Den Bergh, F.; Scholes, R.J. Limits to Detectability of Land Degradation by Trend Analysis of Vegetation Index Data. *Remote Sens. Environ.* **2012**, *125*, 10–22. <https://doi.org/10.1016/j.rse.2012.06.022>.
139. Dube, T.; Mutanga, O.; Sibanda, M.; Seutloali, K.; Shoko, C. Use of Landsat Series Data to Analyse the Spatial and Temporal Variations of Land Degradation in a Dispersive Soil Environment: A Case of King Sabata Dalindyebo Local Municipality in the Eastern Cape Province, South Africa. *Phys. Chem. Earth Parts A/B/C* **2017**, *100*, 112–120. <https://doi.org/10.1016/j.pce.2017.01.023>.
140. Maponya, M.G.; Van Niekerk, A.; Mashimbye, Z.E. Pre-Harvest Classification of Crop Types Using a Sentinel-2 Time-Series and Machine Learning. *Comput. Electron. Agric.* **2020**, *169*, 105164. <https://doi.org/10.1016/j.compag.2019.105164>.
141. Medina-Medina, A.J.; Salas López, R.; Barboza, E.; Tuesta-Trauco, K.M.; Zabaleta-Santiesteban, J.A.; Guzman, B.K.; Oliva-Cruz, M.; Tariq, A.; Rojas-Briceño, N.B. Participation GIS for the Monitoring of Areas Contaminated by Municipal Solid Waste: A Case Study in the City of Pedro Ruiz Gallo (Peru). *Case Stud. Chem. Environ. Eng.* **2024**, *10*, 100941. <https://doi.org/10.1016/j.csee.2024.100941>.
142. Kleynhans, W.; Salmon, B.P.; Olivier, J.C.; Van Den Bergh, F.; Wessels, K.J.; Grobler, T.L.; Steenkamp, K.C. Land Cover Change Detection Using Autocorrelation Analysis on MODIS Time-Series Data: Detection of New Human Settlements in the Gauteng Province of South Africa. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 777–783. <https://doi.org/10.1109/JSTARS.2012.2187177>.
143. Cho, M.A.; Onesimo, M.; Mabhaudhi, T. Using Participatory GIS and Collaborative Management Approaches to Enhance Local Actors' Participation in Rangeland Management: The Case of Vulindlela, South Africa. *J. Environ. Plan. Manag.* **2023**, *66*, 1189–1208. <https://doi.org/10.1080/09640568.2021.2017269>.
144. Weyer, D.; Bezerra, J.C.; De Vos, A. Participatory Mapping in a Developing Country Context: Lessons from South Africa. *Land* **2019**, *8*, 134. <https://doi.org/10.3390/land8090134>.
145. Musakwa, W.; Makoni, E.N.; Kangethe, M.; Segooa, L. Developing a Decision Support System to Identify Strategically Located Land for Land Reform in South Africa. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2014**, *XL-2*, 197–203. <https://doi.org/10.5194/isprsarchives-XL-2-197-2014>.
146. Nhamo, G. *The Increasing Risk of Floods and Tornadoes in Southern Africa*; Sustainable Development Goals Series; Springer International Publishing AG: Cham, Switzerland, 2021.
147. Olanrewaju, C.C.; Reddy, M. Assessment and Prediction of Flood Hazards Using Standardized Precipitation Index—A Case Study of eThekweni Metropolitan Area. *J. Flood Risk Manag.* **2022**, *15*, e12788. <https://doi.org/10.1111/jfr3.12788>.
148. Wessels, K.; Van Den Bergh, F.; Roy, D.; Salmon, B.; Steenkamp, K.; MacAlister, B.; Swanepoel, D.; Jewitt, D. Rapid Land Cover Map Updates Using Change Detection and Robust Random Forest Classifiers. *Remote Sens.* **2016**, *8*, 888. <https://doi.org/10.3390/rs8110888>.
149. Dube, T.; Maluleke, X.G.; Mutanga, O. Mapping Rangeland Ecosystems Vulnerability to *Lantana camara* Invasion in Semi-arid Savannahs in South Africa. *Afr. J. Ecol.* **2022**, *60*, 658–667. <https://doi.org/10.1111/aje.12951>.
150. Mansour, K.; Mutanga, O.; Everson, T.; Adam, E. Discriminating Indicator Grass Species for Rangeland Degradation Assessment Using Hyperspectral Data Resampled to AISA Eagle Resolution. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 56–65. <https://doi.org/10.1016/j.isprsjprs.2012.03.006>.
151. Iloms, E.; Ololade, O.O.; Ogola, H.J.O.; Selvarajan, R. Investigating Industrial Effluent Impact on Municipal Wastewater Treatment Plant in Vaal, South Africa. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1096. <https://doi.org/10.3390/ijerph17031096>.
152. Mudaly, L.; Van Der Laan, M. Interactions between Irrigated Agriculture and Surface Water Quality with a Focus on Phosphate and Nitrate in the Middle Olifants Catchment, South Africa. *Sustainability* **2020**, *12*, 4370. <https://doi.org/10.3390/su12114370>.
153. Jewitt, D.; Goodman, P.S.; Erasmus, B.F.N.; O'Connor, T.G.; Witkowski, E.T.F. Systematic Land-Cover Change in KwaZulu-Natal, South Africa: Implications for Biodiversity. *S. Afr. J. Sci.* **2015**, *111*, 9. <https://doi.org/10.17159/sajs.2015/20150019>.
154. Manjoro, M.; Kakembo, V.; Rowntree, K.M. Trends in Soil Erosion and Woody Shrub Encroachment in Ngqushwa District, Eastern Cape Province, South Africa. *Environ. Manag.* **2012**, *49*, 570–579. <https://doi.org/10.1007/s00267-012-9810-0>.
155. Mararakanye, N.; Le Roux, J.J. Gully Location Mapping at a National Scale for South Africa. *S. Afr. Geogr. J.* **2012**, *94*, 208–218. <https://doi.org/10.1080/03736245.2012.742786>.
156. Nuwarinda, H.; Ramoelo, A.; Adelabu, S. Assessing Natural Resource Change in Vhembe Biosphere and Surroundings. *Environ. Monit. Assess.* **2021**, *193*, 404. <https://doi.org/10.1007/s10661-021-09053-7>.

157. Sibanda, M.; Mutanga, O.; Rouget, M.; Kumar, L. Estimating Biomass of Native Grass Grown under Complex Management Treatments Using WorldView-3 Spectral Derivatives. *Remote Sens.* **2017**, *9*, 55. <https://doi.org/10.3390/rs9010055>.
158. Van Zijl, G.; Van Tol, J.; Bouwer, D.; Lorentz, S.; Le Roux, P. Combining Historical Remote Sensing, Digital Soil Mapping and Hydrological Modelling to Produce Solutions for Infrastructure Damage in Cosmo City, South Africa. *Remote Sens.* **2020**, *12*, 433. <https://doi.org/10.3390/rs12030433>.
159. Mpanyaro, Z.; Kalumba, A.M.; Zhou, L.; Afuye, G.A. Mapping and Assessing Riparian Vegetation Response to Drought along the Buffalo River Catchment in the Eastern Cape Province, South Africa. *Climate* **2024**, *12*, 7. <https://doi.org/10.3390/cli12010007>.
160. Smit, I.E.; Van Zijl, G.M.; Riddell, E.S.; Van Tol, J.J. Downscaling Legacy Soil Information for Hydrological Soil Mapping Using Multinomial Logistic Regression. *Geoderma* **2023**, *436*, 116568. <https://doi.org/10.1016/j.geoderma.2023.116568>.
161. Rebelo, A.J.; Jarman, C.; Esler, K.J.; Cowling, R.M.; Le Maitre, D.C. Water-Use Characteristics of Palmiet (*Prionium Serratum*), an Endemic South African Wetland Plant. *Water SA* **2020**, *46*, 558–572. <https://doi.org/10.17159/wsa/2020.v46.i4.9069>.
162. Bindraban, P.S.; Van Der Velde, M.; Ye, L.; Van Den Berg, M.; Materechera, S.; Kiba, D.I.; Tamene, L.; Ragnarsdóttir, K.V.; Jongschaap, R.; Hoogmoed, M.; et al. Assessing the Impact of Soil Degradation on Food Production. *Curr. Opin. Environ. Sustain.* **2012**, *4*, 478–488. <https://doi.org/10.1016/j.cosust.2012.09.015>.
163. Batwa-Ismail, M.Z.; Moodley, R.; Mutanga, O. Elemental Analysis of Soils along the South African National Road (N3)—A Combined Approach Including Statistics, Pollution Indicators, and Geographic Information System (GIS). *Environ. Monit. Assess.* **2021**, *193*, 559. <https://doi.org/10.1007/s10661-021-09374-7>.
164. Ngarava, S.; Zhou, L.; Mushunje, A.; Chaminuka, P. Vulnerability of Settlements to Floods in South Africa: A Focus on Port St Johns. In *The Increasing Risk of Floods and Tornadoes in Southern Africa*; Nhamo, G., Chapungu, L., Eds.; Sustainable Development Goals Series; Springer International Publishing: Cham, Switzerland, 2021; pp. 203–219.
165. Finca, A.; Linnane, S.; Slinger, J.; Getty, D.; Igshaan Samuels, M. Implications of the Breakdown in the Indigenous Knowledge System for Rangeland Management and Policy: A Case Study from the Eastern Cape in South Africa. *Afr. J. Range Forage Sci.* **2023**, *40*, 47–61. <https://doi.org/10.2989/10220119.2022.2138973>.
166. Degrendele, C.; Klánová, J.; Prokeš, R.; Příbylová, P.; Šenk, P.; Šudoma, M.; Röööli, M.; Dalvie, M.A.; Fuhrmann, S. Current Use Pesticides in Soil and Air from Two Agricultural Sites in South Africa: Implications for Environmental Fate and Human Exposure. *Sci. Total Environ.* **2022**, *807*, 150455. <https://doi.org/10.1016/j.scitotenv.2021.150455>.
167. Nsikani, M.M.; Novoa, A.; Van Wilgen, B.W.; Keet, J.; Gaertner, M. *Acacia saligna* 's Soil Legacy Effects Persist up to 10 Years after Clearing: Implications for Ecological Restoration. *Austral Ecol.* **2017**, *42*, 880–889. <https://doi.org/10.1111/aec.12515>.
168. Bennett, J.E.; Palmer, A.R.; Blackett, M.A. Range degradation and land tenure change: Insights from a 'released' communal area of Eastern Cape Province, South Africa. *Land Degrad. Dev.* **2012**, *23*, 557–568. <https://doi.org/10.1002/ldr.2178>.
169. Halpern, A.B.W.; Meadows, M.E. Fifty Years of Land Use Change in the Swartland, Western Cape, South Africa: Characteristics, Causes and Consequences. *S. Afr. Geogr. J.* **2013**, *95*, 38–49. <https://doi.org/10.1080/03736245.2013.806101>.
170. Gibbs, H.K.; Salmon, J.M. Mapping the World's Degraded Lands. *Appl. Geogr.* **2015**, *57*, 12–21. <https://doi.org/10.1016/j.apgeog.2014.11.024>.
171. Nxumalo, G.S.; Chauke, H. Challenges and Opportunities in Smallholder Agriculture Digitization in South Africa. *Front. Sustain. Food Syst.* **2025**, *9*, 1583224. <https://doi.org/10.3389/fsufs.2025.1583224>.
172. Samuel-Rosa, A.; Dalmolin, R.S.D.; Moura-Bueno, J.M.; Teixeira, W.G.; Alba, J.M.F. Open Legacy Soil Survey Data in Brazil: Geospatial Data Quality and How to Improve It. *Sci. Agric.* **2020**, *77*, e20170430. <https://doi.org/10.1590/1678-992x-2017-0430>.
173. Hengl, T.; De Jesus, J.M.; MacMillan, R.A.; Batjes, N.H.; Heuvelink, G.B.M.; Ribeiro, E.; Samuel-Rosa, A.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; et al. SoilGrids1km—Global Soil Information Based on Automated Mapping. *PLoS ONE* **2014**, *9*, e105992. <https://doi.org/10.1371/journal.pone.0105992>.
174. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. <https://doi.org/10.1016/j.apm.2019.12.016>.
175. Le Roux, J.; Morgenthal, T.; Malherbe, J.; Pretorius, D.; Sumner, P. Water Erosion Prediction at a National Scale for South Africa. *Water SA* **2018**, *34*, 305. <https://doi.org/10.4314/wsa.v34i3.180623>.
176. Vrieling, A. Satellite Remote Sensing for Water Erosion Assessment: A Review. *Catena* **2006**, *65*, 2–18. <https://doi.org/10.1016/j.catena.2005.10.005>.
177. Jin, F.; Yang, W.; Fu, J.; Li, Z. Effects of Vegetation and Climate on the Changes of Soil Erosion in the Loess Plateau of China. *Sci. Total Environ.* **2021**, *773*, 145514. <https://doi.org/10.1016/j.scitotenv.2021.145514>.
178. Chen, H.; Teng, Y.; Lu, S.; Wang, Y.; Wang, J. Contamination Features and Health Risk of Soil Heavy Metals in China. *Sci. Total Environ.* **2015**, *512–513*, 143–153. <https://doi.org/10.1016/j.scitotenv.2015.01.025>.

179. Yang, L.; Ge, S.; Liu, J.; Iqbal, Y.; Jiang, Y.; Sun, R.; Ruan, X.; Wang, Y. Spatial Distribution and Risk Assessment of Heavy Metal(Oid)s Contamination in Topsoil around a Lead and Zinc Smelter in Henan Province, Central China. *Toxics* **2023**, *11*, 427. <https://doi.org/10.3390/toxics11050427>.
180. Zhang, X.; Xue, J.; Chen, S.; Wang, N.; Shi, Z.; Huang, Y.; Zhuo, Z. Digital Mapping of Soil Organic Carbon with Machine Learning in Dryland of Northeast and North Plain China. *Remote Sens.* **2022**, *14*, 2504. <https://doi.org/10.3390/rs14102504>.
181. Tutu, H.; McCarthy, T.S.; Cukrowska, E. The Chemical Characteristics of Acid Mine Drainage with Particular Reference to Sources, Distribution and Remediation: The Witwatersrand Basin, South Africa as a Case Study. *Appl. Geochem.* **2008**, *23*, 3666–3684. <https://doi.org/10.1016/j.apgeochem.2008.09.002>.
182. Nawar, S.; Buddenbaum, H.; Hill, J.; Kozak, J.; Mouazen, A.M. Estimating the Soil Clay Content and Organic Matter by Means of Different Calibration Methods of Vis-NIR Diffuse Reflectance Spectroscopy. *Soil Tillage Res.* **2016**, *155*, 510–522. <https://doi.org/10.1016/j.still.2015.07.021>.
183. Padarian, J.; Minasny, B.; McBratney, A.B. Using Deep Learning for Digital Soil Mapping. *Soil* **2019**, *5*, 79–89. <https://doi.org/10.5194/soil-5-79-2019>.
184. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Amirian-Chakan, A.; Rentschler, T.; Zeraatpisheh, M.; Sarmadian, F.; Valavi, R.; Davatgar, N.; Behrens, T.; Scholten, T. Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. *Remote Sens.* **2020**, *12*, 1095. <https://doi.org/10.3390/rs12071095>.
185. Vågen, T.-G.; Winowiecki, L.A.; Tondoh, J.E.; Desta, L.T.; Gumbrecht, T. Mapping of Soil Properties and Land Degradation Risk in Africa Using MODIS Reflectance. *Geoderma* **2016**, *263*, 216–225. <https://doi.org/10.1016/j.geoderma.2015.06.023>.
186. Behrens, T.; Schmidt, K.; MacMillan, R.A.; Viscarra Rossel, R.A. Multiscale Contextual Spatial Modelling with the Gaussian Scale Space. *Geoderma* **2018**, *310*, 128–137. <https://doi.org/10.1016/j.geoderma.2017.09.015>.
187. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems (NIPS 2017)*; NeurIPS: San Diego, CA, USA, 2017, Volume 30.
188. Topouzelis, K.; Papakonstantinou, A.; Garaba, S.P. Detection of Floating Plastics from Satellite and Unmanned Aerial Systems (Plastic Litter Project 2018). *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *79*, 175–183. <https://doi.org/10.1016/j.jag.2019.03.011>.
189. Ma, W.; Ding, M.; Bian, Z. Comprehensive Assessment of Exposure and Environmental Risk of Potentially Toxic Elements in Surface Water and Sediment across China: A Synthesis Study. *Sci. Total Environ.* **2024**, *926*, 172061. <https://doi.org/10.1016/j.scitotenv.2024.172061>.
190. Biermann, L.; Clewley, D.; Martinez-Vicente, V.; Topouzelis, K. Finding Plastic Patches in Coastal Waters Using Optical Satellite Data. *Sci. Rep.* **2020**, *10*, 5364. <https://doi.org/10.1038/s41598-020-62298-z>.
191. Metternicht, G.I.; Zinck, J.A. Remote Sensing of Soil Salinity: Potentials and Constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20. [https://doi.org/10.1016/S0034-4257\(02\)00188-8](https://doi.org/10.1016/S0034-4257(02)00188-8).
192. Douaoui, A.E.K.; Nicolas, H.; Walter, C. Detecting Salinity Hazards within a Semiarid Context by Means of Combining Soil and Remote-Sensing Data. *Geoderma* **2006**, *134*, 217–230. <https://doi.org/10.1016/j.geoderma.2005.10.009>.
193. Dehaan, R.L.; Taylor, G.R. Field-Derived Spectra of Salinized Soils and Vegetation as Indicators of Irrigation-Induced Soil Salinization. *Remote Sens. Environ.* **2002**, *80*, 406–417. [https://doi.org/10.1016/S0034-4257\(01\)00321-2](https://doi.org/10.1016/S0034-4257(01)00321-2).
194. Brown, G.; Kyttä, M. Key Issues and Research Priorities for Public Participation GIS (PPGIS): A Synthesis Based on Empirical Research. *Appl. Geogr.* **2014**, *46*, 122–136. <https://doi.org/10.1016/j.apgeog.2013.11.004>.
195. Sena-Vittini, M.; Gomez-Valenzuela, V.; Ramirez, K. Social Perceptions and Conservation in Protected Areas: Taking Stock of the Literature. *Land Use Policy* **2023**, *131*, 106696. <https://doi.org/10.1016/j.landusepol.2023.106696>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.