On the Semantics of Big Earth Observation Data for Land Classification

Gilberto Camara

National Institute for Space Research, Brazil

Abstract: This paper discusses the challenges of using big Earth observation data for land classification. The approach taken is to consider pure data-driven methods to be insufficient to represent continuous change. We argue for sound theories when working with big data. After revising existing classification schemes such as FAO's Land Cover Classification System (LCCS), we conclude that LCCS and similar proposals cannot capture the complexity of landscape dynamics. We then investigate concepts that are being used for analyzing satellite image time series; we show these concepts to be instances of events. Therefore, for continuous monitoring of land change, event recognition needs to replace object identification as the prevailing paradigm. The paper concludes by showing how event semantics can improve data-driven methods to fulfil the potential of big data.

Keywords: Big data, Earth observation, Geospatial semantics, LUCC, Land-use change

---

1 Introduction

Satellite images are the most comprehensive source of data about our environment; they provide essential information on global challenges. Images provide information for measuring deforestation, crop production, food security, urban footprints, water scarcity, land degradation, among other uses. In recent years, space agencies have adopted open distribution policies. Petabytes of Earth observation data are now available. Experts now have access to repeated acquisitions over the same areas; the resulting time series improve our understanding of ecological patterns and processes [36]. Instead of selecting individual

images from specific dates and comparing them, researchers can track changes continuously [49]. To handle big data, scientists are developing new algorithms for image time series (for recent surveys, see [17,51,52]). These methods are *data-driven and theory-limited*. However, numbers do not speak for themselves [5]. Data-driven approaches without solid theories can lead to results which will not increase our knowledge [26].

What could be the effect of data-rich research in Geography and GIScience? Miller and Goodchild [33] state: "data-driven research should support, not replace, decision making by intelligent and skeptical humans". Kwan [28] recommends a critical evaluation of big data algorithms. Li et al. [29] call for fresh approaches to obtain "causal and explanatory relationships from big spatial data". In their view, theory-free or theory-poor models are not sufficient for conceptual advances in our knowledge of geographical reality. We need sound theories to deal with big data without drowning in it.

Consider how experts use Earth observation data. Their input are images with resolution ranging from 5 to 500 meters, produced by satellites such as Landsat, Sentinels-1/2/3, and CBERS-4. To extract information, experts use methods that assign a label to each pixel (e.g., 'grasslands'). Labels can represent either *land cover* or *land use*. Land cover is the observed biophysical cover of the Earth's surface; land use concepts describe socio-economic activities [8]. Thus, 'forest' is a type of land cover, while 'corn plantation' is a kind of land use. To support land classification, scientists have proposed ontologies and descriptive schemes [20]. We might thus ask: *Are the current classification systems suitable to represent land change when working with big data? If not, which concepts are needed and how should they be applied?*

In what follows, we present the prevailing consensus on classification systems: FAO's Land Cover Classification System (LCCS) [21]. We argue that LCCS does not meet the challenges posed by big data. To support our views, we consider concepts used on image time series analysis; we show these concepys are related to event recognition and are not representable in LCCS. To improve the theory behind big data, we introduce elements of an event-centered ontology for land classification.

2 Classification systems for Earth observation data: current status

The act of classification raises philosophical questions dating as far back as Aristotle. We use an *a priori* conception of reality to classify the world; what we observe has to fit our categories. Words in our language describe elements of the external reality. However, geographical terms such as 'mountain' and 'river' are imprecise and context-dependent [31, 45,46]. These ambiguities have motivated research on geospatial ontologies and semantics [13, 27, 44]. In the early 2000s, there was much optimism about the impact of unified ontologies on modeling and interoperability of geographical information. However, building such complete ontologies is hard. Janowicz et al. [22] recognize that "geographical concepts are situated and context-dependent, can be described from different, equally valid, points of view, and ontological commitments are arbitrary to a large extent". Work on classification systems has shifted. Rather than using a single ontology, the current consensus argues for domain ontologies based on a common foundational ontology. These domain ontologies

are means of making concepts of specific disciplines explicit and better communicating them [6, 43].

The semantics of Earth observation data are constrained by classification systems. Experts agree on what are the possible descriptions of the objects in the image (e.g., 'forest', 'river', 'pasture'). Each pixel of the image is then labeled using visual or automated interpretation. As an example, for countries reporting greenhouse gas inventories, the International Panel of Climate Change (IPCC) restricts the top-level land classes to 'forest', 'cropland', 'grassland', 'wetlands', 'settlements', and 'others'. This approach is too simplistic. Sasaki and Putz [41] criticize the IPCC base classes for inducing wrong assessments for ecological and biodiversity conservation. The IPCC classes are an example where pre-conceived rules collide with the diversity of the world's ecosystems.

Since land classification provides essential information about our environment, many GIScience researchers have addressed the subject of land use and land cover semantics [2,3,9]. They investigated consistency of classification systems [23], semantic similarity between terms used by different systems [12], and disagreements between results [14]. The current consensus favors ontologies aiming at unambiguous definitions of land cover classes, such as the FAO Land Cover Classification System (LCCS) [21]. For this reason, it is important to discuss whether LCCS works well with big EO data.

FAO has developed the Land Cover Classification System (LCCS) "to provide a consistent framework for the classification and mapping of land cover" [11]. LCCS is a hierarchical system. At its highest level, LCCS has eight major land cover types:

1. Cultivated and managed terrestrial areas.
2. Natural and semi-natural terrestrial vegetation.
3. Cultivated aquatic or regularly flooded areas.
4. Natural and semi-natural aquatic or regularly flooded vegetation.
5. Artificial surfaces and associated areas.
6. Bare areas.
7. Artificial water bodies, snow, and ice.
8. Natural water bodies, snow, and ice.

The division on eight classes considers three criteria: presence of vegetation, edaphic conditions, and artificiality of cover [11]. Specialization of top-level LCCS classes uses properties such as life form, tree height, and vegetation density, setting pre-defined limits (e.g., "tree height > 10 meters"). These subdivisions are *ad hoc* and application-dependent, leading to a combinational explosion with dozens or even hundreds of subclasses [21]. Such high expressive power can lead to incompatible LCCS-based class hierarchies [23].

LCCS is a landmark initiative; it provides a basis for a common understanding of land cover concepts. Many global and regional land mapping products use LCCS, including GLOBCOVER [4] and ESA CCI Land Cover [30]. However, LCCS makes assumptions which limit its use with big data:

1. LCCS describes land properties based only on land cover types, disregarding land use. For example, LCCS does not distinguish 'pasture' from 'natural grasslands'; it labels both as herbaceous land cover types.
2. The LCCS hierarchy uses hard boundaries between its subclasses. At each level of the hierarchy, properties of subclasses use fixed values (e.g., "sparse forests have

between 10% and 30% of trees"). Real-world class boundaries do not fit into such strict definitions. When doing data analysis with machine learning, boundaries between classes are data-dependent and cannot be set a priori [19].

3. Classification in LCCS has no temporal reference. LCCS assumes that subtype properties (e.g., percent of tree cover) are detectable at the moment of classification. These properties do not refer to past or future values. Land use and land cover types whose values require time references (e.g., "forest land cleared in the last decade") are not representable in LCCS.

For example, the UNFCCC Reduction of Emissions by Deforestation and Degradation initiative (REDD+) requires representing and measuring forest dynamics [10]. Static and rigid definitions of 'forest' used by LCCS cannot represent concepts such as 'forest degradation' [40]. Forest degradation happens when a natural forest loses part of its biodiversity and its tree cover. It is not a stable state but an intermediary situation that can lead to different medium-term outcomes. One can restore a degraded forest; degradation may continue and lead to complete loss of forest cover. Whatever the case, LCCS lacks explicit temporal information to capture forest degradation and thus support initiatives such as REDD+. Therefore, LCCS is thus not fit for many critical applications of EO data.

3 Elements of an ontology of land use change

3.1 Overview

To represent change in geographical space, GIScience authors distinguish between *continuants* and *occurrents* [15, 16, 18, 50]. Continuants refer to entities that "endure through time even while undergoing different sorts of changes" [18]. The Amazon Forest and the city of Brasilia are continuants. Occurrents happen in a well-defined period and may have different stages during this time. Cutting down a forest area, cultivating a crop in a season, and building a road are occurrents. Objects are associated to continuants and events to occurrents.Philosophers have debated whether these two kinds of entities can be reduced to a single ontology or whether they are different perspectives of the same reality [42]. We follow Grenon and Smith [18] who distinguish two top-level ontologies: one for continuants and another for occurrents.

Atemporal classification systems such as LCCS refer only to properties of continuants. One can state facts such as "this area has 30% forest cover" using LCCS, but cannot assert that "this area lost 70% of its forest in the last two years". To convey change, classification systems for big data need to include occurrents. In what follows, we discuss concepts used in the analysis of satellite image time series. These time series are extracted from organized collections of Earth observation data covering a geographical area in regular temporal intervals. These concepts include 'land-use change trajectory', 'trend', 'break', 'disturbance', and 'degradation'; they refer to occurrents and should be part of classification systems used in big EO data analysis.

A caveat is in order. Philosophy of Language works such as Vendler [47] and Mourelatos [34] associate events to verbs (e.g., 'run', 'walk', 'swim'). When dealing with land change analysis, we use nouns (e.g., 'trend', 'break'). These nouns describe how properties measured

in images change; as such, they represent occurrences. Although the proposed terms combine events with their measured properties, this is not a problem since the context is clear.
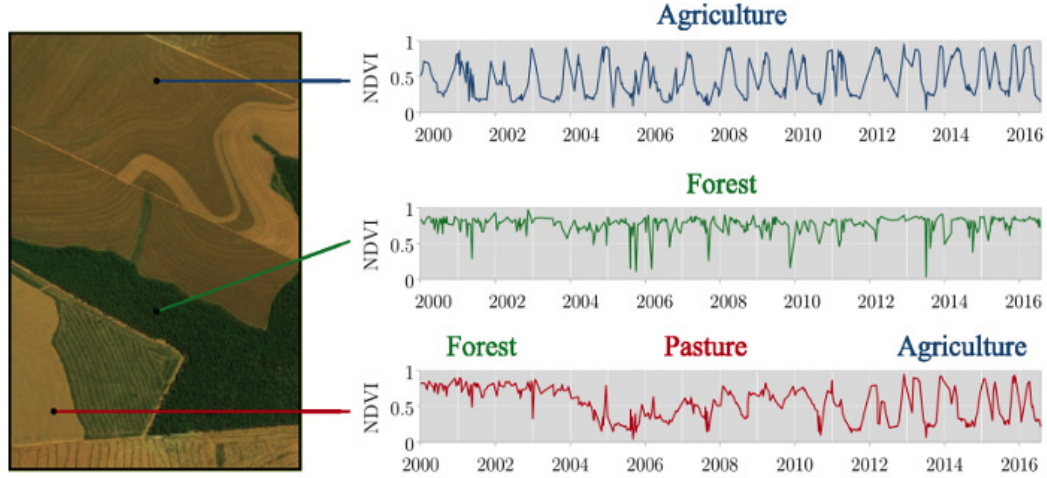
3.2 Land-use change trajectories

Figure 1: Evolution of vegetation index in three locations in an agricultural region in Brazil, measured by MODIS sensor every 16 days from 2000 to 2017 (source: courtesy of Rolf Simões).

As a foundational concept, we propose 'land-use change trajectory' or 'trajectory' for short. A trajectory is a time series $l = \{(v_1, t_1), ..., (v_n, t_n)\}, v_i \in V, t_i \in T$. Each value v_i from a domain V is associated to an interval t_i from a set T. Trajectories can be generic measures ("this dataset captures changes in one area from 2000 to 2020") or denote specific cases ("this dataset corresponds to an undisturbed forest from 2000 to 2010").

Consider Figure 1, which shows three locations in Brazil. On the left, a high-resolution image shows a snapshot of the area. The graphics show the NDVI vegetation index, measured by the MODIS sensor every 16 days from 2000 to 2017. Spikes in the graphs are noise from clouds and can be ignored in the discussion. In the top graph, the vegetation index has a quasi-periodic variation, consistent with measures of agricultural areas. The middle graph shows a quasi-constant signal typical of a forest area. The bottom graphic reveals a more complex pattern. From 2000 to 2002, the signal is compatible with a forest. The vegetation index decreases from 2003 to 2005, indicating a deforestation event. From 2007 to 2010, the signal shows a small annual variation, which suggests the area was used as pasture for cattle raising. From 2011 to 2017, the index becomes similar to the top graph, which shows conversion to agriculture. In all cases, one can speak of land-use change trajectories measured by the sequence of satellite images.

3.3   Patterns

It is useful to identify 'patterns', defined as trajectories l_p representing a known event. Patterns l_p are estimated from the data using statistical approximations [32], curve fitting [24] or other adjustment methods [51]. Phenological models are important subclasses of patterns; they are useful for agricultural monitoring since they capture the onset, growth, maximum, and senescence of croplands [51]. For event recognition, researchers combine patterns with matching functions. Given a pattern l_p and a trajectory l, a matching function $f_m(l_p, l)$ measures how much l is similar to l_p. The most common matching functions are distance metrics $D(l_p, l)$ such as Euclidean distance and dynamic time warping [32, 38]. Pattern matching is the basis for many techniques for change detection using satellite images [52].

Chazdon el al [7] use patterns to address the question: *when is a forest a forest?* The authors argue that current forest definitions used by institutions such as FAO and UNFCCC are incomplete. They state: "To assess and monitor forest and reforests properly requires viewing them as dynamic systems". The authors propose forest definitions based on patterns, as Figure 2 shows. Different situations can occur, depending on how the structural complexity of forests varies in time. Trajectory (2) in Figure 2 shows a complete loss of complexity, resulting in deforestation; trajectory (3) shows a case of forest degradation followed by a recovery. Other trajectories include the case of successful regeneration (case 4) and regeneration interrupted by a deforestation event (case 5). In all cases, the trajectories describe events. Therefore, understanding forest evolution requires event recognition.
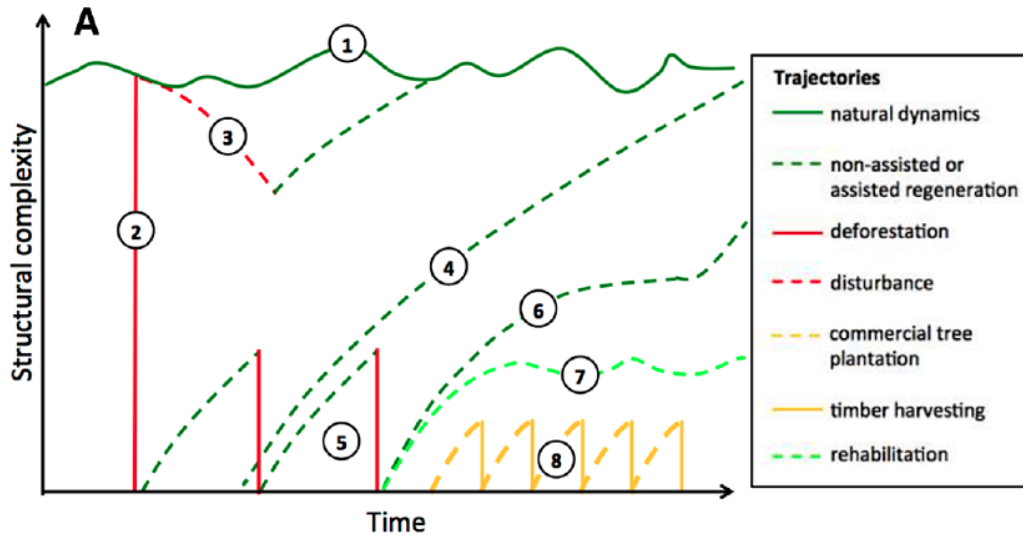

Figure 2: Forest trajectories in terms of their structural complexity over time (source: Chazdon et al [7]).

Chazdon el at. [7] point out that all trajectories shown in Figure 2 are compatible with FAO's definition of forest: "Land with tree crown cover of more than 10% and area of more than 0.5 ha. The trees should be able to reach a minimum height of 5 m at maturity in situ. Also includes areas normally forming part of the forest area which are temporarily unstocked as a result of human intervention or natural causes but which are expected to revert to forest". The FAO forest definition is an awkward attempt to take forest dynamics into account, resulting in confusion and leading to subjective interpretations. To support critical applications such as REDD+, FAO needs to update its definitions of 'forest' to include the temporal element.

3.4 Labeled trajectories

A labeled pattern is a trajectory $l = \{(v_1, t_1), ..., (v_n, t_n)\}$ to which we attach a description. The most common label is a class name such as 'forest' or 'grassland. Consider a trajectory described as "variation of vegetation index in location (-57.32, -11.56) from 2000 to 2010". By assigning a class label to the trajectory, we provide additional information about it. For example, if we find the area has been a forest area (see Figure 1) during this period, the trajectory's description becomes "location (-57.32, -11.56) was a forest from 2000 to 2010".

As described in the previous section, patterns are useful for event recognition and thus suited for labeling trajectories. An alternative technique is machine learning, done in four phases: (a) define the classification scheme; (b) select the time interval; (c) train the model; (d) classify the data. Users first define a set of classes $C = \{c_1, \ldots, c_n\}$. Then they choose a temporal interval for classification of time series (e.g., one year). The training data are sets of trajectories $T = \{L_1, \ldots, L_n\}$; each trajectory fits into the chosen time interval and has a unique label. The result of the training is a function $f : L \to C$, which assigns a class c to each trajectory l. Given the right conditions, machine learning produces good results for event recognition [1, 37, 39].

Machine learning has limitations for event recognition. The machine learning function $f : L \to C$ assigns a unique class label c to a trajectory l. This function only works for event recognition if each trajectory is homogeneous inside the chosen time interval. Consider the situation of the time series in the lower part of Figure 1. There is a transitional period from 2004 to 2006 when the forest was being removed and replaced by pasture. If training data does not include this transition, classification by machine learning will not recognize this event. Since most machine learning algorithms use fixed time intervals and hard boundary classes, they have limited explanatory power when dealing with transitional periods.

3.5 Trends and breaks

Additional concepts for occurrents in time series of EO data include 'trends' and 'breaks'. A trend refers to long-term, average tendency of signals associated to an area on the ground [25]. in most cases, trends are estimated using linear regressions [25, 48]. Given a trajectory $l = \{(v_1, t_1), ..., (v_n, t_n)\}$, a trend is a linear approximation $l_v = \alpha + \beta * t_i, t_i \in \{t_1, \ldots, t_n\}$. Trend analysis methods identify homogeneous subsets of a time series, supporting event recognition.

A break signals abrupt changes in the time series and signals that a significant disturbance happened [48]. The OECD Glossary of Statistical Terms [35], states: "breaks in statistical time series occur when there is a change in the standards for defining and observing a variable over time". Thus, breaks are instantaneous events associated to a discontinuity. Given a trajectory $l = \{(v_1, t_1), ..., (v_n, t_n)\}$, a break corresponds to an event b associated to a value and a time (v_b, t_b) that splits this trajectory in three parts: $l = \{l_{p1}, b, l_{p2}\}$ where l_{p1} and l_{p1} are instances of two patterns, one occurring before and other after the break. Break detection is useful for event recognition; a break is a temporal boundary between two trajectories. Figure 3 shows an example of a break detection algorithm BFAST [48] applied to a MODIS time series of one pixel in a pine plantation. The granularity of the MODIS time series is 16 days. BFAST detects various breaks in the time series, associated to harvesting actions.
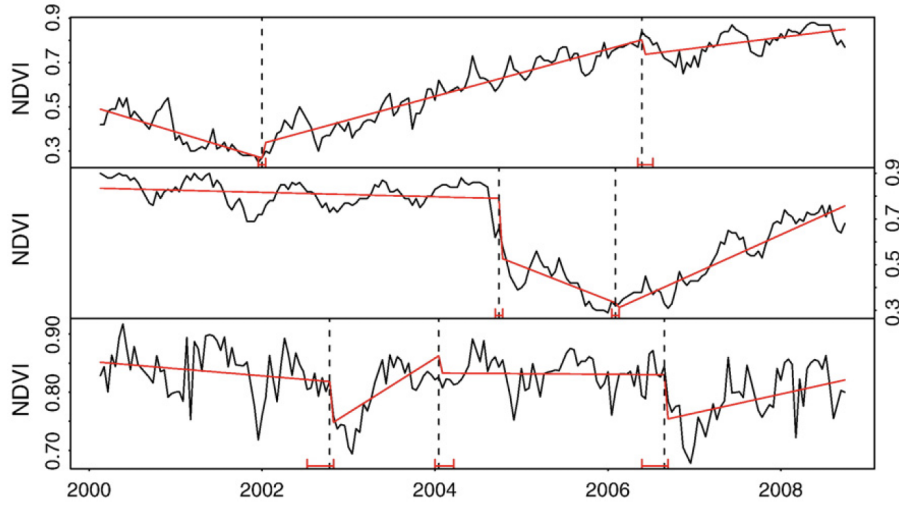

Figure 3: Detected changes in the trend component (red) of 16-day NDVI time series (black) extracted from a single MODIS pixel within a pine plantation, planted in 2001 (top), harvested in 2004 (middle), and with tree mortality occurring in 2007 (bottom) (source: Verbesselt el al. [48]).

3.6 A hierarchy of events for land classification

In the previous sections, we examined concepts used for representing information about occurrents in satellite image time series. These concepts make up a hierarchy of types (see Figure 4). Taking land-use change trajectory as the top concept, we propose 'patterns' and 'labeled trajectories' as subtypes of trajectories. Breaks are different kinds of events. While trajectories span an interval, breaks are instantaneous considering the temporal granularity used in time series. More complex concepts such as 'forest degradation' and 'deforestation' are subtypes of 'labeled trajectories' which can be identified by pattern matching or by classification. For example, consider trajectory (3) in Figure 2, which Chazdon et al [7] use as an example of forest disturbance. A pattern-matching algorithm would compare

different time series with a pattern of forest disturbance. The alternative is to use machine learning with good training data. If the data has spatial and temporal resolution to capture the trajectories, both approaches are adequate for event recognition. Experts can extend the hierarchy further by identifying additional types of events.
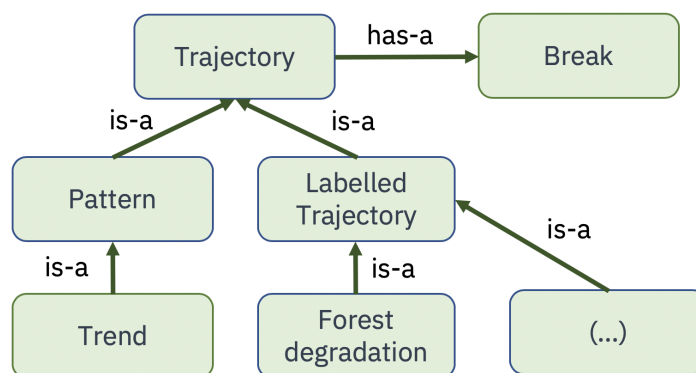
Figure 4: Proposed hierarchy of concepts related to event recognition in satellite image time series (source: author).

4 Conclusion: event recognition as a basis for big Earth observation data analysis

In this paper, we discuss the challenge of supporting big data analysis with sound theory. We argue that time series analysis, including pattern matching, trend analysis, break detection, and time series classification, are subtypes of *event recognition*. When doing continuous monitoring of land change, it is not advisable to use LCCS and similar approaches. Instead of identifying classes such as 'forest' or 'grasslands', data analysis methods need to recognize events such as "this area was a forest from 2000 to 2010, then it was deforested in 2011, and turned into grasslands from 2011 to 2020". When doing continuous monitoring, event recognition replaces object identification as the purpose of land classification.

The emphasis on event recognition has significant consequences for the design of algorithms and classification systems for big data. In particular, machine learning is not a panacea. Continuous monitoring of land dynamics using remote sensing data differs from applications such as spam filters, automatic translation, and object detection. Land systems do not change overnight. There is a period of transition for land cover conversion. Depletion of natural resources such as forests and wetlands can take place over months or even years. Monitoring subtle land transitions is crucial for protecting our environment. Thus, machine learning methods need to be adapted to work with satellite image time series.

Most algorithms for big EO data analysis use techniques that have proven useful in other problems. However, monitoring natural resources is more complex than detecting

spam emails or playing chess. While machine learning and pattern analysis are useful, there is still much to be done to build sound theories for dealing with big EO data. Long-term progress will depend on a new generation of methods that combine machine learning with functional ecosystem models. A better theoretical basis is essential for algorithms that extract information from petabytes of free EO data. This new generation of combined methods will allow a better understanding of the processes that drive landscape dynamics.

Acknowledgments

References

[1] Adarme, M., Queiroz Feitosa, R., Nigri Happ, P., Aparecido De Almeida, C., and Rodrigues Gomes, A. Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery. *Remote Sensing 12*, 6 (Jan. 2020), 910. https://doi.org/10.3390/rs12060910.

[2] Ahlqvist, O. Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science 19*, 7 (2005), 831–857. https://doi.org/10.1080/13658810500106729.

[3] Ahlqvist, O., Varanka, D., Fritz, S., and Janowick, K., Eds. *Land Use and Land Cover Semantics: Principles, Best Practices, and Prospects*. CRC Press, 2017. https://doi.org/10.1201/9781351228596.

[4] Arino, O., Gross, D., Ranera, F., Leroy, M., Bicheron, P., Brockman, C., Defourny, P., Vancutsem, C., Achard, F., Durieux, L., Bourg, L., Latham, J., Di Gregorio, A., Witt, R., Herold, M., Sambale, J., Plummer, S., and Weber, J.-L. GlobCover: ESA service for global land cover from MERIS. In *2007 IEEE International Geoscience and Remote Sensing Symposium* (July 2007), pp. 2412–2415. https://doi.org/10.1109/IGARSS.2007.4423328.

[5] boyd, d., and Crawford, K. Critical Questions for Big Data. *Information, Communication & Society 15*, 5 (June 2012), 662–679. 10.1080/1369118X.2012.678878.

[6] Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., Lewis, S. E., and the ENVO Consortium. The environment ontology: Contextualising biological and biomedical entities. *Journal of Biomedical Semantics 4*, 1 (Dec. 2013), 43. 10.1186/2041-1480-4-43.

[7] Chazdon, R. L., et al. When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio 45*, 5 (2016), 538–50. https://doi.org/10.1007/s13280-016-0772-y.

[8] COMBER, A. The separation of land cover from land use using data primitives. *Journal of Land Use Science 3*, 4 (2008), 215–229. https:/doi.org/10.1080/17474230802465173.

[9] COMBER, A., FISHER, P., AND WADSWORTH, R. What is Land Cover? *Environment and Planning B: Planning and Design 32*, 2 (Apr. 2005), 199–209. 10.1068/b31135.

[10] CORBERA, E., AND SCHROEDER, H. Governing and implementing REDD+. *Environmental Science & Policy 14*, 2 (Mar. 2011), 89–99. https://doi.org/10.1016/j.envsci.2010.11.002.

[11] DI GREGORIO, A. Land Cover Classification System - Classification concepts Software version 3. Tech. rep., FAO, 2016.

[12] FENG, C. C., AND FLEWELLING, D. M. Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems 28*, 3 (May 2004), 229–246. https://doi.org/10.1016/S0198-9715(03)00020-6.

[13] FONSECA, F. T., EGENHOFER, M. J., AGOURIS, P., AND CAMARA, G. Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS 6*, 3 (2002), 231–257. 10.1111/1467-9671.00109.

[14] FRITZ, S., SEE, L., MCCALLUM, I., SCHILL, C., OBERSTEINER, M., VANDERVELDE, M., BOETTCHER, H., HAVLÍK, P., AND ACHARD, F. Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters 6*, 4 (2011), 044005. https://doi.org/10.1088/1748-9326/6/4/044005.

[15] GALTON, A. Fields and Objects in Space, Time, and Space-time. *Spatial Cognition & Computation 4*, 1 (Mar. 2004), 39–68. https://doi.org/10.1088/1748-9326/6/4/04400510.1207/s15427633scc0401_4.

[16] GALTON, A. Experience and History: Processes and their Relation to Events. *Journal of Logic and Computation 18*, 3 (June 2008), 323–340. https://doi.org/10.1088/1748-9326/6/4/04400510.1093/logcom/exm079.

[17] GOMEZ, C., WHITE, J. C., AND WULDER, M. A. Optical remotely sensed time series data for land cover classification: A review. *{ISPRS} Journal of Photogrammetry and Remote Sensing 116* (2016), 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.

[18] GRENON, P., AND SMITH, B. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition & Computation 4*, 1 (Mar. 2004), 69–104. https://doi.org/10.1207/s15427633scc0401_5.

[19] HASTIE, T., TIBSHIRANI, R., AND J, F. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer, New York, 2009.

[20] HEROLD, M., HUBALD, R., AND DI GREGORIO, A. Translating and evaluating land cover legends using the UN Land Cover Classification System (LCCS). Tech. rep., GOFC-GOLD Florence, Italy, 2009.

[21] HEROLD, M., WOODCOCK, C., ANTONIO DI GREGORIO, MAYAUX, P., BELWARD, A., LATHAM, J., AND SCHMULLIUS, C. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing 44*, 7 (July 2006), 1719–1727. https://doi.org/10.1109/TGRS.2006.871219.

[22] JANOWICZ, K., SCHEIDER, S., PEHLE, T., AND HART, G. Geospatial semantics and linked spatiotemporal data – Past, present, and future. *Semantic Web 3*, 4 (Jan. 2012), 321–332. https://doi.org/10.3233/SW-2012-0077.

[23] JANSEN, L. J., GROOM, G., AND CARRAI, G. Land-cover harmonisation and semantic similarity: Some methodological issues. *Journal of Land Use Science 3*, 2-3 (Oct. 2008), 131–160. https://doi.org/10.1080/17474230802332076.

[24] JONSSON, P., AND EKLUNDH, L. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing 40*, 8 (Aug. 2002), 1824–1832. https://doi.org/10.1109/TGRS.2002.802519.

[25] KENNEDY, R. E., YANG, Z., AND COHEN, W. B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sensing of Environment 114*, 12 (Dec. 2010), 2897–2910. https://doi.org/10.1016/j.rse.2010.07.008.

[26] KITCHIN, R. Big Data, new epistemologies and paradigm shifts. *Big Data & Society 1*, 1 (Jan. 2014), 2053951714528481. https://doi.org/10.1177/2053951714528481.

[27] KUHN, W. Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics 3* (2005), 1–24. https://doi.org/10.1007/11496168_1.

[28] KWAN, M.-P. Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge. *Annals of the American Association of Geographers 106*, 2 (Mar. 2016), 274–282. https://doi.org/10.1080/00045608.2015.1117937.

[29] LI, S., DRAGICEVIC, S., CASTRO, F. A., SESTER, M., WINTER, S., COLTEKIN, A., PETTIT, C., JIANG, B., HAWORTH, J., STEIN, A., AND CHENG, T. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing 115* (May 2016), 119–133. https://doi.org/10.1016/j.isprsjprs.2015.10.012.

[30] LI, W., CIAIS, P., MACBEAN, N., PENG, S., DEFOURNY, P., AND BONTEMPS, S. Major forest changes and land cover transitions based on plant functional types derived from the ESA CCI Land Cover product. *International Journal of Applied Earth Observation and Geoinformation 47* (May 2016), 30–39. https://doi.org/10.1016/j.jag.2015.12.006.

[31] MARK, D. M., AND TURK, A. G. Landscape Categories in Yindjibarndi: Ontology, Environment, and Language. In *Spatial Information Theory. Foundations of Geographic Information Science* (Berlin, Heidelberg, 2003), W. Kuhn, M. F. Worboys, and S. Timpf, Eds., Lecture Notes in Computer Science, Springer, pp. 28–45. https://doi.org/10.1007/978-3-540-39923-0_3.

[32] MAUS, V., CAMARA, G., CARTAXO, R., SANCHEZ, A., RAMOS, F. M., AND QUEIROZ, G. R. A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9*, 8 (2016), 3729–3739. https://doi.org/10.1109/JSTARS.2016.2517118.

[33] MILLER, H. J., AND GOODCHILD, M. F. Data-driven geography. *GeoJournal 80*, 4 (Aug. 2015), 449–461. https://doi.org/10.1007/s10708-014-9602-6.

[34] MOURELATOS, A. P. D. Events, processes, and states. *Linguistics and Philosophy 2*, 3 (Jan. 1978), 415–434. https://doi.org/10.1007/BF00149015.

[35] OECD. The OECD glossary of statistical terms. Tech. rep., Organisation for Economic Cooperation and Development, 2003.

[36] PASQUARELLA, V. J., HOLDEN, C. E., KAUFMAN, L., AND WOODCOCK, C. E. From imagery to ecology: Leveraging time series of all available LANDSAT observations to map and monitor ecosystem state and dynamics. *Remote Sensing in Ecology and Conservation 2*, 3 (2016), 152–170. https://doi.org/10.1002/rse2.24.

[37] PELLETIER, C., VALERO, S., INGLADA, J., CHAMPION, N., AND DEDIEU, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment 187* (2016), 156–168. https://doi.org/10.1016/j.rse.2016.10.010.

[38] PETITJEAN, F., INGLADA, J., AND GANCARSKI, P. Satellite Image Time Series Analysis Under Time Warping. *IEEE Transactions on Geoscience and Remote Sensing 50*, 8 (2012), 3081–3095. https://doi.org/10.1109/TGRS.2011.2179050.

[39] PICOLI, M., CAMARA, G., SANCHES, I., SIMOES, R., CARVALHO, A., MACIEL, A., COUTINHO, A., ESQUERDO, J., ANTUNES, J., BEGOTTI, R. A., ARVOR, D., AND ALMEIDA, C. Big earth observation time series analysis for monitoring Brazilian agriculture. *ISPRS journal of photogrammetry and remote sensing 145* (2018), 328–339. https://doi.org/10.1016/j.isprsjprs.2018.08.007.

[40] PUTZ, F., AND REDFORD, K. The Importance of Defining 'Forest': Tropical Forest Degradation, Deforestation, Long-term Phase Shifts, and Further Transitions. *Biotropica 42*, 1 (2010), 10–20. https://doi.org/10.1111/j.1744-7429.2009.00567.x.

[41] SASAKI, N., AND PUTZ, F. Critical need for new definitions of "forest" and "forest degradation" in global climate change agreements. *Conservation Letters 2*, 5 (2009), 226–232. https://doi.org/10.1111/j.1755-263X.2009.00067.x.

[42] SIMONS, P., AND MELIA, J. Continuants and Occurrents. *Proceedings of the Aristotelian Society, Supplementary Volumes 74* (2000), 59–92. https://doi.org/10.1111/1467-8349.00063.

[43] SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A., MUNGALL, C. J., LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SCHEUERMANN, R. H., SHAH, N., WHETZEL,

P. L., AND LEWIS, S. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology 25*, 11 (Nov. 2007), 1251–1255. https://doi.org/10.1038/nbt1346.

[44] SMITH, B., AND MARK, D. M. Ontology and Geographic Kinds. In *Proceedings of the 8th International Symposium on Spatial Data Handling* (1998).

[45] SMITH, B., AND MARK, D. M. Geographical categories: An ontological investigation. *International Journal of Geographical Information Science 15*, 7 (Oct. 2001), 591–612. https://doi.org/10.1080/13658810110061199.

[46] SMITH, B., AND MARK, D. M. Do Mountains Exist? Towards an Ontology of Landforms. *Environment and Planning B: Planning and Design 30*, 3 (June 2003), 411–427. https://doi.org/10.1068/b12821.

[47] VENDLER, Z. Verbs and Times. *The Philosophical Review 66*, 2 (1957), 143–160. https://doi.org/10.2307/2182371.

[48] VERBESSELT, J., HYNDMAN, R., NEWNHAM, G., AND CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment 114*, 1 (2010), 106–115. https://doi.org/10.1016/j.rse.2009.08.014.

[49] WOODCOCK, C. E., LOVELAND, T. R., HEROLD, M., AND BAUER, M. E. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment 238* (Mar. 2020), 111558. https://doi.org/10.1016/j.rse.2019.111558.

[50] WORBOYS, M. Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science 19*, 1 (Jan. 2005), 1–28. https://doi.org/10.1080/13658810412331280167.

[51] ZENG, L., WARDLOW, B. D., XIANG, D., HU, S., AND LI, D. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment 237* (Feb. 2020), 111511. https://doi.org/10.1016/j.rse.2019.111511.

[52] ZHU, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing 130* (2017), 370–384. https://doi.org/10.1016/j.isprsjprs.2017.06.013.