

# On ‘ground’ truth and why we should abandon the term

Iain H. Woodhouse \*

University of Edinburgh, School of Geosciences, Edinburgh, Scotland, United Kingdom

**Abstract.** I make a case to abandon the use of the term “ground truth.” After a brief history of the term that indicates its military origins, two main arguments are presented against the use of the term. The first is from measurement theory. The second is by way of presenting three examples: the first is in crop mapping, in which “ground truth” may have some validity. This is presented as the exception to prove the rule. The second example looks at albedo, in which the measurand is the same for the local and remote measurements, and evidence is provided to support this paper’s argument. The third example looks at the scattering phase center in radar. In this case, the measurand is different for the local and remote measurements, and an argument is given for why this does not warrant the use of the term. Finally, a heuristic checklist that seeks to guide readers to reflect on whether or not their particular field measurements should be referred to as “ground truth” is provided. © 2021 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JRS.15.041501](https://doi.org/10.1117/1.JRS.15.041501)]

**Keywords:** remote sensing; Earth observation; ground truth; calibration; field measurements; in situ; validation.

Paper 210526V received Aug. 16, 2021; accepted for publication Oct. 27, 2021; published online Nov. 18, 2021.

## 1 Introduction

“Exactitude is not truth.” — Henri Matisse (1869–1954)

The term “ground truth” has been used in remote sensing for more than half a century. Since 1965, there have been ~6088 remote sensing articles that have used this expression in the title, abstract, or keywords. The term has also spread to other disciplines, with a further 27,348 articles using the term published across disciplines as diverse as computer science and engineering (22,302 combined), psychology (117), and neuroscience (598). Even articles in nursing (39), veterinary science (11), and economics (57) have taken to using “ground truth” as a meaningful expression.

In this paper, I aim to persuade fellow remote sensing scientists that it is now time for us to rethink the use of this term and nudge us toward retiring it entirely from our lexicon. And, perhaps, if it is abandoned in our community, we may even finally kill it off altogether.

The main thrust of my argument is that “ground truth” can distort our understanding of what remote sensing has achieved in the past and risks constraining our imagination of what it might achieve in the future. It primes us, and predisposes our students, to believe that the “right answer,” the most truthful one, is the one that is measured on the ground not the one measured from a distance. When we use the term “ground truth,” it does not allow for the possibility that the most valid measurement, the one that we are most interested in, is actually the one made from above.

### 1.1 Summary of the Argument

Remote sensing is a practical discipline intent on making progress through application. But it is exactly this pragmatism to which I aim to appeal in this paper. I wish to make the case that

---

\*Address all correspondence to Iain H. Woodhouse, [i.h.woodhouse@ed.ac.uk](mailto:i.h.woodhouse@ed.ac.uk)

“ground truth” as a term is misleading, disingenuous, and distorts newcomers’ understanding of the science of remote sensing. In using the term, we introduce new students to the subject in a way that risks curtailing their imagination and hinders the possibility of step changes in remote sensing science by narrowing their conceptualization of what a “remote measurement” entails. This is not (just) about being philosophically pedantic. It is about how we think about the fundamental principles of what we do when we make a remote measurement and then making sure that we communicate this effectively to those people who are just starting to learn how to use data from aircraft, drones, and satellites to measure planet Earth.

While this may be the first paper of its kind to focus on the term “ground truth” and argue for its demise, as this paper will note, it is certainly not the first to question the use of the term. It is also important to note that this paper aims to complement, rather than distract from, the incredible work carried out by the teams of people working on calibration and validation (cal/val) of satellite products, who constantly deal with the challenge of linking *in-situ* to remote measurements.<sup>1</sup>

I aim to demonstrate in the sections that follow that the term “ground truth” is indeed redundant—in fact, it was likely never appropriate, bar for a small number of unique measurement scenarios.

Remote sensing as a discipline may have originally introduced the term, but it is now time for this discipline to drop it.

## 1.2 Structure of the Paper

The paper is structured in three parts. The first is a brief resumé of the term, exploring its possible origins and examining its various meanings. Second, I set out an argument against the term from the position of measurement theory, in that trueness is a less tangible concept than the term “ground truth” would lead you to believe. In this sense, I argue that the term is outdated and not fit for purpose. I then present three case studies to support this argument: the first actually offers an example of when the term is appropriate—this is the exception that proves the rule. The other two are real remote sensing examples in which the use of the term is clearly invalid.

The third part is a heuristic device that provides a checklist of criteria by which to judge whether the term is valid or not for any given scenario. This is presented more in the spirit of provocation than enlightenment, with the aim of encouraging deeper reflection and critical appraisal of the term by remote sensing scientists before they simply default to its use.

## 1.3 Methods Used for the Review

In trying to establish the origin of the term “ground truth,” the following databases were searched for each combination of “ground truth,” “ground-truth,” and “groundtruth:”

- Web of Science/Web of Knowledge,<sup>2</sup>
- Scopus,<sup>3</sup>
- The New York Times, The U.S. Military Intelligence Reports (1911–1944) from ProQuest History Vault,<sup>4</sup>
- Records of the Central Intelligence Agency (CIA) Electronic Reading Room,<sup>5</sup>
- The Oxford English Dictionary (OED), and
- Google Scholar and Google Books.

## 2 Resumé of ‘Ground Truth’

### 2.1 Definition the First

The Oxford English Dictionary (OED)<sup>6</sup> gives three definitions of the term “ground truth,” The first is “a fundamental truth; the real or underlying facts; information that has been collected at source.” This is the oldest use of the term, with its earliest recorded use being cited as Henry Ellison’s (1811–1880) poem “The Siberian Exile’s Tale” in 1833 (contained within Ref. 7). The earliest recorded use of the term in this manner within a science text (that this author could

find) is in a *Science* article in 1896 by H. R. Greene, relating to fundamental principles standing or falling on the “ground truth or truths” on which they rest.<sup>8</sup> This is a very general use of the term, not related to measurement directly. The OED suggests that this term may originate from the German *Grundwahrheit*, meaning a basic or fundamental truth (first used around 1650, but less common in modern German, although still used in some contexts, e.g., digital humanities<sup>9</sup>).

A similar meaning is also found in 19th century religious texts, although in this context it refers to the truth as grounded on the Earth—the experience of “Man” [*sic.*] in the everyday world on the ground as opposed to the spiritual truth of God in heaven as evidenced by scripture.<sup>10</sup>

## 2.2 Definition the Second

The second OED definition is the one most pertinent to the current discussion: “In remote sensing: information obtained by direct measurement at ground level, rather than by interpretation of remotely obtained data (as aerial or satellite images, etc.), especially as used to verify or calibrate remotely obtained data. Also: the actual characteristics of a terrain being surveyed.” Similar definitions specifically linking it back to a remote sensing origin are found in, for example, *A Dictionary of Geography*:<sup>11</sup> “In remote sensing, ground truth refers to the comparison of remotely sensed data with real features and materials on the ground.”

Similarly, two short quotes from authors commenting from a biological sciences standpoint reiterate the case. First, from *Standards in Genomic Sciences*:<sup>12</sup>

“The concept of ground truth is a well-established principle in cartography, where data collected at a distance are confirmed by measurements made on location. Those local measurements are used to calibrate remote sensing devices, verify or correct experimental inferences, and update geographic databases. Ground truth observations also provide a means of training and supervising image classification software and resolving errors of omission or commission. Cartographic methods have improved significantly because of the development of precise positioning methods (GPS), the development of interoperable data standards for rapid exchange of precise and highly interlinked information, and the development of various devices and visualizations that serve-up information on-demand to different classes of end-users.”

And second, from an editorial in *Nature Methods*:<sup>13</sup>

“Researchers using satellite imaging to remotely observe features on the Earth enjoy the luxury of a simple solution for verifying the interpretation of their data with the truth on the ground or ‘ground truth.’ They or a surrogate can go observe it firsthand. In contrast, researchers using an algorithm to analyze data on complex biological phenomena rarely have the luxury of a straightforward ground truth.”

The exact point at which the term “ground truth” originated within the remote sensing lexicon seems elusive. The earliest published work in the civilian domain that uses the term in this way appeared in the 1960s. The OED cites the earliest as 1965, with a conference proceedings entry by R. H. Alexander in *Proceedings of the Third Symposium on Remote Sensing of Environment 1964* (but published in 1965): “Talking as space scientists we can say that it will be essential to get some direct observations at the planetary surface to establish ‘ground truth’ and calibration of the data obtained by the remote sensors.”<sup>14</sup>

Similarly, this author found a Texas Instruments technical report entitled, “Ground-Truth Research For An Airborne Multisensor Survey Program,” from 1966<sup>15</sup> and a flurry of peer-reviewed articles that started to appear in 1968 (e.g., Ref. 16). What is noteworthy is that these authors use the term “ground truth” in such an unremarkable way that it would seem clear that the term predates these publications by quite some time.

A search of the CIA archives online reveals, not unsurprisingly, that the term was in military intelligence as early as 1964 in the context of aerial reconnaissance. One report<sup>17</sup> dated February 20, 1964, for instance, focuses on using aerial surveillance for “suspected military build-up,” assessing “aerial targets” (by which they mean targets identified from the air) and evaluating a remote sensing system, the exact nature of which is still redacted but looks to be considering thermal imagery. Here, the term “ground truth” is even defined in a footnote, which would indicate that it was not yet sufficiently widely used that the term could go unremarked upon.

The footnote defines it as “Determining the actual state of terrestrial surface environment in support of airborne remote sensor operations.”<sup>17</sup>

However, another document (Ref. 18) dated even earlier (January 15 of the same year) uses the term so matter-of-factly that it as if it was in common use (offering no raw definition in the text).<sup>18</sup> Interestingly, a third document (Ref. 19) from February 10 also uses it without remark, but it does explain that “ground truth data . . . [is] complete, accurate, and fully documented data on the actual target conditions.”<sup>20</sup> So, it would seem for these authors that ground truth is not simply the *in situ* measurement of a specific metric but the entire gamut of conditions: hourly weather station report from 24 h prior to mission (including temperature, precipitation, wind, cloud cover, cloud levels, and dew point), flash photographs and description of target elements, airport logs, and a “full analysis of all exposed materials in each target area.”<sup>19</sup>

At the time of writing, this author could not find conclusive evidence of the original use of the term in a remote sensing context, but the path very clearly leads to military reconnaissance. With the earliest recorded use being within this field, and the military decision-making context having a focus on “what is happening on the ground,” where “on the ground” refers to a distant location where the action is taking place, it seems most likely that this is its origin. It would seem plausible that, given that military reconnaissance personnel have an express interest in the relationship between the reconnaissance and what is actually happening on the ground below, the term “ground truth” might emerge as a suitable term. [It would be understandable for the reader to also ask if this use of “on the ground” links to the expression “boots on the ground” (meaning military personnel physically located at the location of interest). However, while that expression might be expected to have a long-established history in military jargon history, it is much younger than “ground truth.” According to Matthew Seelinger, chief historian of the Army Historical Foundation (as reported in Ref. 19), the earliest documented use of the term “boots on the ground” does not seem to predate a 1980 edition of the *Christian Science Monitor*, in which General Volney F. Warner gave an interview concerning the Iranian Hostage Crisis. It does, however, confirm that military minds think of “the ground” as where the action is taking place.]

### 2.3 Definition the Third

The third OED definition is “Information obtained by direct observation of a real system, as opposed to a model or simulation; a set of data that is considered to be accurate and reliable, and is used to calibrate a model, algorithm, procedure, etc. Also: (specifically in image recognition technologies) information obtained by direct visual examination.” This is the extension of the original reconnaissance meaning, whereby a remote or simulated result is compared with the result “on the ground.” Such a definition amplifies the conceptual inference that “ground truth” is the “right answer” because in these studies, by design, it is the right answer.

It is also interesting to note how this last definition has also migrated to popular culture—see, for example, the 2009 book, “The Ground Truth: The Untold Story of America Under Attack on 9/11,” by John Farmer.<sup>21</sup> The title ingeniously plays with the terms “ground zero” and “evolving situation ‘on the ground’” (to mean what was happening at the point of action), while combining them with a conspiratorial hint that only the book has “the truth.”

### 2.4 Other Challenges to the Term

This article’s questioning of the appropriateness of the term “ground truth” is not new. Challenging the use of “ground truth” coincides with the early evolution of remote sensing making more physical, rather than cartographic, measurements. In a NASA report from as early as 1972, Roger M. Hoffer makes a critique of the term,<sup>22</sup> despite entitling his report “The Importance of ‘Ground Truth’ Data in Remote Sensing.” His use of quotation marks in the title is elaborated in the text:

“‘Ground truth’ involves the collection of measurements and observations about the type, size, condition and any other physical or chemical properties believed to be of importance concerning the materials on the earth’s surface that are being sensed remotely. Lately, the term

“ground truth” has fallen into disfavor, for several reasons. Sometimes, errors in data collection have caused ‘ground truth’ to be false data; in addition, there are often so many variables involved that one wonders what the “truth” of the situation really is. Also, if you are obtaining data through interpretation of large-scale photos collected from the air or if you are obtaining measurements of the temperature of a water body, should such data really be referred to as “ground truth”? Therefore, it seems more logical to call such procedures “surface observations” or some similar term to refer to the collection of data about the materials on or near the earth’s surface.”<sup>22</sup>

As our discipline grew, this perspective seems to have waned, but it has not disappeared. One example is from 2019 in the context of forest biomass mapping:<sup>23</sup> “field-based estimates—often improperly seen as ‘ground truth’—most likely convey more important errors than generally assumed.”

This current article might therefore be considered an attempt to reignite the critical perspectives that have been offered over the last half century and to formulate a coherent argument in one paper that might be cited whenever critique is warranted.

### 3 Wherein Lies the True Value?

#### 3.1 From Measurement Theory

A key argument against the term “ground truth” is based on validity within standard measurement theory, and in this section the term is examined more closely, with examples.

Clearly the use of the term “ground” in “ground truth” can be addressed quickly—it is clear that in a remote sensing context this need not refer to ground *per se* but is in reference to making a measurement *in situ*. This can be on the surface of the land, the depths of the oceans, or high in the atmosphere. In so far as it need not refer to “the ground” at all, this might immediately invalidate the use of the term; however, it seems to have stuck (and indeed, has been applied to all manner of nonground results as noted above).

The second term is perhaps more contentious. The term “truth” implies a realist ontology (that there is a value of the measurand, the thing being measured, that has a conceivable value independent of human knowledge or models) and equates to the term “trueness” in measurement theory. “Trueness” is well defined in measurement theory. The ISO 5725-1:1994<sup>24</sup> standard states that:

“The “trueness” of a measurement method is of interest when it is possible to conceive of a true value for the property being measured. Although, for some measurement methods, the true value cannot be known exactly, it may be possible to have an accepted reference value for the property being measured; for example, if suitable reference materials are available, or if the accepted reference value can be established by reference to another measurement method or by preparation of a known sample. The trueness of the measurement method can be investigated by comparing the accepted reference value with the level of the results given by the measurement method. Trueness is normally expressed in terms of bias.”<sup>24</sup>

It is the combination of “trueness” (the bias from the actual value) and precision (the randomness inherent in the measurement) that together formally describe the “accuracy” of the measurement.<sup>25</sup>

In remote sensing, therefore, the term “ground truth” should be considered to be merely a colloquialism used to describe “the accepted reference value” of the property being measured. In remote sensing, it is often the case that this “accepted reference value” is considered to be the “true” value. Since all measurements are an approximation, in practice, what this means is that the “accepted reference value” has been determined using some method that is considered to have much less uncertainty than the remote measurement. Or, perhaps it is more correct to say that what we mean by ground truth is the process of collecting measurements leading to a set of measurements known to be the most accurate using the current instruments.

The best estimate is usually, by definition, the expectation (the expected value or mean) of the distribution. The dispersion of the results are characterized by the positive square root of the variance, called the standard deviation (or standard uncertainty<sup>26</sup>).

“True value cannot be found by experimental means and is defined as the average of measured values derived from a sequence of repeated measurements. In contrast, measured value is a single measurement of an object that is intended to be as accurate as possible. The difference between these two measurement concepts is referred to as the error, or total error.”<sup>27</sup>

In some contexts, we might revisit whether using the term “ground truth” means “a sequence of repeated measurements” of low uncertainty, in contrast to the remote measurement, which is a single measurement. However, often field measurements are also just one value, but they use a method deemed to have lesser uncertainty.

In some fields, they explicitly prefer using the term “best available measurement” rather than “ground truth,” and this is perhaps a suitable alternative in remote sensing.<sup>27</sup>

It is worth noting here that the term “ground truth” significantly predates the change from Error Theory to Uncertainty Theory. In the last decade of the 20th century, the Theory of Error was gradually replaced by the Theory of Uncertainty,<sup>28</sup> and this had an influence on measurement theory, even if it did not noticeably impact remote sensing. An uncertainty-focused approach accepts that the measurand can never be known exactly but is only ever an idealized concept that is impossible to evaluate without uncertainty. In this context, you acknowledge the dispersion of the values attributed to the measurand and record a probability distribution function, rather than a single number.

So, measurement theory introduces two situations for remote sensing that are worth looking at in a little detail: one case in which the conceived value is the same for both the *in situ* and the remote measurement and another in which they are different. In both cases, we are assuming here that the value we are trying to measure is a measurand of which “it is possible to conceive of a true value.” While there are some conceptually challenging examples (what exactly is “soil moisture” of an agricultural field), most remote sensing parameters are sufficiently well defined so that they can be uniquely “conceived of.”

There are certainly some properties of the Earth system that can both be conceived of and be straightforwardly determined by *in situ* observation. Yet, there are also elusive properties that fail to be observed as direct measurements. In the following sections are three examples to illustrate the argument. The first case—the exception to prove the rule—is for classification of a discrete measurand. This is one scenario in which “ground truth” may well be considered an appropriate term. The second case looks at an example in which the conceived value is the same for both the *in situ* and the remote measurements. The third describes a scenario in which they are different. In both of these latter cases, the term is not valid.

### 3.2 Case 1: Measurands on the Nominal Scale

The least controversial use of the term ground truth is when you have a measurand that is on a nominal scale of measurement. The nominal scale only satisfies the identity property of measurement—that is, each value on the measurement scale has a unique, unambiguous meaning. Values on this scale correspond to variables that represent a descriptive category but have no inherent numerical value—they have a label but no magnitude.

In civilian remote sensing, this is best exemplified by crop mapping—the measurand is not a numeric value but membership of a discrete category, such as wheat, barley, potatoes, etc. In a military surveillance context, this might be types of targets, such as a tank, jeep, etc. The conceived true value is a category. For these kinds of examples, the measurand is genuinely on a nominal scale, and the targets in question are easily assigned into discrete classes.

Methods that use such a classification approach to characterize a target area that is actually on a continuous numeric (ratio) scale are more appropriately called thresholding, rather than classification. An example would be, say, classifying the boundaries between forests and woodland, which, even on the ground, can be poorly defined and difficult to measure in natural environments. This kind of problem uses a classification approach (a nominal scale of output classes) to draw arbitrary boundaries within an otherwise continuous variable. A mixed pixel or fuzzy logic approach would not address this issue since it still must deal with the mismatch between a continuous measurand that has a final data product based on discrete classes.

With the measurand on the nominal scale, the measurement problem can be rephrased as a Boolean logic question: The field is wheat: true or false? For well-defined discrete classes,



there is therefore a valid logic behind claiming to have “truth on the ground” because close inspection of a crop by a human observer is certainly the “best available measurement.” Likewise in a military surveillance context, the target is either a tank or it is not. Given the similarity of these two scenarios, it should not be surprising to note that the transition of “ground truth” from military to civilian use would have occurred just prior to the development of the Landsat program, a mission specifically designed to map agricultural production (supposedly to help the CIA to predict crop yields in the then Soviet Union). I conjecture that the term easily (and with validity) translated from target classification to crop classification and then subsequently migrated to other forms of remote measurement thereafter, despite the insightful protestations of scientists such as Roger Hoffer.<sup>22</sup> Classification of agricultural fields is often the first lesson that a remote sensing student will learn, but such students are rarely asked to unlearn it later, when the approach is no longer valid, as in the next two examples.

### 3.3 Case 2: Identical Measurand In Situ and Remote

The second case study is the measurement of albedo. Albedo determines what proportion of incident radiation is scattered back to space from a surface or object. As a measurand, it is well conceived and it lies on the ratio scale of measurement, meaning that it satisfies all four properties of measurement: identity (it is a well-conceived measurand), it is represented by a number (magnitude) rather than a label, the scale has equal intervals (units along the albedo scale are equal to one another), and the minimum value is zero. Having a true zero allows us to know how many times greater one value is compared with another. Having a zero albedo means that there is no albedo.

Measuring albedo across Earth’s surface is important because it is a fundamental climate variable—it influences the proportion of solar radiation that is absorbed as opposed to scattered back into space and therefore plays a key role in the Earth radiation budget. Measurement of albedo makes a good example for our purposes for several reasons. First, it is a well-defined measurand with a clear physical meaning, so “it is possible to conceive of a true value for the property being measured.” Second, it is an uncontroversial measurand in so far as it is a property of the direct measurement of radiation (coming and going) that our remote sensing instruments are primarily designed to measure. And third, both the field measurements and the remote measurements are designed to directly measure the albedo through detecting the upwelling radiance from the surface. This is in contrast to, say, a measurement of soil moisture or water turbidity, in which the principal measurements are different on the ground compared with the ones from a distance (as in the third case study below).

To support the argument of this paper, let us consider the work by Ryan et al.,<sup>29</sup> whereby MODIS measurements of albedo are compared with *in situ* albedo measurements recorded by automatic weather stations (AWS) across Greenland. The major source of uncertainty in the remote measurements is the impact of the atmosphere, so it makes sense to use *in situ* measurements to test and evaluate the methods used to remove the atmospheric component. The AWS pyranometers (radiometers optimized for measuring solar irradiance) measure downward and upward shortwave radiation fluxes (with an uncertainty of <5%) from the ground surface at a distance of 2.8 m (snow height not considered).<sup>29</sup> Since these instruments have a field of view of 150 deg, they have a maximum ground footprint diameter of 21 m, equating to an area of 346 m<sup>2</sup> (although in practice the effective footprint is smaller since the radiometers’ cosine response means that their sensitivity is not uniform). Ryan et al. considered how these measurements correspond to MODIS measurements using the MODIS daily albedo (300 to 3000 nm) product, MOD10A1 collected by NASA’s Terra satellite, with a pixel size of 463 m (circa 214,000 m<sup>2</sup>). The key assumptions in using these AWS measurements as “ground truth” are that they meet the following criteria: (1) the pyranometer is more accurate than the satellite measurements (best available measurement) and (2) the pyranometer measurement is representative of what the satellite measures. If you are measuring the same measurand both *in situ* and remotely, from the point of view of probability theory, the measurements occur in a different “reasoning environment”—what Estler refers to as an “apples and oranges” comparison.<sup>26</sup> Although the measurand is the same (radiation), the methods differ. This introduces a component of uncertainty that requires careful attention to detail.

In the case of Ryan et al., their “attention to detail” was to test condition (2) using a fixed wing drone with a camera to make a detailed “point to pixel” comparison and evaluate the impact of the heterogeneity of the albedo across a single MODIS pixel. Their study exposed a fundamental flaw in the matching of *in situ* to satellite measurements: the issue of scale. Even though the assumption that the local measurement was “the best available” answer (criterion 1, above), they found that the *in-situ* measurements were not representative—the albedo was overestimated by the AWS pyranometers because topography and larger scale variation reduced the km-scale albedo with an overestimation of up to 10%. The significance of this result to the current argument is summarized by the authors:

“... because the *in situ* measurement is assumed to be accurate, and indeed, is often considered a “ground truth,” discrepancies are frequently attributed to bias in the satellite-derived albedo product [...]. This results in loss of confidence in satellite-derived albedo retrieval due to incorrect error attribution, thereby diminishing the statistical significance of long-term albedo trends and diluting capacity to accurately monitor the Earth’s cryosphere.”<sup>29</sup>

The pyranometer footprints are insufficiently large to capture the spatial heterogeneity within the single pixel. In this case, if the ambition is to measure the impact of albedo over a large scale—for instance, for use in regional or global climate system models—the remote measurements are more reliable than the supposed “ground truth.” It is important to stress that this is not to diminish the value of *in situ* measurements in improving, calibrating, and correcting satellite data, but rather it demonstrates that these authors also recognize that the term ground truth implies a level of confidence in the measurements that may not be justified—not because the measurement is inaccurate, but because it does not actually measure the parameter one is trying to measure.

With similar reasoning, one might also examine the measurements of many other ratio scale measurands wherein both the *in situ* and the remote measurements are measuring the same physical parameters but at such different scales that it is the remote measurement that is actually capturing the metric one wishes to know. This might include spectroradiometer measurements over land or sea surface temperature measurements. In the latter case, the broad area measurement is what is of interest not the point sample. Of course, it should be noted that the use of near-range radiometers from ships will not suffer as much as the AWS measurements in Ryan et al. because they benefit from a correlation length of surface temperature that is sufficiently long as to be much larger than the pixel size, in most cases.

### 3.4 Case 3: Different Measurands In Situ and Remote

The third case study addresses the situation in which two different measurands are being measured—one *in situ* and one remotely (the differences between temporal and physical dimensions being addressed in Sec. 3). This is perhaps the most common scenario in satellite remote sensing because in many applications we are trying to determine the value of a physical or biophysical measurand on (or near) the surface using electromagnetic radiation alone. Measurands such as leaf area index (LAI), NO<sub>x</sub> concentration, soil moisture content, above ground biomass, and snow equivalent water content are all on the ratio scale and can all be measured *in situ* by some kind of direct measure (the “best available measurement”). Yet, in remote sensing we aim to determine the value of these measurands by indirectly measuring the properties of electromagnetic radiation. Both measurements may be accurate and valid, yet much of remote sensing is trying to correlate one to the other, either through a statistical or a physical model. In these cases, the issue with the term “ground truth” is that it puts greater emphasis on the quality of the *in situ* measurement, without any consideration given to the nature of the remote measurement, which may be equally “true” but is simply measuring a different measurand.

Here is an example from my own field of research to illustrate the potential problems with using “ground truth” too loosely.

Synthetic aperture radar used in a single-pass interferometric configuration provides an estimate of what is called the scattering phase center (or scattering phase height, SPH).<sup>30</sup> Over unvegetated surfaces (with a sufficiently high dielectric constant to ensure that scattering is from the surface), the SPH will correspond to the actual height of the surface (above some reference surface) to within absolute accuracies of less than a meter and relative height accuracies in the



region of centimeters or less (i.e., on a scale of the wavelength), even when measured from space.<sup>28</sup> Since we are only talking about height, it is clearly “possible to conceive of a true value for the property being measured.” The measurand of the SPH is a consequence of determining the aggregate phase of all of the signals returning from the radar range and azimuth location (what we might refer to as “the pixel”). The SPH is an extremely well-defined, unambiguous measurand, and the radar measurement is the only way to measure this parameter.

Height, as a measurand, lies on the ratio scale. The radar provides the “best available measurement” of the SPH, and even when uncertainties in the phase measurements at the instrument are taken into account, it is still the *best measurement* you will get of the SPH. That the SPH does not always correspond to what might be measured on the surface is neither here nor there at this point, as the SPH is a measurand in its own right. For most applications in topographic mapping, the correspondence to the actual ground height as measured by other means on the ground is reassuring, but this would not be considered a “calibration” of the SPH (other than to convert relative heights to absolute heights, but that is not a requirement, *per se*).

If we then consider the case of vegetated surfaces, SPH has a different role. Over forest areas in particular, SPH has been used effectively to determine the height of forest canopies (e.g., Ref. 30). The SPH is usually presented as the inferior measurement of canopy height, with the *in situ* measurements using instruments such as terrestrial laser scanners or hypsometers being considered the “ground truth.”<sup>31</sup>

Let us unravel this situation.

First, the implication of the language is that, when there are mismatches between *in situ* measurements and radar measurements, the “ground truth” wins over the SPH. But “canopy height” cannot actually be measured directly from the ground—the height of individual trees (or other metrics of individual trees) are what is measured. These measurements are then collated to represent some value of “canopy height.” There are at least seven different ways to characterize a “canopy height” (which we take to be defined as an aggregate value that represents some average across a population of trees), and new methods continue to be proposed.<sup>32</sup> These include the following:

- Arithmetic mean height: the arithmetic average of all trees (or more usually, all trees above a certain height or greater than a certain diameter).
- Weighted average (i.e., Lorey’s mean height): the average height of all trees (above a threshold size) weighted by their basal area. (Lorey’s mean height is often used as the ground truth of mean tree height of forest stands in remote-sensing studies.)<sup>32</sup>
- Dominant height: the arithmetic mean of the dominant (tallest) trees.
- Top height, H100: a particular kind of dominant height that uses the mean height of the 100 tallest trees per hectare (although in practice, the height of the 100 trees with the largest diameter at breast height).
- Dominant and codominant height: the arithmetic mean of the tallest and second tallest trees.
- Mean timber height: the arithmetic mean of the timber height of all trees or all trees above a threshold).
- Form height: the product of the mean height and the form factor (using top height as the mean).

Reference 32 summarizes the problem: “. . . the differences of the measures between different definitions can sometimes reach several meters even if they are for the same stand.” The conclusion is that canopy height is variable and ill-defined, but SPH has only one, well-defined definition.

SPH is dependent on many of the other key parameters related to vegetation structural complexity: number density of trees, basal area, LAI, stand density, horizontal and vertical heterogeneity, tree architecture, and canopy closure.<sup>33</sup> Perhaps the approach here should not be to statistically correlate SPH to field-based measurements but to understand exactly what the SPH is revealing about the collective three-dimensional distribution of material in a forest canopy. If the interest is in structural complexity, canopy density, and light penetration, then perhaps SPH is a better-quality measurement than the measurements on the ground. It is certainly worth questioning whether canopy height is actually “a true value” that is being conceived or is in fact

a multitude of conceived ideas, none of which necessarily captures the physical structural properties of the forest under observation. Meanwhile, SPH is a very precise, unambiguous, repeatable, and very accurate measure of the combined contribution from the scattering elements in the canopy.

Similar arguments might be made, for example, for soil moisture (in which field measurement themselves may be inconsistent, such as in Ref. 34), forest biomass (in which there is disagreement in the direct nature of remote measurements<sup>35,36</sup> as well as field measurements),<sup>37</sup> or LAI (in which there are various direct and indirect approaches to measuring LAI).<sup>38</sup> But this is not the place to offer an exhaustive list. The purpose here is to prompt readers to reflect on their own measurands of interest in their own field and decide for themselves whether “ground truth” is meaningful for their observation.

## 4 Heuristic Checklist

This section aims to provide a practical starting point to consider whether “ground truth” is an appropriate term to use in any particular case. Presented here are four features of the measurement that should be reflected upon before using the term.

### 4.1 Measurement Type

The proposed argument is that the only legitimate case for using the term “ground truth” is when the measurand of interest is on the nominal scale—i.e., a labeled class, the membership of which is either true or false, as in case 1 above. When the measurand does not comply with this condition, the term is not appropriate. The important feature here is being specific about the measurand. For instance, in the case of crop mapping, the measurand is not a map of the ground features (which incurs complications such as mixed-pixels or ambiguities of field boundaries) but is only concerned about classifying the crop type per field.

### 4.2 Timeliness

The simultaneity of the *in situ* and remote measurement has a marked impact on the appropriateness of the term. In an ideal field campaign, the time difference between field collection and remote measurement is zero. However, this is not always the case, so consideration here must include not only the time difference between the measurements but also the correlation length in times of changes in the target property (i.e., how often the measurand changes and at what rate). When the time between measurements is much greater than the correlation length, then the simultaneity is lost, the value of the *in situ* measurements is minimized, and the “ground truth” is no longer a valid expression. (The correlation length is a metric derived from the autocorrelation of a sequence of numbers. This might be a sequence in time or in space. Its detailed derivation can be found in Ref. 39.)

### 4.3 Spatial Congruence

As with time, we must also consider the spatial closeness of the measurements. This must consider factors of both proximity and scale. Proximity refers to how close the location of the field measurement is to the remote measurement. This time, it is the spatial correlation length of the measurand that must be considered—the important thing is that the coincidence of the *in situ* and remote measurements must be much smaller than the correlation length.

Scale is an important factor that is often incorporated into discussions when considering the extreme example of trying to compare an *in situ* point measurement with an area-averaged remote measurement. Again, the correlation length is important here: is there a measurand that can be conceived of being measured and is the variability appropriately accounted for? For point measurements, it is often the case that several are measured within a certain area. To validly claim “ground truth,” the point measurements would have to meet the Nyquist sampling criteria for the correlation length of the measurand. (The Nyquist theorem states that the highest

frequency that can be represented accurately is one half of the sampling rate.<sup>40</sup> In spatial terms, the spatial frequency is the wavelength or the correlation length. This then sets a theoretical limit to the frequencies that can be detected with a given sampling rate. A grid of point measurements in a transect that are spaced 100 m apart, for example, could not detect patterns with correlation lengths <200 m.)

Note that spatial congruence and scale apply vertically as well as horizontally.

#### 4.4 *Strength of the Model between Different Measurands*

Remote sensing often does not measure the same physical property as that being measured *in situ*. This is the issue of case 3 above. There may well be a correlation between a good quality measurement of some physical property (reflectance, brightness temperature, normalized radar cross section, etc.) and another good quality measurement of a physical property measured on the ground (LAI, soil moisture, etc.). Each measurement is “truthful,” and it is not always the case that the ground measurement is the unique, well-defined parameter that is actually required for the question at hand. For instance, the normalized radar cross section value that is recorded in a pixel of an SAR image is a well-calibrated measurement of a genuine physical property. But it is usually a different physical property that we wish to retrieve: soil moisture, biomass, surface roughness, or temperature, for example. While these ground-based measurements may be well conceived and well measured, they may lack uniqueness, or they may not be the parameter that is actually required. A key question therefore is how uncertain is our knowledge (usually expressed through a physical model) of the link between the two measurands and does it make sense to try to correlate them?

In the case of SAR, the best available measurement of NRCS is the radar data, not any measurement on the ground.<sup>30</sup> Researchers are often simply correlating two “best available measurements” of two separate measurands, linked in some way via a model (either statistical or theoretical). To imply that one of these is more truthful than the other is misleading. You are merely formulating a statistical correlation between the two. This is often referred to as “calibration,” a term in itself which is misleading since remote data are often extremely well calibrated (in the strict sense) already.<sup>1</sup> The term calibration suggests matching an uncertain measurement to a more certain one.

It may be argued by some that this is the very condition under which many readers currently use the term “ground truth”—it is the name given to the truthful measurement on the ground, rather than the truthful measurement from afar. Even if this is the case, it remains misleading to use the term.

To supplement these heuristics, it is also appropriate here to suggest some alternative words or phrases to replace “ground truth.” Hoffer<sup>21</sup> offers “surface observations,” with “field measurement” or “*in situ* measurement” being common alternatives. “Validation measurement,” “reference measurement,” or “best available measurement” are all suitable alternatives. It is left to the reader, then, to decide which is most appropriate for their task at hand.

## 5 Discussion

This article has summarized the history of the term “ground truth” and discussed when it might be considered appropriate in remote sensing. It offers a heuristic checklist to decide when the measurement approach may allow for the use of the term.

In essence, this paper tries to promote an uncertainty theory approach (sometimes referred to as a Bayesian approach or a probability theory approach) to the environmental measurement techniques that we collectively refer to as “remote sensing.” The value of such an approach is that it emphasizes that the measurand can never be known with certainty and that this applies to both the *in situ* and remote measurement. In the traditional framework, the *in situ* measurement is assumed to be the one with the least uncertainty. However, described here are two examples in which this is clearly not the case and arguments for why the comparative “truthfulness” of *in situ* versus remote measurement are dependent on the context and purpose of the measurement.

Furthermore, it is important to recognize that at the heart of the use of “ground truth” as a framework for conceptualizing remote sensing lies an (often ignored) assumption that the human scale is the *de facto* scale by which we should interrogate the nature of our planet. While *in situ* measurements do not use anthropic units *per se* (units of measure based on parts of the body), we cannot help but make these measurements at an anthropic scale. The use of the term “ground truth” ignores the implications of this assumption.

There are some disciplines that have engaged with this issue. In ecology, for example, this is both fundamental and problematic. In his influential paper from 1992 entitled, “The Problem of Pattern and Scale in Ecology,” Simon Levin remarks, “The observer imposes a perceptual bias, a filter through which the system is viewed.”<sup>41</sup> With microscopes and environmental DNA,<sup>42</sup> ecologists recognize that understanding any system requires it to be studied at the appropriate scale, and, while not entirely new, the vision of a “macroscope” in support of macroecology appears to be gaining some interest.<sup>43</sup>

The goal of satellite remote sensing should be to reduce uncertainties on the macroscopic parameters not the human scale parameters that we too easily accept as “truth.” Perhaps if we renamed our discipline “macroscopics” instead of remote sensing, we would create a generation of new practitioners who reach beyond the anthropic scale and embrace the value of the high-quality measurements that we can measure at a distance across the entire planet. (It should not go unnoticed that atmospheric scientists, particularly those looking at the upper atmosphere, and astronomers, would argue that they have been doing exactly this for generations.)

## 6 Conclusion

It would be easy to dismiss objection to the term “ground truth” on the basis that it is simply a convenient shorthand and everybody knows it is not really “the truth.” But language can define, and constrain, how we think about what we do. Just as instrumentalists are right to make us reflect on how our choice of measurement technique influences our results, so too do the words we choose. Ground truth is prescriptive—it *a priori* defines the ground measurements as the best measurement. It leaves no room for doubt and so diverts us from engaging with the fact that it is the macroscopics that are the best available measure for the properties that we are ultimately trying to determine.

This paper presented “ground truth” in its wider context, explaining its origins (in so far as they can be determined with certitude) and reviewing the way it has been used in different remote sensing applications. While there are occasions when the term is appropriate, it is clearly not the best term to use for most satellite measurements. While the aim here is ultimately to stimulate reflection and debate, not to impose rigid thinking, I cannot help but conclude that it is time to retire “ground truth.”

## Acknowledgments

The author has no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose. The author wishes to thank the anonymous reviewers for their helpful suggestions.

## References

1. S. Sterckx et al., “Towards a European Cal/Val service for earth observation,” *Int. J. Remote Sens.* **41**(12), 4496–4511 (2020).
2. [www.webofknowledge.com](http://www.webofknowledge.com).
3. [www.scopus.com](http://www.scopus.com).
4. [www.proquest.com](http://www.proquest.com).
5. <https://www.cia.gov/readingroom/>.
6. OED Online, “ground truth, n,” Oxford University Press, 2021, <https://www.oed.com/view/Entry/249130> (accessed 6 August 2021).

7. H. Ellison, *Mad Moments: Or First Verse Attempts by a Born Natural*, 1st ed., Vol. 2, Malta, (1833).
8. E. L. Greene, “Some fundamentals of nomenclature,” *Science* **3**(53), 13–16 (1896).
9. M. Boenig et al., “Ground Truth: Grundwahrheit oder Ad-Hoc-Lösung? Wo stehen die Digital Humanities?” Abstract zur Konferenz Digital Humanities im deutschsprachigen Raum (2018).
10. H. R. Reynold, *Notes of a Christian Life: A Selection of Sermons*, Macmillan and Co., London (1865).
11. S. Mayhew, *A Dictionary of Geography*, 5th ed., Oxford University Press, Oxford (2015).
12. G. M. Garrity, “Ground truth,” *Stand Genom. Sci.* **1**, 91–92 (2009).
13. “Ground-truth data cannot do it alone,” *Nat. Method* **8**(11), 885, (2011).
14. R. H. Alexander, “Geographic research potential of earth satellites,” in *Proc. Third Symp. Remote Sens. Environ.*, University of Michigan, pp. 453–459 (1965).
15. R. Ludlum, “Ground-truth research for an airborne multisensor survey program,” AD0641894, Texas Instruments Inc., Dallas, Texas (1966).
16. J. E. Wilson, “Ground truth procedures for aiding interpretation of remote sensor data,” *Photogramm. Eng.* **34**(9), 992 (1968).
17. CIA, “Proposal No. 112-GD64 A&B,” Sanitized copy approved for release 2011/04/27: CIA-RDP78B04747A001000020031-2 (1964).
18. CIA, “Multi-sensor imagery research (around-the-clock aerial surveillance of levels in a suspected military build-up),” Sanitized copy approved for release 2005/12/23 CIA-RDP78B04747A0010000020033-0 (1964).
19. CIA, “Engineering passes and aerial targets,” Approved for release 2005/12/23: CIA-RDP79B00314A000500040008-5 (1964).
20. W. Safire, “On language: let’s do this,” *N.Y. Times* (7 November 2008).
21. J. Farmer, *The Ground Truth: The Untold Story of America Under Attack on 9/11*, Riverhead Books, New York (2009).
22. R. M. Hoffer, “The importance of ‘ground truth’ data in remote sensing,” Notes from UN Panel Meeting on the Estab. and Implementation of Res. Programs in Remote Sensing, NASA Technical Reports, Purdue University (1972).
23. M. Réjou-Méchain et al., “Upscaling forest biomass from field to satellite measurements: sources of errors and ways to reduce them,” *Surv. Geophys.* **40**, 881–911 (2019).
24. “Accuracy (trueness and precision) of measurement methods and results—part 1: general principles and definitions,” ISO 5725-1:1994.
25. A. Menditto, M. Patriarca, and B. Magnusson, “Understanding the meaning of accuracy, trueness and precision,” *Accred Qual Assur.* **12**, 45–47 (2007).
26. W. T. Estler, “Measurement as inference: fundamental ideas,” *CIRP Ann.* **48**(2), 611–631 (1999).
27. D. Yan et al., “Validation and ground truths,” in *Exploring Occupant Behavior in Buildings*, A. Wagner, W. O’Brien, and B. Dong, Eds., pp. 394–428, iBooks, Springer International, Cham (2018).
28. S. Salicone, *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence*, p. 228, Springer Science & Business Media, New York (2007).
29. J. C. Ryan et al., “How robust are in situ observations for validating satellite-derived albedo over the dark zone of the Greenland Ice Sheet?” *Geophys. Res. Lett.* **44**, 6218–6225 (2017).
30. J. M. Kellndorfer et al., “Vegetation height estimation from shuttle radar topography mission and national elevation datasets,” *Remote Sens. Environ.* **93**, 339–358 (2004).
31. M. Schmitt and X. X. Zhu, “Demonstration of single-pass millimeterwave SAR tomography for forest volumes,” *IEEE Geosci. Remote Sens. Lett.* **13**(2), 202–206 (2016).
32. T. Nakai et al., “A comparison between various definitions of forest stand height and aerodynamic canopy height,” *Agric. For. Meteorol.* **150**(9), 1225–1233 (2010).
33. D. I. Forrester, “Linking forest growth with stand structure: tree size inequality, tree growth or resource partitioning and the asymmetry of competition,” *For. Ecol. Manage.* **447**, 139–157 (2019).



34. G. C. Heathman et al., “Field scale spatiotemporal analysis of surface soil moisture for evaluating point-scale in situ networks,” *Geoderma* **170**, 195–205 (2012).
35. I. Woodhouse et al., “Radar backscatter is not a ‘direct measure’ of forest biomass,” *Nat. Clim. Change* **2**, 556–557 (2012).
36. S. Saatchi et al., “Forest biomass and the science of inventory from space,” *Nat. Clim. Change* **2**, 826–827 (2012).
37. D. B. Clark and J. R. Kellner, “Tropical forest biomass estimation and the fallacy of misplaced concreteness,” *J. Veg. Sci.* **23**, 1191–1196 (2012).
38. N. J. J. Bréda, “Ground-based measurements of leaf area index: a review of methods, instruments and current controversies,” *J. Exp. Bot.* **54**(392), 2403–2417 (2003).
39. G. Franceschetti and D. Riccio, “Surface classical models,” Chapter 2 in *Scattering, Natural Surfaces, and Fractals*, G. Franceschetti and D. Riccio, Eds., pp. 21–59, Academic Press, Amsterdam, Netherlands (2007).
40. R. Oshana, “Overview of digital signal processing algorithms,” Chapter 4 in *Embedded Technology, DSP Software Development Techniques for Embedded and Real-Time Systems*, R. Oshana, Ed., pp. 59–121, Newnes, Newton, Massachusetts (2006).
41. S. A. Levin, “The problem of pattern and scale in ecology: the Robert H. MacArthur Award Lecture,” *Ecology* **73**, 1943–1967 (1992).
42. P. Taberlet et al., “Environmental DNA,” *Mol. Ecol.* **21**, 1789–1793 (2012).
43. M. Dornelas et al., “Towards a macroscope: leveraging technology to transform the breadth, scale and resolution of macroecological data,” *Global Ecol. Biogeogr.* **28**, 1937–1948 (2019).

**Iain Woodhouse** is a professor of applied earth observation at the University of Edinburgh. His research focuses on the use of radar and lidar for mapping forests, and he is author of the textbook, *Introduction to Microwave Remote Sensing*. He is passionate about getting Earth observation into the mainstream, including getting involved in training worldwide and helping to start three satellite data analytics companies in and around Edinburgh: Ecometrica, Carbomap, and Earth Blox.