

# **Seismic Event Location Accuracy Codes: We Can't Handle the (Ground) Truth**

Eric Bergman  
Global Seismological Services  
bergman@seismo.com

September 28, 2013

## **Prologue**

*This article deals with the classification system (commonly known as the “GTx” system) currently in wide use to characterize the accuracy of hypocenters of seismic events, and proposes an improved nomenclature. This obscure topic has important consequences for a wide range of seismological research, as it strongly influences the selection and usage of seismic sources for investigations aimed at resolving Earth structure, addressing seismotectonic questions, and reducing earthquake hazard.*

*The ideas presented below are based on my own research experience in the area of earthquake location accuracy. My opinions on the matter have been strongly influenced by conversations and collaborations with other researchers in this field but the specific recommendations given here are strictly my own. They are inevitably colored by my own approach to the research and I will not be surprised to learn that I have not anticipated someone else's dire need or heartfelt belief. My primary purpose in writing this article has been to clarify my own thinking and to devise a system of nomenclature for my own use, but I will be very willing to consider adjustments to better meet the needs of other researchers, whether they be producers or consumers of calibrated seismic locations.*

*I encourage the community, e.g., the IASPEI Commission on Seismological Observation and Interpretation, to consider adopting a system of nomenclature that addresses the issues raised here. This article may be freely circulated and comments are welcome.*

## **From Ground Truth to GT25**

The terminology “ground truth” as it relates to seismology goes back at least to the late 1980s, in the context of developing a monitoring system for a proposed Comprehensive Test Ban Treaty (CTBT). The International Monitoring System (IMS) which has been developed to support the CTBT utilizes multiple technologies, but seismological analysis is a major component of the system, as it is for all national programs for nuclear monitoring. In recent years the concept of ground truth data sets has moved out of the nuclear monitoring community into the wider seismological research community. Although it has been used for several aspects of seismological research, the concept of ground truth is most often used in the context of event

location, as a way to solve the “chicken and egg” problem of seismology: an accurate location cannot be determined without an accurate velocity model of the Earth, and an accurate velocity model cannot be determined without accurate source locations. Although the terminology “ground truth” was apparently not in use at the time, this approach was employed in Herrin (1968) in developing a P-wave travel time model for global earthquake location using arrival time data from nuclear tests.

Thus, ground truth events came to be thought of as seismic events for which all hypocentral parameters (origin time, epicenter, focal depth) are known with exceptional accuracy, i.e., having zero bias at the scales of interest to the relevant community, which I take here to be the community of earthquake seismologists, which largely includes the community of seismologists working on nuclear monitoring. In practice, the set of ground truth events is dominated by man-made explosions, especially nuclear tests, with the complication that this information has not always been shared by testing states. Such events are of exceptional value in research, and so there is great pressure (i.e., funding) to extend the geographic coverage of ground truth events beyond the, thankfully, limited number of nuclear test sites.

One way to do this is to relax the requirement that hypocentral parameters must be known with essentially zero uncertainty. A second approach loosens the requirement that all hypocentral parameters be known to equivalently small uncertainty. In between true ground truth events and events with significant, unknown errors in all hypocentral parameters (most seismic catalogs, for example) there is a class of events whose hypocentral parameters (or at least some of those parameters) can be arguably determined with uncertainties that are usefully small, if not zero.

This approach was reinforced by the needs of the portion of the CTBT community that deals with on-site inspections. Such inspections must be limited in geographical extent, (e.g., 1000 km<sup>2</sup>) so the monitoring system must be able to pinpoint suspected violations of the treaty to within small, quantifiable limits. As a result the concept of “GT” events with significantly larger uncertainties in their locations than traditional ground truth events became entrenched in the monitoring research community, most notably in the form of the classification “GT5”. GT5 is taken to mean that the epicenter of an event is known within 5 km, and it has assumed the status of a litmus test in the earthquake location business. If an event qualifies as GT5 it’s of interest but there is often little effort made to try to say how much better than GT5 it might be. Nor is there much interest in events that fall just short of achieving GT5 status.

Traditional ground truth events, mainly nuclear tests and large chemical explosions, meet much tougher standards than GT5, of course, but everyone knows that that data set is unlikely to expand much. They are usually categorized as GT0, although GT1 or GT2 are sometimes encountered. The EHB global catalog of earthquakes (Engdahl et al., 1998), which has been analyzed specifically to eliminate the worst location bias, is thought to have an average epicentral accuracy level of about 15 km, except for subduction zones, where location bias can be considerably larger, so the GT5 standard represents a considerable challenge to seismologists. There are several important points to be made about GT5:

- Unlike GT0, where hypocentral parameters are expected to be known *a priori*, the designation of GT5 involves some kind of estimation or regression process, which always carries uncertainties. Therefore it should carry an indication of the level of uncertainty, such as GT<sub>90</sub>5. This is often neglected in practice. The need for the introduction of such uncertainties to the once rather pure concept of “ground truth” is a huge, ugly camel’s nose under the tent.
- The uncertainty of an epicenter is normally treated, not as a circle, but as an ellipse, derived from a covariance matrix in a least-squares analysis, although grid-search methodologies can yield even more irregular shapes. There is no standard on how the single scale length attached to a “GT” designation should relate to that ellipse. Should it be the semi-major axis? The semi-minor axis? The average?
- The concept of GT5 is easily extended to other scale lengths (“GTX”), and thus it has been extended, to as far as GT25. A cynic may wonder if this process has been encouraged by the desire to improve the apparent global coverage of “GT Databases” in Powerpoint presentations. If the EHB catalog (Engdahl et al., 1998) can be reasonably thought of as GT15 or so, what does GT25 mean? How far from the original concept of ground truth are we prepared to go?
- GTX classifications deal only with epicenters and have nothing to say about the accuracy and uncertainties of focal depth or origin time. For the narrow application of guiding on-site inspections under a CTBT, that is all that is required. For many other important applications, such as developing improved crustal models, it is a crucial omission, and modelers are prone to assuming the best about their data sets.
- The determination of meaningful confidence ellipses, or other measures of location uncertainty, is critically dependent on correct knowledge of the uncertainties in the individual arrival time data. The gap between original ground truth and GTX is further widened by the degree (extensive in most cases) to which location algorithms fall short of this standard.

## **The Unfortunate Influence of the Bondar et al. Criteria**

Bondar et al (2004), on which I am a co-author, has had the unfortunate effect of exacerbating these issues by presenting what many seismologists have taken to be an easy way to discover ground truth events, using nothing more than network geometry criteria. I’m quite sure that each of the authors has a somewhat different opinion on the matter of how the results reported in that paper should be used, but there always seemed to be good agreement among us that we were trying to come up with some convenient rule-of-thumb guidelines for estimating epicentral accuracy from seismic bulletins (i.e., containing the arrival time data as well as hypocentral parameters) that normally carry no useful estimate of that property, and in particular those estimates would attempt to account for location bias as well as random uncertainty.

To my mind the criteria were most suitable for use as a screening mechanism to select events which could be considered as good candidates for further analysis to determine location accuracy

at levels (with 5 km as a target, but always trying for better) relevant in research that depends on seismic sources with small and well-characterized uncertainties. I view it as extremely unfortunate that we did not propose a different nomenclature than GTX for classifying events processed with network geometry criteria. The information yielded by a network geometry analysis can be quite useful but it should not be co-mingled with either legitimate ground truth data, which are *a priori*, rare and precious, or with the results of detailed relocation analyses carried out specifically to determine locations with the greatest possible accuracy for as many hypocentral parameters as possible.

## **Beyond GTX**

Based on the above considerations I believe the GTX formulation has become an impediment to seismological research and should be abandoned. Therefore I have attempted to design a new nomenclature that more clearly and accurately reflects what is actually known about a hypocenter. These categories and criteria are based on the current practices and capabilities in seismic source location and earth model construction in the nuclear monitoring and seismotectonic research community. These practices and standards will undoubtedly evolve. The categories are not intended to obviate the need for full specification of uncertainties in hypocentral parameters, only to provide a fairly convenient way to categorize seismic sources in ways that are of importance in research, and especially, to prevent the inadvertent assumption that some parameters are known more accurately than they really are.

Of necessity a new nomenclature must be more complex than the GTX formulation. The endpoint of more complexity is a full description of the location methodology and data analysis, especially regarding error analysis, and the means taken to reduce bias in all parameters. There is clearly no practical way to reflect that information in a simple nomenclature. Given the variety of possibilities that deserve to be distinguished I have found that eleven categories are needed to adequately distinguish the different levels of knowledge concerning location accuracy that exist in today's research environment. Eight of these categories deal with events that would currently be lumped under the GTX formulation. The other three categories deal with uncalibrated events, such that all events in any seismic catalog can be assigned a category that clarifies their status with regard to location accuracy.

A fundamental goal of the new nomenclature system is to re-establish the distinction between traditional ground truth and what are more accurately termed "calibrated" data sets. It is therefore helpful to think of the eleven categories as members of four classes: the ground truth class (two categories), the calibrated class (four categories), the network geometry criteria class (2 categories) and the uncalibrated class (three categories). In my usage, calibrated data sets are the result of location analyses that aggressively seek to minimize bias in one or more (but as many as possible) hypocentral parameters. The degree to which this can be done depends on the available data and is therefore quite variable. Researchers who would use calibrated data sets as input for other investigations, such as in tomography, should be certain they understand the nature of their source data with regard to location accuracy. Seismologists who generate such

data sets should assist those researchers by using a well-designed nomenclature to characterize the accuracy of their results.

## **Ground Truth, the GT Class**

The GT0 nomenclature is reserved for what I have termed legitimate (or original or traditional) ground truth, events for which all four hypocenter coordinates are known *a priori* at levels of accuracy which are negligible for the purpose at hand, which is taken to be among the purposes of most current earthquake seismology and nuclear monitoring research. For the location parameters (epicenter and depth) these uncertainties are typically less than a few hundred meters. At a typical crustal P velocity of 6 km/s 100 meters represents 0.015 s, so origin time should be known to several hundredths of a second in order to be compatible. These limits may not be suitable for some engineering purposes or specialized source studies.

There is a need for a somewhat relaxed GT category, because even though the hypocentral parameters of a man-made explosion may be given *a priori*, the uncertainties may not meet the stricter requirements given above. This may be the case because of inadequate record keeping or the difficulty in carrying out suitably accurate surveying or timing prior to the availability of GPS technology. The GT1 category is meant for such cases. This still implies near-certain knowledge of location within a kilometer or so, with comparable uncertainty in origin time (several tenths of a second). Industrial explosions and even some nuclear tests may not meet this standard. Such events ought to be treated in the calibrated class of events, as discussed below, rather than being assigned GT status with inflated scale lengths.

GT5 should not ever be used in this classification system, which will help avoid confusion with the current GTX system of classification.

## **Calibrated Events: the C Class**

In contrast to the GT class, where the concern is primarily with the scale of random error in the hypocentral parameters, the class of calibrated events is dominated by concern that the estimation process which has been used to determine hypocentral parameters may have introduced significant bias. Therefore we are very much concerned about minimizing bias and understanding which hypocentral parameters may be treated as effectively bias-free. Obviously we also desire to estimate the hypocentral parameters such that the formal uncertainties (driven by uncertainty in the data), usually expressed in Gaussian terms, are as small as possible; this will be handled similarly to the “X” in the GTX formulation, discussed below in the section “Scale Length”.

A very important point about the calibrated class of events, is that it includes only events for which the epicenter (at least) has been determined in such a way as to minimize bias. Although a bit unsatisfying in a logical sense, this policy reflects the reality that the seismological

community overwhelmingly thinks of GT events (using the popular current nomenclature) as referring only to the epicenter. The other important point to be made is that this class requires an actual location analysis, not just the application of some set of network geometry criteria such as those presented by Bondar et al. (2004). In other words, application of network geometry criteria to estimate location accuracy is a precursor to calibration analysis, not a substitute for it.

Given that we do not know the Earth's velocity structure to sufficient accuracy, the only way to reduce bias for an event that was not engineered by man is to keep path lengths through the unknown Earth structure as short as possible. In other words only near-source data should be employed for estimating calibrated parameters. "Near source data" is not restricted to seismological stations at short epicentral distance, although that is by far the most common case. Mapped surface faulting, treated with all due geological sensitivity, may serve as near source data for the purpose of constraining an epicenter, as may InSAR or other types of remote sensing analyses, since the ultimate signal (e.g., surface deformation) is not subject to bias from unknown Earth structure. InSAR analysis, through determination of the distribution of rupture on a fault plane, may be used to reduce bias in focal depth. Waveform modeling (even at regional or teleseismic distances) may similarly provide useful near-source constraint on focal depth through analysis of the interference of direct and near-source surface-reflected phases.

Unfortunately, there is no methodology for obtaining usefully-calibrated hypocenters for deep earthquakes because every available data type must propagate through an excessive volume of material with insufficiently well-known velocity. The exact definition of "deep" in this context must be evaluated on a case-by-case basis, but it probably includes any event deeper than about 100 km. If uncertainties in velocity structure (and their effect on raypath geometry) are honestly propagated into the uncertainties of the derived location parameters, then the issue will be resolved by the increasing uncertainty of the location, leaving aside the question of bias.

The nomenclature I propose for the calibrated class is based on the following practical considerations about the calibration of the various hypocentral parameters:

### ***Epicenter***

Bias in epicentral coordinates can be minimized by means of seismological analysis (typically a location analysis), as well as by other means, including geological and remote-sensing analyses and *a priori* knowledge of human-engineered sources that may be too weak for GT status. It is quite common for the epicenter to be the only hypocentral parameter of an event that can be usefully constrained with minimal bias.

### ***Depth***

Focal depth is more difficult to constrain than the epicentral coordinates. In the location analysis, it requires data at epicentral distances comparable to the focal depth itself, a few tens of kilometers for crustal events, a much stricter requirement than that for the epicenter, which can

be usefully constrained with stations 100 km or so away. This distance requirement can be ignored for waveform modeling, however, as well as for analyses of teleseismic depth phases, most famously emphasized by the EHB algorithm (Engdahl et al., 1998). Therefore the minimization of bias in focal depth can be part of the general location analysis, coupled with the estimate of a minimally-biased epicenter, or it can be constrained independently, even when the epicenter may be uncalibrated.

## Origin Time

Calibration of origin time is only fully possible when both the epicenter and focal depth can be calibrated. Unless it has been specified *a priori* for a human-engineered event it must be estimated from seismic arrival time data at the shortest possible epicentral distances, and any bias in the location parameters would propagate into origin time. It is quite common, however, to encounter cases where the epicenter and origin time of an event can be constrained with near-source data (not necessarily for the event in question but through linkage to other events in a multiple event analysis), but the focal depth of the event cannot be usefully constrained, other than as an average depth for a cluster of events, some of which have well-constrained depths, or through regional seismotectonic considerations. In this case the origin time itself cannot be considered to be unbiased, but since it is reliably coupled to the assumed focal depth, the combined hypocentral coordinates can still provide valuable information on empirical travel times from a specific point in the Earth.

Given the above considerations there are three cases that need to be distinguished in the calibrated class of the nomenclature. In the following table, the asterisk indicates parameters that have been calibrated in some manner:

	<b>Epicenter</b>	<b>Focal Depth</b>	<b>Origin Time</b>
<b>CH</b>	*	*	*
<b>CT</b>	*		*
<b>CF</b>	*	*	
<b>CE</b>	*		

CH (“H” refers to hypocenter). All four hypocentral coordinates have either been inferred by means that yield minimally-biased estimates or constrained *a priori* (as in some human-engineered events that don’t quite qualify for GT1 status or better).

CT (“T” refers to travel time). Epicenter has been calibrated; depth has been fixed at some assumed value (e.g., the average depth of nearby events with constrained depths); the estimate of origin time is based on local-distance data, but relative to an uncalibrated depth. Neither the focal depth nor origin time can be considered calibrated in themselves but the combination can be used to estimate empirical travel times from the specific point in the Earth. Such events are not quite

as valuable as CH events but still have considerable value as input to model-building exercises or as validation events.

CF (“F” refers to focal depth). Epicenter and focal depth have been calibrated, but not origin time. An example could be an InSAR location for an event and depth calibrated either by an additional analysis of surface deformation to infer distributed displacement on a fault surface, or through waveform analysis. The estimate of origin time is not based on near-source readings. These events can be used in validation exercises where their epicenters are compared with locations done with ray-traced travel-times through a model.

CE (“E” refers to epicenter). The epicenter is calibrated. As with the CT class, depth has been fixed at some assumed (albeit reasonable) value. If the calibration of the epicenter has not been based on near-source seismic data (e.g., an InSAR location), the estimate of origin time must be based on regional or teleseismic arrivals and therefore cannot be considered calibrated, nor can it be used for estimate of empirical travel times. These events can be used in validation exercises where their epicenters are compared with locations done with ray-traced travel-times through a model.

## **Network Geometry Criteria: The N Class**

Events in the N class are not considered to be calibrated in the sense defined here, but the arrival time data set has been processed with some network criteria (e.g., Bondar et al. (2004), but others are developing similar criteria for different source regions) based on simple metrics such as number of readings and distribution of reporting stations, in order to provide an estimate of epicentral accuracy that is expected to account for systematic location bias. The assumption here is that 1) the data do not permit a calibration analysis because there are insufficient near-source data, or 2) that such an analysis has simply not yet been done (i.e., a bulletin has simply been scanned for candidate calibration events). If a careful relocation analysis has been done to standards that can arguably justify classification as a calibrated event, the C class should be used.

NE (“E” from epicenter). The epicentral accuracy has been estimated with an appropriate network geometry criteria. Focal depth and origin time are uncalibrated. Many so-called “GT Catalogs” are dominated by events in this category. Requires a scale length, confidence level optional.

NF (“F” from focal depth). As NE but focal depth is calibrated. Requires a scale length, confidence level optional.

## **Everything Else: The U Class**



All seismic events that do not fit into one of the GT, C or N classifications are considered uncalibrated. That does not mean that none of the hypocentral coordinates are calibrated, only that the epicenter is not considered to be calibrated. The following classifications are defined:

UE (“E” from epicenter). No hypocentral parameters are calibrated but there is a credible estimate of epicentral accuracy from a location analysis (confidence ellipse), leaving aside the question of systematic location bias. Requires a scale length, confidence level optional.

UF (“F” from focal depth). As UE, but focal depth is calibrated. The subset of events in the EHB catalog (Engdahl et al., 1998) that carries depth estimates based on analysis of teleseismic depth phases would fall into this category, as would any event that has been the subject of a waveform modeling exercise that solves for focal depth. Requires a scale length, confidence level optional.

U (uncalibrated). Simply a dot on a map. No credible information is available on location accuracy, epicentral or otherwise. No scale length or confidence level is used.

## Scale Length

With the exception of the “U” category all classifications should carry a scale length, equivalent to the “X” in the GTX formulation. The ground truth (GT) class categories are defined with specific scale lengths, which refer to the uncertainty in both the epicenter and focal depth.

For the Calibrated (C) and Network Geometry Criteria (N) classes the scale length is related to the uncertainty in epicenter only. For the CH class one would have to refer to a more detailed description of the data set to learn anything quantitative about the uncertainty in focal depth. The scale length is an integer, in kilometers, related to the uncertainty of the epicenter. Network geometry criteria always yield a single value for scale length. For the C class, as discussed above, there is no consensus about how the 2-dimensional uncertainty in an epicenter should be reduced to a single number. Three possibilities that seem reasonable when dealing with an ellipse with semi-minor axis **a** and semi-major axis **b** are:

- Nearest integer to the semi-major axis length of the confidence ellipse:  $\text{nint}(\mathbf{b})$ .
- Nearest integer to the average of the two semi-axis lengths:  $\text{nint}((\mathbf{a}+\mathbf{b})/2)$ .
- Nearest integer to the radius of the circle with the same area as the ellipse:  $\text{sqrt}(\mathbf{ab})$ .

For a circular confidence region all three methods are equal. As the ellipticity of the confidence region increases, there will be substantial differences between the different scale lengths, but the first method will always yield the largest value. For a confidence ellipse with semi-axis lengths 1 and 5 km, for example, the scale length calculated with the three methods would be 5, 3, and 2 km, respectively. I have adopted the first method for my own work, in order to be more conservative in the “GT5” wars, and I recommend the same to others.

Scales lengths larger than 9 are permitted, but I think they have rapidly diminishing value in the current research environment. When the scale length of confidence ellipses moves into double digits, one ought to begin to worry about the legitimacy of the assumptions underlying the statistical analysis. I would consider moving such events into one of the uncalibrated categories.

## **Confidence Levels**

As Bondar et al. (2004) pointed out, it is necessary to specify the confidence level that has been used in determining epicentral uncertainties, e.g., as a subscript in the form “GT<sub>90</sub>5” to indicate that the confidence ellipse was calculated for a 90% confidence level. Compliance on this point seems to be casual at best. It is admittedly awkward to include the subscript in computer output, and since the nomenclature I am proposing here is primarily intended to be carried in digital files, I leave it as optional in that context, with the strong recommendation to clarify the issue in accompanying documentation. I strongly encourage the use of the subscript in published reports to ensure wide understanding.

## **Nomenclature of Nomenclatures**

It is inevitable that some shorthand will be needed to refer to the set of classifications I have defined above, in particular with respect to the widely-known GTX system. I suggest “GTCNU” or “GTCNU System of Classification” or “GTCNU System of Classification for Location Accuracy”, depending on the degree of abbreviation desired.

## **References**

- Bondar, I., Myers, S. C., Engdahl, E. R., & Bergman, E. A. (2004). Epicentre accuracy based on seismic network criteria. *Geophysical Journal International*, 156, 483–496. doi:10.1111/j.1365-246X.2004.02070.x
- Engdahl, E. R., van der Hilst, R. D., & Buland, R. P. (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America*, 88(3), 722–743.
- Herrin, E. (1968). Introduction to "1968 Seismological Tables for P Phases." *Bulletin of the Seismological Society of America*, 58(4), 1193–1195.