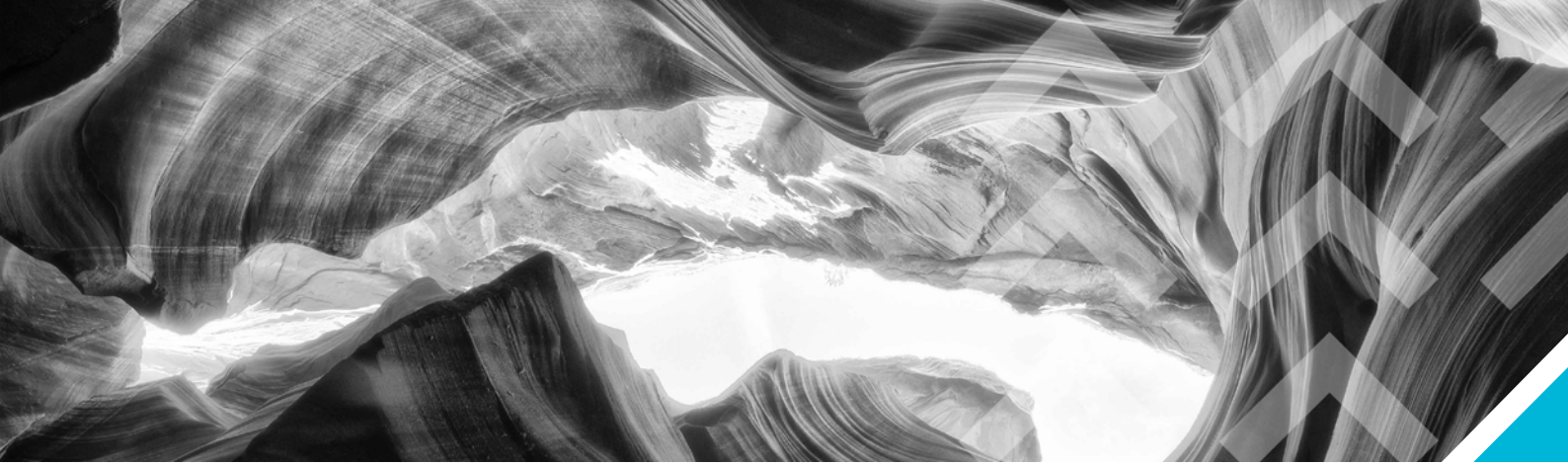
Axim GEOSPATIAL
AN NV5 COMPANY

DATA FITNESS™

Geospatial Data Quality Assessment

What Is Your Geospatial Data Good For?

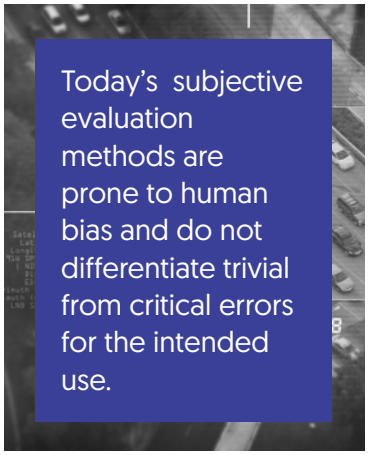
Today's subjective evaluation methods are prone to human bias and do not differentiate trivial from critical errors for the intended use.



Terms & Conditions

Introduction

There is currently no definitive way to measure geospatial data quality and its fitness-for-use. A majority—typically more than 90%—of QA\QC effort is spent evaluating subjective criteria *with subjective evaluation methods*.



These methods rely on acceptance based on "gold standard" use parameters and are prone to human bias, do not differentiate trivial errors from those critical to the intended use of the data, and don't scale to meet today's dynamic update and use requirements. Methods are needed to evaluate datasets that have mixed provenance and lineage upon input or at any point in the development cycle. In addition, these mechanisms need to be adjustable by the end users for them to determine what is "good enough" for their intended use. Axim Geospatial has developed a solution to address this problem: Data Fitness™.

This scalable, cloud-based service provides a statistical and repeatable measure of geospatial data quality, tracks lineage and assesses fitness–"for-use" against user-provided criteria.

Why Data Fitness Assessment is Critical Now

The National Geospatial-Intelligence Agency (NGA) has proposed a GEOINT Broker platform to leverage geospatial data from numerous sources and deliver GEOINT content and services to diverse customers. To do this, data filters that deliver a normalized, repeatable and understandable diagnostic of a dataset's quality is needed. This "map filter" can be applied to thematic inputs, combinations of datasets at different stages of production or integration, and any combination of final outputs. The value of such a tool is apparent: a reliable, repeatable diagnostic process that allows users to determine what data are fit for their intended use.

The QA/QC Problem

An Army commander uses a map differently from an Air Force pilot. Representations of features critical to one may not be important to the other. As a result, each user judges the quality of a map based on its fitness for their application. While numerous automated methods exist for correctly rating objective qualities such as topology, positional accuracy, and metadata integrity, evaluating subjective qualities such as omission and commission and attribute correctness are typically performed manually and fraught with problems.

Those reviews and data improvements are not captured leaving the end user unclear as to the lineage of data quality. Also, users are awash in data and are often challenged to compare and contrast datasets to determine which is better for both quick turnaround, tactical missions or longer-term planning.

Reviewing the quality of subjective interpretations focuses primarily on completeness of feature identification and accuracy of attribute classification. Have all the roads been identified? Has each one been classified properly? These subjective interpretations have a significant impact on data's fitness-for-use. A dirt path incorrectly attributed as a paved road matters for a platoon leader, but not to a tank commander.

Our research shows that current accept/reject methods used in the community are heavily biased towards rejecting geospatial data that should be accepted. More than 95% of errors that lead to rejection are not critical to end-user applications, often preventing delivery of data to people in the field who could have used them safely and effectively for their intended purposes.

Acceptance criteria are based on only one set of presumed use criteria and do not allow for flexibility based on other uses. This bias slows the production and QA/QC process and drives up the cost. Compounding the bias issue is the inflexibility for the reviewer to weight errors differently depending on use. Errors are judged equally – or at least statically - in acceptance or rejection of the map because the reviewer has no way of differentiating critical errors from harmless ones relative to an intended use.

A theoretical solution was presented in a 2007 technical paper, "Geospatial Statistical Quality Management," written by Charles E. Gillies. It laid out a framework for marrying principals for describing the quality of geospatial data (ISO 19157-2013) with military standards for product quality testing (MIL-STD-1916) to create an integrated geospatial sampling model (IGSM). From that paper and years of experience, Axim Geospatial developed Data Fitness as a cloudbased, software solution to meet this need.

Figure 1: Fitness Center provides use case data models and allows users to set target MTPDs and move feature classes between criticality categories per requirements

Data Fitness™ provides:

- Tools for highly optimized review of subjective data quality errors of omission, commission, attribute correctness, and thematic accuracy
- A Fitness-for-use Starter Kit that categorizes feature classes by criticality (i.e., critical, important, and referential) and provides suggested Map Tolerance Percent Defect (MTPD) values for defined use cases. See Figure 1
- A Fitness Center for reviewing error rates by criticality and what-if capabilities to further assess data quality. See Figure 2
- Proprietary statistical sampling techniques to efficiently evaluate any dataset at any point in the workflow from source evaluation, in-workflow quality evaluation, to final QA,
- Portable tracking of all error calls and data quality lineage accessible to anyone at any time,
- The ability to identify trends early with source data providers and internal resource error propagation
- Socialization. Ability to compare how the fitness of your data compares to your peers, and
- Data quality reporting of all errors found to drive process improvements and data fixes.

Data Fitness™ allows a user to select a use case and review default settings of feature class criticality and suggested MTPD levels. The user can then adjust those settings and save it as their own. Once sampling is complete, the user is presented with the same settings and scores. Yellow highlighted feature classes have the highest errors, and the error chart at the bottom shows how those errors are distributed amongst omissions, commissions, etc. Users can then move feature classes between the three criticality levels to see if it impacts the datasets ability to meet their needs.

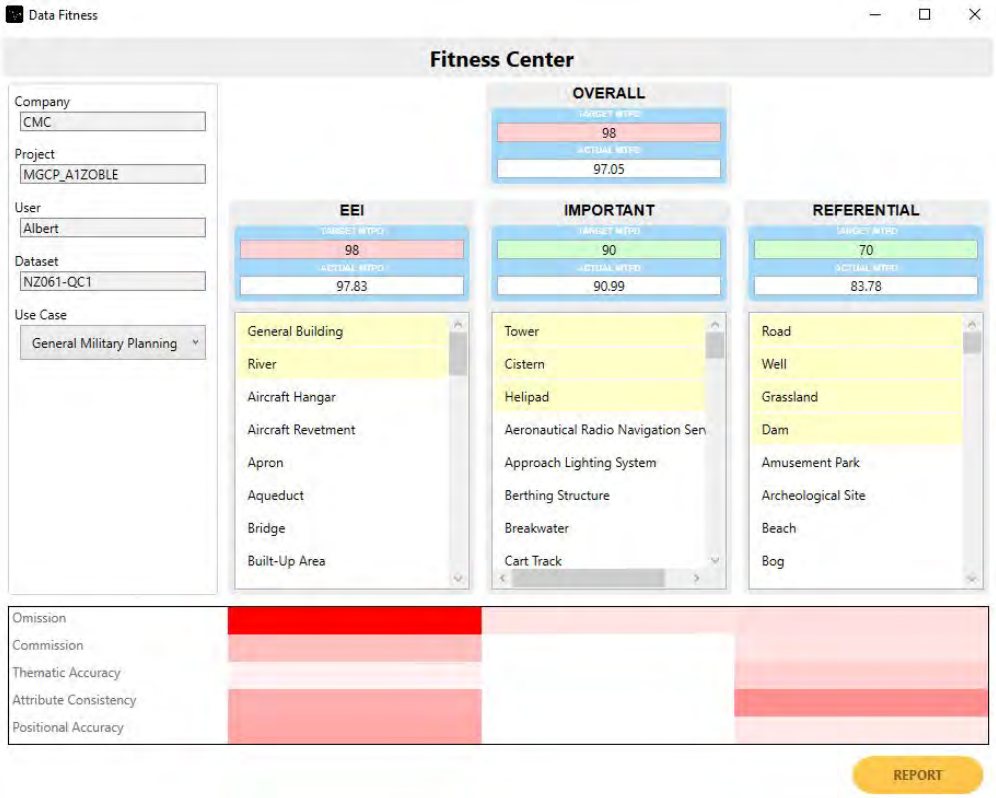


Figure 2: After processing Fitness Center displays target and actual MTPDs and error distributions against user requirements

The QA/QC Problem

Data Fitness™ meets the GEOINT community's need for reliable and repeatable tools to measure quality, track lineage and assess a geospatial dataset's fitness–for-use as defined by a user community. Data Fitness™ reduces review time by up to 90% (see Figure 3) while achieving a higher rate of accurately accepting and rejecting datasets for a defined set of use requirements.

The lineage of all quality review and changes are embedded within the dataset, so they are portable and ready for interrogation at any time. Additionally, the MTPD diagnostic provides confidence in the dataset that was not there before. Data Fitness™ provides a statistically relevant, repeatable measure. With the need for dynamic map updates using a wide range of sources and methods, understanding "what is this data good for" now has a tool suite ready for deployment.

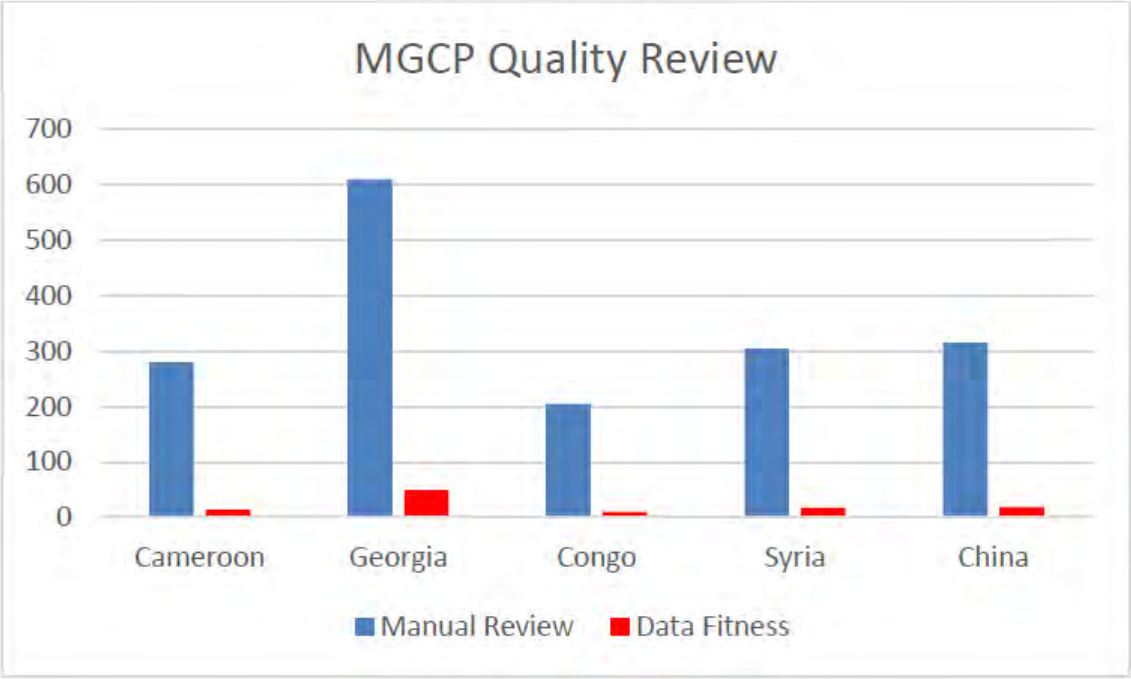


Figure 3: Data Fitness cuts review times by up to 90%

Reach Out To Us

The staff at Axim Geospatial is interested to hear your thoughts on how Data Fitness™ can help you improve your geospatial data quality and understand its fitness-for-use.
Schedule a meeting to talk more.

SCHEDULE A MEETING