

## A REVIEW OF BENCHMARKING IN PHOTOGRAMMETRY AND REMOTE SENSING

K. Bakula<sup>1,4</sup>, J. P. Mills<sup>2,4</sup>, F. Remondino<sup>3,4</sup>

<sup>1</sup>Department of Photogrammetry, Remote Sensing and Spatial Information Systems, Faculty of Geodesy and Cartography, Warsaw University of Technology, Warsaw, Poland – krzysztof.bakula@pw.edu.pl

<sup>2</sup>School of Engineering, Newcastle University, Newcastle upon Tyne, UK – jon.mills@newcastle.ac.uk

<sup>3</sup>3D Optical Metrology (3DOM), Bruno Kessler Foundation (FBK), Trento, Italy – remondino@fbk.eu

<sup>4</sup>European Spatial Data Research EuroSDR, Leuven, Belgium

### Commission I, WG I/10

**KEY WORDS:** algorithms, benchmark, data quality, evaluation, sensors

### ABSTRACT:

This paper is a preface of the ISPRS/EuroSDR Workshop on the “Evaluation and Benchmarking of Sensors, Systems and Geospatial Data in Photogrammetry and Remote Sensing” that was held in Warsaw University of Technology on 16-17 September 2019. The paper reviews benchmarking in photogrammetry and remote sensing found in the literature relating to geodata. The first part of the paper is a bibliographic analysis based on queries in Scopus and Web of Science databases which shows an increase in research activities based on benchmarking data. In the second part, a review of past and ongoing benchmarking initiatives is presented, providing examples of initiatives within e.g. ISPRS and EuroSDR. The topic of evaluating data, sensors and algorithms with benchmarking activities is interesting as it provides the opportunity to compare, with a unique approach, research results from independent scientists. As hereafter reported, benchmarking has increased in recent years, with many benchmarks being presented in the photogrammetric and remote sensing communities.

### 1. INTRODUCTION

A benchmark can be broadly defined as a standard or point of reference against which entities may be compared. It is also understood as a test designed to evaluate or compare the performance of computer hardware or software. Nowadays, benchmarks are a useful tool for evaluating and measuring sensors and platforms or to assess the quality of geospatial data (geodata) and processing algorithms.

The origins of benchmarking can be found in process management. Benchmarking is the practice of comparing processes, i.e. business and performance metrics, to industry standards and best practice from other companies. Dimensions that can be typically measured include quality, time and cost. Benchmarking is used to measure performance using a specific indicator (for example, cost per unit of measure, productivity per unit of measure, cycle time per unit of measure, or defects per unit of measure), resulting in a metric of performance that is then compared to others (Fifer, 1989).

Benchmarking may be a one-off, single action but may also be treated as a continuous process. There is no single benchmarking process that has been universally adopted. However, the wide appeal and acceptance of benchmarking has led to the emergence of benchmarking methodologies (Boxwell Jr, 1994), with benchmarking methodology typically comprising 12 stages (Camp, 1989):

- Select subject;
- Define the process;
- Identify potential partners;
- Identify data sources;
- Collect data and select all partners;
- Determine the gap;
- Establish process differences;
- Target future performance;

- Communicate;
- Adjust goal;
- Implement;
- Review and recalibrate.

Such approaches from process management, which primarily relate to management in commercial companies, may also be adopted in benchmarking in the geospatial industry. There is no doubt that scientific problems must be identified, and existing solutions should be explored with potential partners. After these initial steps, test data should be prepared and final partners (ideally via an open call for participation) should be found. It is an important requirement that participating organizations are leaders in selected areas to ensure that the planned benchmark is in accordance with up-to-date issues which are related to that subject.

Benchmarks in Photogrammetry and Remote Sensing are typically undertaken as an evaluation of:

- a given sensor /or system of sensors;
- geodata collected using a particular platform / sensor;
- algorithms or methods for geodata processing.

A benchmark can be assessed:

- by many users (benchmark participants) with a proposed methodology – i.e. if geodata from the sensor is provided as a reference sample for processing in evaluation or comparison by benchmark participants;
- by many users (benchmark participants) with a defined methodology – i.e. if the evaluation or comparison of algorithms is the aim of the benchmark;
- by a benchmark proposer or defined algorithm and analysed by benchmark participants – i.e. if the sensor or data are evaluated by analysing different samples collected, with benchmark participants submitting them for evaluation, perhaps within an organised contest.

Test data in benchmarking research is usually prepared as a reference dataset provided as open data for participants. It may also be necessary to define how samples should be collected to guarantee objectivity of results obtained under similar conditions and experimental setups at (potentially many) participants' institutions.

As a tool in the hands of scientists, the concept of benchmarking can be widely understood, and it allows for the exchange of thoughts and theories, joint experiments and independent research, often leading to universal conclusions. This short review of the development of benchmarking in photogrammetry and remote sensing reveals the growth in popularity of benchmarks in the presented bibliographic analysis of published articles and conference papers. It helps with retrieving information concerning completed, ongoing and potential future scientific initiatives related to benchmarking.

## 2. ANALYSIS OF BIBLIOGRAPHY DATABASES

To gain an overview on benchmarking techniques, it is worth tracing how often the phrase “benchmark” appears in the bibliographic databases. There are many bibliographic databases, but only a few allow the discovery of such information through an advanced query. To illustrate the popularity of benchmarking, query results are presented based on Scopus and Web of Science databases. These databases contain the most significant journal and conference papers in photogrammetry and remote sensing (including the ISPRS Archives and Annals).

Elsevier’s Scopus database (<https://www.scopus.com/>) permits an advanced search for articles by title, abstract and/or keywords. This search can therefore be combined. A combined query searching for the word “benchmark” and related words such as “benchmarking” or “benchmarks” in titles, abstracts or keywords returned 1353 papers (as accessed on 6 September 2019) related to, or mentioning, benchmarks (Figure 1). These results are connected to many subject areas (Figure 2) which include photogrammetry and/or remote sensing. Analysis of the results of this query show a growing trend when communication between scientists significantly increased in the era of Internet and digital bibliographic databases.

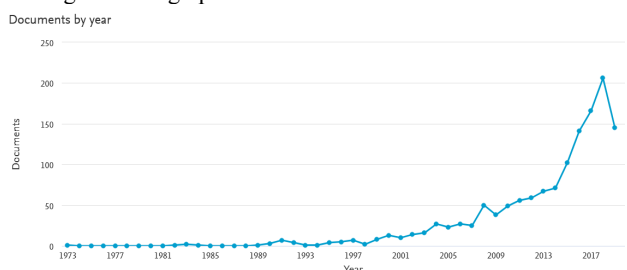


Figure 1. Number of results for query: *TITLE-ABS-KEY*("benchmark\*" AND ("photogrammetry" OR "remote sensing")) in Scopus database.

The query discussed above allows the researcher to identify a lot of published work that refers to benchmarks, but does not directly relate to them. Therefore, a further search was performed regarding the occurrence of the word “benchmark” or its related form in a paper’s title and reference to photogrammetry or remote sensing in any other field (keywords or abstract). The result showed 90 items which are included in the Appendix. The growth in the number of items directly related to benchmarks is clearly noticeable in the last 10 years (Figure 3).

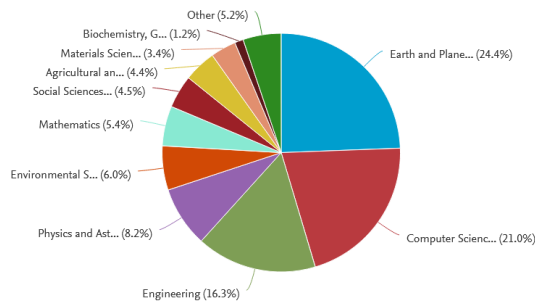


Figure 2. Subject areas of results for query: *TITLE-ABS-KEY*("benchmark\*" AND ("photogrammetry" OR "remote sensing")) in Scopus database.

The Web of Science database (<https://clarivate.com/webofsciencegroup/solutions/web-of-science/>) was originally produced by the Institute for Scientific Information (ISI) and is currently maintained by Clarivate Analytics. Search tools include the possibilities to produce search results comparable to the first query performed with Scopus. The results of searching for papers with the phrase “benchmark” or related forms in photogrammetry and remote sensing in all fields returned 1699 results (accessed on 6 September 2019) (Figure 4), and dates of publication confirm an increase in number over the last decade. Division of the results into the Web of Science categories (Figure 5) shows that this research technique is popular in many groups from various disciplines related to remote sensing and photogrammetry.

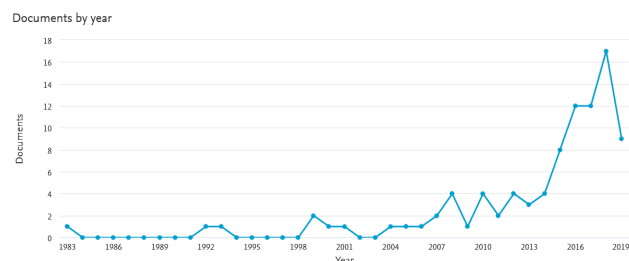


Figure 3. Results of query: *TITLE*("benchmark\*") and *TITLE-ABS-KEY*("photogrammetry" or "remote sensing") in Scopus database.

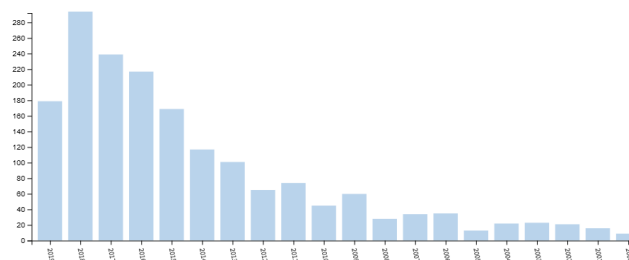


Figure 4. Number of results for query: *All* = (benchmark\* and (photogrammetry or remote sensing)) in the Web of Science database.

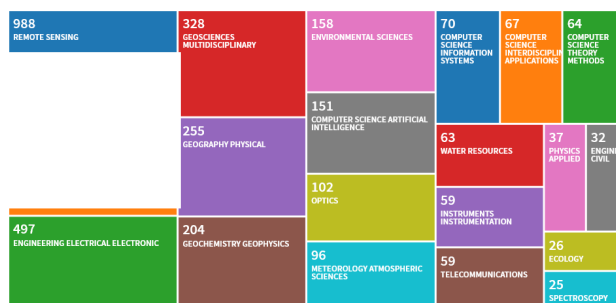


Figure 5. Disciplines of results for query:  $All = (benchmark * and (photogrammetry or remote sensing))$  in the Web of Science database.

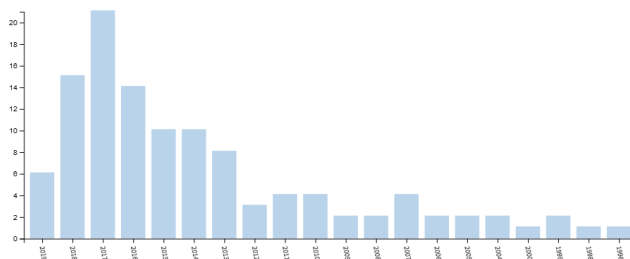


Figure 6. The results of query:  $TI=benchmark * AND all = (photogrammetry or remote sensing)$  in Web of Science database.

Searching for documents titled with the word “benchmark” and related to photogrammetry and remote sensing from the Web of Science retrieved 117 documents (accessed on 6 September 2019) (Figure 6). The Scopus and the Web of Science results are comparable and indicate that this method of scientific research activity has become increasingly popular in recent years.

### 3. OVERVIEW OF BENCHMARKING INITIATIVES

Organisations such as the International Society for Photogrammetry and Remote Sensing (ISPRS) and European Spatial Data Research (EuroSDR) have repeatedly supported the organisation of benchmarks. In the following, selected examples of completed (Section 3.1), ongoing and forthcoming (Section 3.2) benchmarks are presented.

#### 3.1 Completed benchmark initiatives

Benchmarking activities are in the interest of many ISPRS working groups, with the following a summary of some of the most recent and high profile benchmarks related to sensors, algorithms and methods.

##### 3.1.1 ISPRS Test Project on Urban Classification, 3D Building Reconstruction and Semantic Labeling

This benchmark commenced in 2011 and is still ongoing. The most important application of the tested photogrammetric data was object detection and 3D building reconstruction in urban areas. Participants in this project could choose of the following tasks:

- 1) Urban Object Detection: participants may choose to detect single object classes, or can try to extract several object classes simultaneously;
- 2) 3D Building Reconstruction: participants shall reconstruct detailed 3D roof structures in the test areas.

In both cases reference data were created by the organiser of the benchmarks. The participants submitted their results to the organisers of the tests: urban classification and 3D reconstruction, 2D and 3D Semantic Labeling. The results from participants were compared to the reference data and participants were informed about the results of the evaluation. Some outcomes of these benchmarks were published by Rottensteiner (2013) and the initiative is still ongoing. Reference datasets were published in 2018. More information about this benchmarks on: <http://www2.isprs.org/commissions/comm3/wg4/tests.html>

##### 3.1.2 Joint ISPRS/EuroSDR Benchmark on High Density Aerial Image Matching

The project (Haala, 2014) aim to evaluate the potential of photogrammetric 3D data capture, with reference to ongoing developments in software for automatic image matching. The basic scope of this benchmark was the evaluation of 3D point clouds and digital surface models (DSMs) generated from aerial images in different software systems. The benchmark considered state-wide-generation of high quality DSMs and the applicability of open and commercial tools. Three aerial image blocks were provided as test data.

##### 3.1.3 ISPRS/EuroSDR Benchmark for Multi-Platform Photogrammetry

The aim of the project (Nex et al, 2015) was to assess the accuracy and reliability of calibration methods and image orientation from different platforms as well as their integration for image matching and dense point cloud generation. In current research, an important issue is that large changes in perspective and scale differences need to be tackled in image orientation and (dense) image matching and this is generally not approached systematically. By providing a new benchmark dataset consisting of state-of-the-art sensor data and covering different relevant tasks and scenarios, the current status of research was identified. Two test data sets of Dortmund (Germany) and Zurich (Switzerland) were released for this benchmark in 2014/2015.

##### 3.1.4 ISPRS Benchmark on UAV Semantic Video Segmentation

This 2017 ISPRS Scientific Initiative aims to promote and advance the video segmentation task of VHR UAV sequences. With this project in 2017-2018, the authors wanted to pave the way for a unified framework towards meaningful quantification of semantic segmentation from UAV imagery and video. In this project, labelling image segmentation algorithms were developed. Eight classes (building, road, tree, low vegetation, moving car, static car, human and background) were detected based on datasets of 10 videos with 75,000 frames captured by DJI Phantom systems in Germany and China. In future, there are plans to develop a comprehensive benchmark creating the largest dataset of UAV scenes with high-quality annotations, a sound evaluation methodology for pixel-level semantic labelling, and a corresponding challenge associated with these. (Ying Yang & Yilmaz, 2018).

##### 3.1.5 ISPRS Benchmark Challenge on Large Scale Classification of VHR Geospatial Data

Another 2017 ISPRS Scientific Initiative, the goal of this benchmark was the generation of a publicly available, large-scale, VHR, multi-spectral dataset for training and the evaluation of sophisticated machine learning models. The authors were aiming at the provision of a complex and realistic benchmark in

segmentation or VHR imagery in change detection. Worldview 2 satellite images of Toulouse were annotated into six land cover classes: 'impervious surface', 'building', 'pervious surface', 'high vegetation', 'cars', and 'water'. In the project implementation, a fully automatic evaluation tool for a comprehensive accuracy assessment of semantic instance segmentation and automatic evaluation tool for change detection were prepared. (Rosche, 2018).

### 3.1.6 Benchmark on Terrestrial Laser Scanning for Forestry Applications

This EuroSDR benchmark (Liang et al., 2018), aims to evaluate the quality, accuracy, and feasibility of automatic, semi-automatic or manual tree extraction methods based on high-density TLS data. Provided test datasets (24 sample plots of size 32 m x 32 m) were collected in Finnish forests with terrestrial scanners.

## 3.2 Ongoing and Future Initiatives

Many benchmarking initiatives are still ongoing, and the following are examples that have been found or linked on the ISPRS webpage as open initiatives. There are many initiatives related to photogrammetry and remote sensing data processing that provide benchmark data for tests performed by other scientists. For example, the KITTI benchmark (Geiger, et al., 2012) and the "MLS 1 - TUM City Campus" benchmark (Gehring et al., 2017) deliver many types of data from mobile mapping platforms. Some datasets collected on-board a Micro Aerial Vehicle can be found in the EuRoC benchmark (Burri et al, 2016). The Large-Scale Point Cloud Classification Benchmark (Hackel et al. 2017) provided a large labelled 3D point cloud data set of natural scenes with over 4 billion points of diverse urban scenes. The goal of the initiative was to help data-demanding methods like deep neural nets unleash their full power and to learn richer 3D representations. The aim of the "3D sensors and systems for metrology and industrial vision" benchmark, is to provide free samples of 3D sensor data suitable for industrial vision and close range 3D measurements. Moreover, the call for participants' "data for corridor mapping" was also announced. All of these benchmark datasets are still available despite the fact that they were initiated several years ago. In this section, a number of further initiatives are introduced in more detail.

### 3.2.1 Benchmark on Indoor Modelling

This ISPRS Benchmark on Indoor Modelling aims to deal with the fact that while results on various 3D indoor models from point cloud data have been reported in literature, a comparison of the performance of different methods has not been possible. Spatial models of indoor environments are needed in a growing number of applications including navigation, emergency response and a range of location-based services. Test data in this benchmark comprises five point clouds captured by different sensors in indoor environments of various complexities which should be processed to the model used for the purpose of indoor navigation (Khoshelham et al., 2017).

### 3.2.2 The FINE Benchmark

The Fisheye Indoor Narrow Spaces Evaluation (FINE) benchmark (<http://www.3d-arch.org/FINE-benchmark.pdf>) provides fisheye images and terrestrial laser scanning point clouds of complex environments. The aim of the project is to

evaluate the performance of different image-based processing methods when surveying complex spaces. The main questions that the benchmark poses concern the potential of image-based fisheye techniques for 3D reconstruction of indoor narrow spaces and whether these can be considered as valid low-cost alternatives to static or mobile/hand-held laser scanning instruments.

### 3.2.3. The HyRANK Hyperspectral Dataset and Benchmark Framework

The HyRANK Dataset and Benchmark is another ISPRS scientific initiative. The main objective is to fill the current gap regarding the limited availability of hyperspectral datasets and benchmarking frameworks for validating new classification methods against the state-of-the-art. The testing dataset contains two hyperspectral images from the Hyperion sensor (EO-1, USGS) and the validation is a set of three images. The purpose for the benchmark participant is to use their classification algorithms to produce a land cover map in accordance to nomenclature that follows the CORINE Land Cover principles (14 classes) (Karantzalos et al., 2019).

### 3.2.3. GeoBIM benchmark 2019

This ISPRS and EuroSDR benchmark started in 2019. The aim is to investigate the available technical solutions to support research and activities related to the GeoBIM topic (integration and interoperability of data concerning 3D geoinformation and 3D building information models) (Noardo et al., 2019). This is the first project that will provide insight into the current state-of-the-art of the open standards implementation in the 3D geo and BIM domain, and also identify remaining issues. It will compare the ability of existing software tools to use and process CityGML and IFC models and understand their performance. In this benchmark, three aspects are investigated in four tasks related to the types of data mentioned above, their georeferencing and conversion between different types of geodata formats.

### 3.2.4 3DOMcity benchmark

A novel multi-purpose benchmark for assessing the performance of the entire image-based pipeline for 3D urban reconstruction and 3D data classification is presented during the Workshop on "Evaluation and Benchmarking of Sensors, Systems and Geospatial Data in Photogrammetry and Remote Sensing" (Özdemir et al., 2019). This benchmark called "3DOMcity" is photogrammetric contest and is publicly available at <https://3dom.fbk.eu/3domcity-benchmark>. It covers the entire 3D reconstruction pipeline, including image orientation, dense image matching, point cloud classification and 3D building reconstruction. The performance assessment is performed within a metrological context. Provided datasets offer 2D and 3D data collected at a very high spatial resolution.

### 3.2.5 ISPRS/EuroSDR Single-Photon LiDAR benchmark

Single Photon and Geiger-Mode LiDAR (SPL/GML) are relatively new types of airborne laser scanning sensors, and are technologies that can potentially provide very accurate, high-resolution 3D mapping in a more efficient way with respect to conventional airborne laser scanning. However, such LiDAR sensors can cause problems with noise and lower accuracy than linear full-waveform systems. Many tests already conducted by national mapping and cadastral agencies cannot comprehensively answer the question regarding a potential change in production pipelines and upgrading existing airborne sensors due to different

times of acquisition, different land cover and various scanning parameters. Bernard et al. (2019) describes the activities that have so far been conducted as part of a EuroSDR initiative, including a global on-line questionnaire to understand the interest in SPL/GML technologies.

#### 4. SUMMARY AND CONCLUSION

The research into the creation of benchmarks described in this paper is not an exhaustive list of all such initiatives. The wide variety of sensors and systems available on the market for collecting geospatial data renders the evaluation of derived information, calibration of sensors and benchmarking of systems a critical task. It is also an important scientific issue for many professionals. In daily work, the assessment of tools for collecting geospatial data resources is a crucial issue for all professionals handling such data. Although benchmarks can be understood differently, starting from preparing datasets for other scientists for evaluation of their methods to developed programmes of research, benchmarking can help in complex investigations and can build scientific networks, thereby bringing researchers together to conduct joint research projects.

This paper was prepared to introduce the background to a joint ISPRS/EuroSDR Workshop on Evaluation and Benchmarking of Sensors, Systems and Geospatial Data in Photogrammetry and Remote Sensing, held in Warsaw, Poland in September 2019. The benchmarking workshop provides a good opportunity for attendees who want to extend their knowledge in the fields of photogrammetry and remote sensing. During the planned workshop, research on evaluation sensors and systems, and assessment of data quality will be presented in selected examples of applications. In this event, some benchmarking proposals will be also announced and presented for release to interested researchers from the photogrammetric and remote sensing community.

#### REFERENCES

Bernard, C., Mills, J. P., Talaya, J., Remondino, F., 2019. Investigation into the potential of single photon airborne laser scanning technology, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W13, 927-934.

Boxwell Jr, R. J., 1994. Benchmarking for competitive advantage. New York: McGraw-Hill. p. 225.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., et al., 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10), 1157-1163.

Camp, R., 1989. The search for industry best practices that lead to superior performance. Productivity Press.

Fifer, R. M., 1989. Cost benchmarking functions in the value chain. *Strategy & Leadership*, 17(3), 18-19.

Gehring, J., Hebel, M., Arens, M., Stilla, U., 2017. An approach to extract moving objects from MLS data using a volumetric background representation. *ISPRS Annals of the*

*Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1, 107-114.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354-3361.

Haala, N., 2014. Dense Image Matching Final Report. EuroSDR Publication Series, Official Publication No. 64, 115-145.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., & Pollefeys, M., 2017. Semantic3d. net: A new large-scale point cloud classification benchmark. arXiv preprint arXiv:1704.03847.

Karantzos, K., Karakizi, C., Kandylakis, Z., Antoniou, G. 2018. HyRANK Hyperspectral Satellite Dataset I (Version v001), <http://doi.org/10.5281/zenodo.1222202>.

Khoselham, K., Díaz Vilariño, L., Peter, M., Kang, Z., Acharya, D., 2017. The ISPRS benchmark on indoor modelling, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W7, 367-372.

Liang, X., Hyyppä, J., Kaartinen, H., Lehtomäki, M., Pyörälä, J., Pfeifer, N. et al., 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS journal of photogrammetry and remote sensing*, 144, 137-179.

Nex, F., Gerke, M., Remondino, F., Przybilla H.-J., Bäumker, M., Zurhorst, A., 2015. ISPRS Benchmark for Multi-Platform Photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W4, 135-142.

Noardo, F., Ohori, K. A., Biljecki, F., Krijnen, T., Ellul, C., Harrie, L., Stoter, J., 2019. GeoBIM benchmark 2019: design and initial results. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2/W13).

Özdemir, E., Toschi, I., Remondino, F., 2019. A Multi-Purpose Benchmark for Photogrammetric Urban 3D Reconstruction in a Controlled Environment, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Rosche R., 2018. Report for 2017 Scientific Initiative: ISPRS Benchmark Challenge on Large Scale Classification of VHR Geospatial Data. [https://www.isprs.org/society/si/SI-2017/ISPRS-SI2017-TC2\\_WG6\\_Roscher\\_Report.pdf](https://www.isprs.org/society/si/SI-2017/ISPRS-SI2017-TC2_WG6_Roscher_Report.pdf)

Rottensteiner, F., 2013. ISPRS Test Project on Urban Classification and 3D Building Reconstruction: Evaluation of Building Reconstruction Results. Technical report.

Ying Yang, M, Yilmaz A., 2018. Report for Scientific Initiative 2017 ISPRS Benchmark on UAV Semantic Video Segmentation [https://www.isprs.org/society/si/SI-2017/ISPRS-SI2017-TC2\\_WG5\\_Yang\\_Report.pdf](https://www.isprs.org/society/si/SI-2017/ISPRS-SI2017-TC2_WG5_Yang_Report.pdf)

APPENDIX

List of the papers being the result of the query *TITLE("benchmark\*") and TITLE-ABS-KEY("photogrammetry" or "remote sensing")* in Scopus database:

Authors	Title	Year	Source
Helber P., Bischke B., Dengel A., Borth D.	Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification	2019	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
Hou D., Xing H.	V-rsir: A web-based tool and benchmark dataset for remote sensing image retrieval	2019	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Ahuir-Torres J.I., Bausch N., Farrar A., Webb S., Simandjuntak S., Nash A., Thomas B., Muna J., Jonsson C., Mathew D.	Benchmarking parameters for remote electrochemical corrosion detection and monitoring of offshore wind turbine structures	2019	Wind Energy
Condorelli F., Rinaudo F.	Benchmark of metric quality assessment in photogrammetric reconstruction for historical film footage	2019	ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Stofferahn E., Fisher J.B., Hayes D.J., Schwalm C.R., Huntzinger D.N., Hantson W., Poulter B., Zhang Z.	The Arctic-Boreal vulnerability experiment model benchmarking system	2019	Environmental Research Letters
Riquelme A., Cano M., Tomás R., Jordá L., Pastor J.L., Benavente D.	Digital 3D Rocks: A Collaborative Benchmark for Learning Rocks Recognition	2019	Rock Mechanics and Rock Engineering
Ma N., Szilagyi J.	The CR of Evaporation: A Calibration-Free Diagnostic and Benchmarking Tool for Large-Scale Terrestrial Evapotranspiration Modeling	2019	Water Resources Research
Zhou W., Newsam S., Li C., Shao Z.	PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval	2018	ISPRS Journal of Photogrammetry and Remote Sensing
Collier N., Hoffman F.M., Lawrence D.M., Keppel-Aleks G., Koven C.D., Riley W.J., Mu M., Randerson J.T.	The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation	2018	Journal of Advances in Modeling Earth Systems
Shen Y., Zhou X., Liu J., Chen J.	CSRS-SIAT: A benchmark remote sensing dataset to semantic-enabled and cross-scales scene recognition	2018	International Geoscience and Remote Sensing Symposium (IGARSS)
Huang B., Lu K., Audebert N., Khalel A., Tarabalka Y., Malof J., Boulch A., Saux B.L., Collins L., Bradbury K., Lefèvre S., El-Saban M.	Large-scale semantic classification: Outcome of the first year of inria aerial image labeling benchmark	2018	International Geoscience and Remote Sensing Symposium (IGARSS)
Helber P., Bischke B., Dengel A., Borth D.	Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification	2018	International Geoscience and Remote Sensing Symposium (IGARSS)
Dumitru C.O., Schwarz G., Datcu M.	Evaluation of retrieved categories from a TerraSAR-X benchmarking data set	2018	International Geoscience and Remote Sensing Symposium (IGARSS)
Sanches I.D., Feitosa R.Q., Achancearay P., Montibeller B., Luiz A.J.B., Soares M.D., Prudente V.H.R., Vieira D.C., Maurano L.E.P.	LEM benchmark database for tropical agricultural remote sensing application	2018	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Goldberg H., Brown M., Wang S.	A benchmark for building footprint classification using orthorectified RGB imagery and digital surface models from commercial satellites	2018	Proceedings - Applied Imagery Pattern Recognition Workshop
Scott G.J., Hagan K.C., Marcum R.A., Hurt J.A., Anderson D.T., Davis C.H.	Enhanced fusion of deep neural networks for classification of benchmark high-resolution image data sets	2018	IEEE Geoscience and Remote Sensing Letters
Hurt J.A., Scott G.J., Anderson D.T., Davis C.H.	Benchmark Meta-Dataset of High-Resolution Remote Sensing Imagery for Training Robust Deep Learning Models in Machine-Assisted Visual Analytics	2018	Proceedings - Applied Imagery Pattern Recognition Workshop
Shao Z., Yang K., Zhou W.	A benchmark dataset for performance evaluation of multi-label remote sensing image retrieval	2018	Remote Sensing
Shao Z., Yang K., Zhou W.	Correction: A benchmark dataset for performance evaluation of multi-label remote sensing image Retrieval	2018	Remote Sensing
Dumitru C.O., Schwarz G., Datcu M.	SAR Image Land Cover Datasets for Classification Benchmarking of Temporal Changes	2018	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
Battle L., Chang R., Heer J., Stonebraker M.	Position statement: The case for a visualization performance benchmark	2018	2017 IEEE Workshop on Data Systems for Interactive Analysis, DSIA 2017
Fresner J., Krenn C.	Theoretical minimum consumption calculation as starting point for cleaner production option identification as a new approach to benchmarking	2018	Journal of Cleaner Production
Dai S., Zheng X., Zuo S., Ren Y.	Improving the prediction accuracy of forest aboveground biomass benchmark map by integrating machine learning and spatial statistics	2018	Proceedings of Spatial Accuracy 2018
Moreira A., Krieger G., Zink M.	TandEM-X & Tandem-L: Setting benchmarks in radar remote sensing	2018	Proceedings of the International Astronautical Congress, IAC
Munoz-Mari J., Izquierdo-Verdiguier E., Campos-Taberner M., Perez-Suay A., Gomez-Chova L., Mateo-Garcia G., Ruescas A.B., Laparra V., Padron J.A., Amoros-Lopez J., Camps-Valls G.	HyperLabelMe : A Web Platform for Benchmarking Remote-Sensing Image Classifiers	2017	IEEE Geoscience and Remote Sensing Magazine
Kang X., Li S., Benediktsson J.A.	Hyperspectral image classification: A benchmark	2017	International Geoscience and Remote Sensing Symposium (IGARSS)
Maggiori E., Tarabalka Y., Charpiat G., Alliez P.	Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark	2017	International Geoscience and Remote Sensing Symposium (IGARSS)
Barelli L., Paolini P., Forti G.	The XII century towers, a benchmark of the Rome countryside almost cancelled: The safeguard plan by low cost UAV and terrestrial DSM photogrammetry surveying and 3d web GIS applications	2017	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Karmas A., Karantzalos K.	Benchmarking server-side software modules for handling and processing remote sensing data through Rasdaman	2017	Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing

Tits L., Somers B., Stuckens J., Coppin P.	Validating nonlinear mixing models: Benchmark datasets from vegetated areas	2017	Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing
Cheng G., Han J., Lu X.	Remote Sensing Image Scene Classification: Benchmark and State of the Art	2017	Proceedings of the IEEE
Ito A., Nishina K., Reyer C.P.O., François L., Henrot A.-J., Munhoven G., Jacquemin I., Tian H., Yang J., Pan S., Morfopoulos C., Betts R., Hickler T., Steinkamp J., Ostberg S., Schaphoff S., Ciais P., Chang J., Rafique R., Zeng N., Zhao F.	Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: Benchmarking for impact assessment studies	2017	Environmental Research Letters
Xia G.-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L., Lu X.	AID: A benchmark data set for performance evaluation of aerial scene classification	2017	IEEE Transactions on Geoscience and Remote Sensing
Bolting J., Stolle M., Fergani S., Biannic J.-M., Defay F.	An open benchmark for distributed formation flight control of Fixed-Wing Unmanned Aircraft Systems	2017	IFAC-PapersOnLine
Calle M.A.G., Verleysen P., Alves M.	Benchmark study of failure criteria for ship collision modeling using purpose-designed tensile specimen geometries	2017	Marine Structures
van Oosterom P., Martinez-Rubi O., Tijssen T., Gonçalves R.	Realistic benchmarks for point cloud data management systems	2017	Lecture Notes in Geoinformation and Cartography
Koch T., Dangelo P., Kurz F., Fraundorfer F., Reinartz P., Korner M.	The TUM-DLR Multimodal Earth Observation Evaluation Benchmark	2016	IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops
Swales A., Denys P., Pickett V.I., Lovelock C.E.	Evaluating deep subsidence in a rapidly-accreting mangrove forest using GPS monitoring of surface-elevation benchmarks and sedimentary records	2016	Marine Geology
Papadomanolaki M., Vakalopoulou M., Zagoruyko S., Karantzas K.	Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data	2016	ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Talaminos-Barroso A., Estudillo-Valderrama M.A., Roa L.M., Reina-Tosina J., Ortega-Ruiz F.	A Machine-to-Machine protocol benchmark for eHealth applications - Use case: Respiratory rehabilitation	2016	Computer Methods and Programs in Biomedicine
Bonnema M.G., Sikder S., Hossain F., Durand M., Gleason C.J., Bjerklie D.M.	Benchmarking wide swath altimetry-based river discharge estimation algorithms for the Ganges river system	2016	Water Resources Research
Evans I.S., Miki J., Ho J., Robinson D.T.	Benchmarking stereo-derived DSMs to lidar by land cover and slope for resource development in northern Ontario, Canada	2016	International Journal of Remote Sensing
Witharana C., LaRue M.A., Lynch H.J.	Benchmarking of data fusion algorithms in support of earth observation based Antarctic wildlife monitoring	2016	ISPRS Journal of Photogrammetry and Remote Sensing
Hendrickx J.M.H., Allen R.G., Brower A., Byrd A.R., Hong S.-H., Ogden F.L., Pradhan N.R., Robison C.W., Toll D., Trezza R., Umstot T.G., Wilson J.L.	Benchmarking Optical/Thermal Satellite Imagery for Estimating Evapotranspiration and Soil Moisture in Decision Support Tools	2016	Journal of the American Water Resources Association
Gerke M., Nex F., Remondino F., Jacobsen K., Kremer J., Karel W., Huf H., Ostrowski W.	Orientation of oblique airborne image sets - Experiences from the ISPRS/EuroSDR benchmark on multi-platform photogrammetry	2016	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Zhang H., Wang C., Chen Y., Jia F., Li J.	Benchmark on outdoor scenes	2016	Proceedings of SPIE - The International Society for Optical Engineering
Pirotti F., Sunar F., Piragnolo M.	Benchmark of machine learning methods for classification of a Sentinel-2 image	2016	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Nikolov I., Madsen C.	Benchmarking close-range structure from motion 3D reconstruction software under varying capturing conditions	2016	Lecture Notes in Computer Science
Hu J., Jiang T., Tong X., Xia G.-S., Zhang L.	A benchmark for scene classification of high spatial resolution remote sensing imagery	2015	International Geoscience and Remote Sensing Symposium (IGARSS)
Lagrange A., Le Saux B., Beaupere A., Boulch A., Chan-Hon-Tong A., Herbin S., Randrianarivo H., Ferecatu M.	Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks	2015	International Geoscience and Remote Sensing Symposium (IGARSS)
Inglada J., Arias M., Tardy B., Morin D., Valero S., Hagolle O., Dedieu G., Sepulcre G., Bontemps S., Defourny P.	Benchmarking of algorithms for crop type land-cover maps using Sentinel-2 image time series	2015	International Geoscience and Remote Sensing Symposium (IGARSS)
Julien Y., Sobrino J.A.	CloudSim: A fair benchmark for comparison of methods for times series reconstruction from cloud and atmospheric contamination	2015	2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images, Multi-Temp 2015
Tucci G., Bonora V., Conti A., Fiorini L.	Benchmarking range-based and image-based techniques for digitizing a glazed earthenware frieze	2015	ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Mikes S., Haindl M., Scarpa G., Gaetano R.	Benchmarking of Remote Sensing Segmentation Methods	2015	IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
Nex F., Gerke M., Remondino F., Przybilla H.-J., Baumker M., Zurhorst A.	ISPRS benchmark for multi-platform photogrammetry	2015	ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Koch J., Jensen K.H., Stisen S.	Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study	2015	Water Resources Research
Kuchmister J., Ćmielewski K., Gotuch P.	The application of the optoelectronic technique of transferring heights from the recessed benchmarks in networks in the examination of rock mass deformation	2014	Acta Geodynamica et Geomaterialia
Rottensteiner F., Sohn G., Gerke M., Wegner J.D., Breitkopf U., Jung J.	Results of the ISPRS benchmark on urban object detection and 3D building reconstruction	2014	ISPRS Journal of Photogrammetry and Remote Sensing
Guo Y., Zhang J., Lu M., Wan J., Ma Y.	Benchmark datasets for 3D computer vision	2014	Proceedings of the 2014 9th IEEE Conference on Industrial Electronics and Applications, ICIEA 2014

Kuschik G., D'Angelo P., Qin R., Poli D., Reinartz P., Cremers D.	DSM accuracy evaluation for the ISPRS commission i image matching benchmark	2014	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Singha S., Vespe M., Trieschmann O.	Automatic Synthetic Aperture Radar based oil spill detection and performance estimation via a semi-automatic operational service benchmark	2013	Marine Pollution Bulletin
Kelley D.I., Prentice I.C., Harrison S.P., Wang H., Simard M., Fisher J.B., Willis K.O.	A comprehensive benchmarking system for evaluating global vegetation models	2013	Biogeosciences
Revercomb H., Best F., Knuteson R., Tobin D., Taylor J., Gero J.	Status of high spectral resolution IR for advancing atmospheric state characterization and climate trend benchmarking: A period of both opportunity realized and squandered	2013	AIP Conference Proceedings
Mikeš S., Haindl M., Scarpa G.	Remote sensing segmentation benchmark	2012	2012 IAPR Workshop on Pattern Recognition in Remote Sensing, PRRS 2012
Abuzar M., McAllister A., Whitfield D., Sheffield K.	Satellite-based measurements for benchmarking regional irrigation performance in Goulburn-Murray catchment	2012	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Zhang K., Wu S., Sang J., Norman R., Carter B., Bennett J.	Platform technologies for space atmosphere and climate - A benchmark geospatial project from the Australian space research program	2012	CEUR Workshop Proceedings
Kaartinen H., Kukko A., Hyypä J., Jaakkola A.	Benchmarking mobile laser scanning systems using a permanent test field	2012	International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences
Roberts R.S., Pope P.A., Vatsavai R.R., Jiang M., Arrowood L.F., Trucano T.G., Gleason S., Cheriyyadath A., Sorokine A., Katsagelos A.K., Pappas T.N., Gaines L.R., Chilton L.K.	Design of benchmark imagery for validating facility annotation algorithms	2011	International Geoscience and Remote Sensing Symposium (IGARSS)
Wang S., Jiang S., Ning H.	The effect of benchmark change to the cylinder index of diesel engine	2011	2011 International Conference on Remote Sensing, Environment and Transportation Engineering, RSETE 2011 - Proceedings
Sabater N., Blanchet G., Moisan L., Almansa A., Morel J.-M.	Review of low-baseline stereo algorithms and benchmarks	2010	Proceedings of SPIE - The International Society for Optical Engineering
Zwart S.J., Bastiaanssen W.G.M., de Fraiture C., Molden D.J.	A global benchmark map of water productivity for rainfed and irrigated wheat	2010	Agricultural Water Management
Dziegielewski B., Kiefer J.C.	Appropriate design and evaluation of water use and conservation metrics and benchmarks	2010	Journal / American Water Works Association
Anagnostou E.N., Maggioni V., Nikolopoulos E.I., Meskele T., Hossain F., Papadopoulos A.	Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates	2010	IEEE Transactions on Geoscience and Remote Sensing
Brenning A.	Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection	2009	Remote Sensing of Environment
Cai X.L., Thenkabail P.S., Platonov A.	Benchmarking cotton water use and productivity using spectral indices	2008	29th Asian Conference on Remote Sensing 2008, ACRS 2008
Jäger U., Höpken M., Dürr B., Metzler J., Willersinn D.	Multisensor benchmark data for riot control	2008	Proceedings of SPIE - The International Society for Optical Engineering
Widlowski J.-L., Pinty B.	Canopy reflectance model benchmarking: RAMI and the ROMC	2008	Lecture Notes in Computational Science and Engineering
González-Dugo M.P., Mateos L.	Spectral vegetation indices for benchmarking water productivity of irrigated cotton and sugarbeet crops	2008	Agricultural Water Management
Bellmann A., Hellwich O., Rodehorst V., Yilmaz U.	A benchmarking dataset for performance evaluation of automatic surface reconstruction algorithms	2007	Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition
Suresh Babu A.V., Venkateshwar Rao V., Muralikrishna I.V.	Satellite remote sensing derived spatial water utilisation index (wui) for benchmarking of irrigation systems	2007	Journal of the Indian Society of Remote Sensing
Mannucci A.J., Ao C.O., Yunc K.T.P., Young L.E., Hajj G.A., Iijima B.A., Kuang D., Meehan T.K., Leroy S.S.	Generating climate benchmark atmospheric soundings using GPS occultation data	2006	Proceedings of SPIE - The International Society for Optical Engineering
Christophe E., Léger D., Mailhes C.	Quality criteria benchmark for hyperspectral imagery	2005	IEEE Transactions on Geoscience and Remote Sensing
Anderson J.G., Dykema J.A., Goody R.M., Hu H., Kirk-Davidoff D.B.	Absolute, spectrally-resolved, thermal radiance: A benchmark for climate monitoring from space	2004	Journal of Quantitative Spectroscopy and Radiative Transfer
Bastiaanssen W.G.M., Brito R.A.L., Bos M.G., Souza R.A., Cavalcanti E.B., Bakker M.M.	Low cost satellite data for monthly irrigation performance monitoring: Benchmarks from Nilo Coelho, Brazil	2001	Irrigation and Drainage Systems
Thenkabail P.S., Nolte C., Lyon J.G.	Remote sensing and GIS modeling for selection of a benchmark research area in the inland valley agroecosystems of West and Central Africa	2000	Photogrammetric Engineering and Remote Sensing
Le Marshall J.F., Simpson J.J., Jin Z.	Satellite calibration using a collocated nadir observation technique: Theoretical basis and application to the GMS-5 pathfinder benchmark period	1999	IEEE Transactions on Geoscience and Remote Sensing
Thenkabail P.S.	Characterization of the alternative to slash-and-burn benchmark research area representing the congolese rainforests of africa using near-real-time spot hrv data	1999	International Journal of Remote Sensing
Stonebraker M., Frew J., Gardels K., Meredith J.	The SEQUOIA 2000 storage benchmark	1993	ACM SIGMOD Record
Narayanan R.M., Green S.E., Alexander D.R.	Mid-Infrared Backscatter Characteristics of Various Benchmark Soils	1992	IEEE Transactions on Geoscience and Remote Sensing
Morse Anthony	Benchmark data on the separability between orchards and vineyards in the southern San Joaquin valley of California.	1983	Technical Papers of the American Society of Photogrammetry,