



**HAL**  
open science

# Aligning Geospatial AI for Disaster Relief with The Sphere Handbook

Sagar Verma, Siddharth Gupta, Kavya Gupta

► **To cite this version:**

Sagar Verma, Siddharth Gupta, Kavya Gupta. Aligning Geospatial AI for Disaster Relief with The Sphere Handbook. 2022. <hal-03861146>

**HAL Id: hal-03861146**

**<https://hal.science/hal-03861146v1>**

Preprint submitted on 19 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

---

# Aligning Geospatial AI for Disaster Relief with The Sphere Handbook

---

**Sagar Verma**  
Granular AI  
sagar@granular.ai

**Siddharth Gupta**  
Granular AI  
sid@granular.ai

**Kavya Gupta**  
Université Paris-Saclay, CentraleSupélec, Inria, Centre de Vision Numérique  
kavya.gupta100@gmail.com

## Abstract

The Sphere handbook and its core premise of *right to life with dignity* have been broadly adopted, establishing a standard operating procedure for global humanitarian intervention. Plenty of machine learning methods aim to aid in disaster relief. While performing exceptionally on a machine learning task, these methods fail to deliver targeted effort to the victims of natural disasters. We argue that this is due to the misalignment of such methods with real-world relief practices. This paper presents the alignment of the Sphere guidelines with Geospatial AI solutions. We show several limitations in different machine learning methods proposed for disaster relief in recent years. We take the case of WASH requirements during flood disasters, extend these models to align with Sphere guidelines, and build a solution that has a much better potential to serve individuals stuck in disasters.

## 1 Application Context

Natural hazards have the potential to cause catastrophic damage and significant socioeconomic loss. The actual damage and loss observed in the recent decades have shown an increasing trend. As a result, disaster managers must take a growing responsibility to protect their communities by developing efficient management strategies. Some set principles or guidelines should govern these strategies. The Sphere Handbook is such a guideline coming out of The Sphere Project<sup>1</sup> based on two core beliefs quoted as:

- *People affected by disaster or conflict have the right to life with dignity and, therefore, the right to assistance; and*
- *All possible steps should be taken to alleviate human suffering arising out of disaster or conflict.*

In the last few years, we have seen rapid advances in using Geospatial AI techniques for disaster management. Most of these techniques use deep learning methods to automatically extract information from satellite imagery [1–4]. These methods extract information like building and road damages, flood maps, population maps, and wildfire maps. It has been claimed that such derived information can help automate decision-making during a disaster for quick and better response. In the last three years of HADR workshops, there have been 13 papers proposing building damage assessment, 3 proposing flood mapping, and 6 proposing better ways of relief management. The context of

---

<sup>1</sup><https://spherestandards.org/handbook-2018/>

this paper is to check the alignment of existing methods with the actual need of disaster victims. We analyze these methods' disaster time usability with domain experts at US Air Force Research Laboratory (AFRL). We find that the current research is directed toward obtaining better metrics that satisfy machine learning criteria and are not very useful at achieving the core beliefs mentioned above.

## 2 Sphere AI Guidelines

The Sphere Project was started in 1997 by a group of humanitarian non-governmental organizations and the Red Cross and Red Crescent Movement. They aim to improve the quality of humanitarian responses and be accountable for actions. The core beliefs of Sphere includes : 1) the right to life with dignity, 2) the right to humanitarian assistance, and 3) the right to protection and security. Humanitarian standards are developed by humanitarian practitioners with specific areas of expertise and are formulated based on evidence, experience, and learning. The standards are developed by consensus and reflect accumulated best practices at a global level. They are revised regularly to incorporate developments across the sector. Humanitarian standards are valuable for practitioners to plan, implement, and evaluate humanitarian responses. They provide a solid basis to advocate for the rights of crisis-affected communities. By applying them in preparedness and response operations, humanitarian agencies and practitioners commit to quality and make themselves accountable to the populations they serve.

The Sphere handbook sets principles and minimum humanitarian standards in four technical areas of humanitarian response: 1) Water supply, sanitation, and hygiene promotion (WASH), 2) Food security and nutrition, 3) Shelter and settlement, and 4) Health. It also discusses the vulnerabilities, gaps, and difficulties from various perspectives that need to be considered and taken care of for an effective humanitarian response. The presented standards and guidance in the Sphere Handbook are supported by research and evidence that reflect 20 years of field testing by humanitarians worldwide. The principal users of the Sphere Handbook are practitioners involved in planning, managing, and implementing a humanitarian response. They identify the key indicators in all the four technical areas necessary for providing minimum standards in case of humanitarian response. For example, in the case of WASH, key indicators include "Percentage of water quality tests meeting minimum water quality standards" and "Percentage of toilets reported as safe by women and girls." These key indicators are essential in pre- and post-disaster scenarios. Also, Sphere identifies the needs and requirements of the people at both individual and community levels. Aligning the guidelines in the Sphere handbook with Satellite imagery, AI can help identify water points, shelter points, waste disposal units, disease breeding sites, difficult-to-access areas, and nearby health centers. Engaging and following such guidelines will help provide more pragmatic and useful solutions.

In the AI community, technologists and researchers are moving towards developing a more socially responsible AI, which is of utmost importance when deploying the systems for areas as critical and vulnerable as HADR. The social responsibility of AI includes efforts devoted to addressing technical and societal issues. The four fundamental responsibilities of social AI [5] include 1) functional, 2) legal, 3) ethical, and 4) philanthropic. For the most part, AI practice considers the fundamental responsibilities of AI systems to be functional and legal, whereas ethical and philanthropic responsibilities are somewhat neglected in practice. The Sphere handbook can be an excellent starting point for AI practitioners to align their upcoming work with humanitarian response's ethical, legal, and practical basis. This will help in developing systems that are good *AI Citizens*. Sphere handbook also discusses Humanitarian responses in different contexts and under various circumstances, such as the differences across populations and diversity among individuals, operational and logistical realities that will affect how and what kind of humanitarian response is delivered. Considering these facts, it is also crucial for AI systems to keep up with their functional responsibilities and make more generalized and robust systems.

## 3 Current Status of Geospatial AI in Disaster Relief

Artificial Intelligence (AI), specifically machine vision and deep learning, allow us to interpret imagery at scale in near-real-time [6–8]. Using AI centered tools like GeoEngine [9, 10], we can further increase humanitarian capacity building by simplifying access to ground truth data. This is particularly important after a disaster when compromised physical infrastructure could threaten

ground responders and limit access to lines of communication. Remotely sensed data can provide immediate data to triage and plan. It can also provide pre-disaster baseline data. Further, autonomous analysis techniques can introduce novel spatial metrics for a nuanced understanding of a given region’s condition. Table 1 lists out some very famous datasets and methods proposed in last few years that aim at disaster relief and management.

Method/Dataset	Disasters	ML Task	Alignment
xView2 (xBD) [2]	E F L V W	Detection	No
CDF [11]	E F L V W	Detection	No
Benosn et al., 2020 [12]	E F L V W	Concept Drift	Yes
Khvedchenya et al., 2021 [13]	E F L V W	Detection	No
Xia et al., 2021 [14]	E F L V W	Detection	No
Shin et al., 2021 [15]	E F L V W	Explainability	Yes
Kuzin et al., 2021 [16]	E F L V W	Human-in-the-loop	Yes
FloodNet [1]	F	Detection	No
SpaceNet-8 [17]	F	Detection	No
Franceschini et al., 2021 [18]	F	Acquisition	Yes
Ochoa et al., 2021 [19]	E F	Detection	No
Seo et al., 2019 [20]	E	Data Imbalance	No
BRAILS [21]	E	Soft-Story Buildings	No
Xu et al., 2019 [4]	E	Detection	No
Boin et al., 2020 [22]	E	Data Imbalance	No
Lee et al., 2020 [23]	E W	Detection	No
FireNet [24]	W	Fire Perimeter	No
Huot et al., 2020 [25]	W	Wildfire Prediction	No

Table 1: This table shows what type of machine learning problems existing geospatial datasets and methods try to solve in disaster relief and management goals. It also shows if they align with the guidelines in The Sphere Handbook. Following disaster types have been considered: E arthquake/Tsunami, F lood, L andslide, V olcanic Eruption, and W ildfire.

Research community started with the intention to detect after-effects of disasters like damage assessment and flood mapping. The goal has been to automatically analyze a large swath of area in the least possible time with very little human intervention. Datasets like OSCD [6, 7] and QFabric [8] enable training change detection models to detect construction and destruction of man made structures. Datasets like xView2 [2] and FloodNet [1] enable training large deep learning models for building damage detection and classification. Several methods have been proposed on these datasets that aim to improve the F1-score [11, 13, 14] or solve some niche problems [12, 15, 16].

Benson et al. [12] argues that models trained on datasets like xView2 perform worse when new images come, and they try to solve the problem of concept drift. This is a critical problem to understand and resolve. This has also been discussed in the Sphere handbook as the problem where a disaster occurs first time in a new area for which existing models do not exist or are bound to fail. Similarly, [16] proposed using human-in-the-loop to adapt existing damage detection models to a new area by using sparse inputs in the form of point labels by experts available on the ground. Decision makers on the ground during disaster relief and management are key people, and as much we would like to use machine learning based methods to make things fast, their eyes on such ML outputs are quite important. Explainable AI comes to the rescue for such requirements as shown by [15]. Another important factor the Sphere handbook mentions is that different places have different rules on data acquisition and decimation. Furthermore, high-quality data is often unavailable and expensive to acquire, especially in the case of aerial images. To solve such problems often, small unmanned aerial vehicles (sUAV) mounted with inexpensive handheld cameras like GoPro are used [18]. One crucial technical area of humanitarian response is water supply and hygiene promotion (WASH). [3] proposes leveraging satellite images and public data to map water availability and population dependency during the Venezuelan migration crisis.

After analysing all the current papers specially the one published in last three years of HADR, we found that most methods are focused on detecting building damage. Only Tingzon et al. [3] work on WASH preparedness using satellite imagery. Building damage assessment methods provide very macro analysis of disaster destruction. They do not provide intervention-level insights aligned with the need of policymakers and development organizations. We propose that Geospatial AI capabilities be extended/mapped to align critical indicators in the Sphere handbook.

## 4 Case Study of WASH

In this study we aim to calculate key indicators for water safety during flooding. We use guidelines mentioned in WASH technical area in the Sphere handbook. For this we collect SkySat and Pleiades satellite images (50cm resolution) of flooded regions in Pakistan between July - September 2022. To date, more than 1.14 million houses have been damaged and over 765,000 houses have been destroyed across the country, while over 1,500 deaths and more than 12,800 injuries were recorded since mid-June. Figure 1 shows satellite images captured over Sehwan Sharif airport before and after the floods. Before and after images for two more cities, Kairpur and Qambar were also collected.

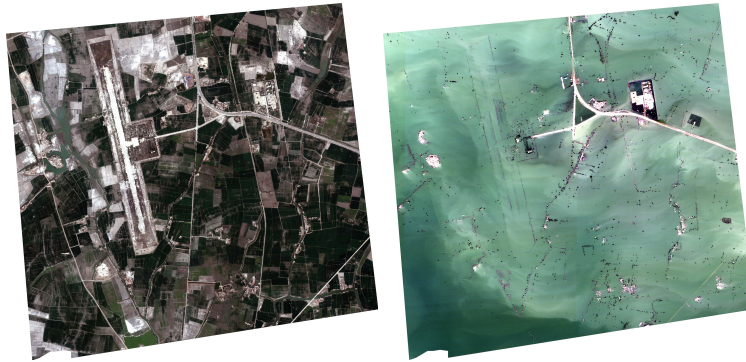


Figure 1: Sehwan Sharif Airport and nearby villages captured on 2022-01-27 and after the flood on 2022-09-24.

### 4.1 Key Indicators for Water Safety

The Sphere handbook sets certain indicators for water safety under the technical area of WASH. We aim to derive these indicators from satellite imagery using existing deep learning methods. Following are the key indicators for water safety:

- Average volume of water used for drinking and domestic hygiene per household. Minimum of 15 liters per person per day and determine quantity based on context and phase of the response.
- Maximum number of people using the water-based facility. 250 people per tap, 500 people per hand pump, 400 people per open hand well, 100 people per laundry facility, and 50 people per bathing facility.
- Percentage of household income used to buy water for drinking and domestic hygiene: target 5 percent or less.
- Percentage of targeted households who know where and when they will next get their water.
- Distance from any household to the nearest water point: <500 meters.
- Queuing time at water sources: <30 minutes.
- Percentage of communal water distribution points free of standing water.
- Percentage of water systems/facilities that have functional and accountable management system in place.

## 4.2 Flood Detection

The first step in obtaining the above key indicators is identifying houses, schools, drinking water bodies, and flooded areas in satellite images. We use BiDate UNet [6, 7] trained on xView2 dataset [2] to detect buildings and classify their damage levels. xView2 dataset only detects buildings and their damage levels. We also train flood foundation network available as SpaceNet-8 baseline<sup>2</sup> to get flood maps. SpaceNet-8 contains images from New Orleans, Louisiana, following Hurricane Ida in August 2021 and Dernau, Germany, during the June 2021 floods across Western Europe. The trained networks turned out to be heavily biased toward the urban planning of these cities and failed to give results for the small towns and villages of Pakistan.

## 4.3 Other Essential Information

Population maps were derived for Shewan, Qambar, and Khairpur from Data for Good<sup>3</sup>. To obtain existing locations where safe water can be obtained we look into mWater<sup>4</sup> and UN Wash<sup>5</sup> projects. UN Wash map only shows country-level metrics around the availability of safe water during normal times. mWater shows a more up-to-date map of safe water availability but is available for only 0.1% area of Pakistan.

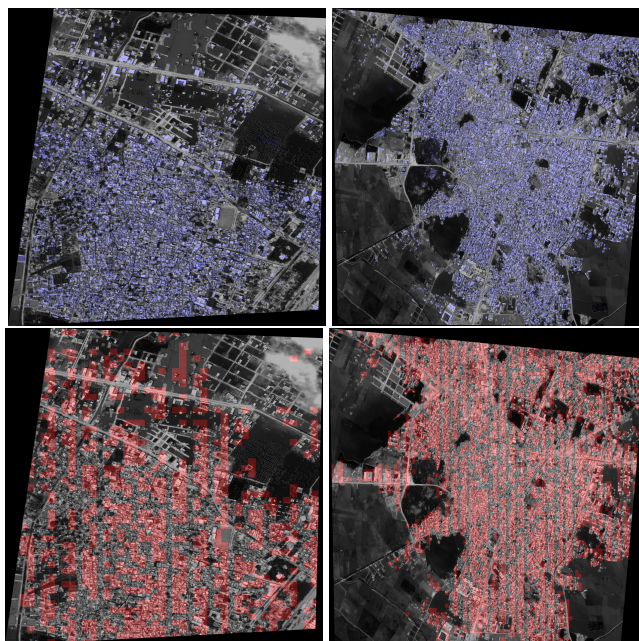


Figure 2: Derived information from satellite images. The left column shows Khairpur, and the right shows Qambar. The top row shows detected buildings, and the bottom row shows population density per  $30m \times 30m$  grid.

Figure 2 shows information derived using the trained model and Data for Good population maps. We manually identify government buildings to obtain vital areas where safe water can be available during a flood. Qambar and Khairpur have 5 and 2 government buildings within the observed area. We use all this information to calculate key indicators. Within the satellite image captured for Qambar (6.04 km-square) and Khairpur (1.68 km-square), there are approximately 61187 and 11664 people, respectively. Assuming that all the government buildings have safe tap water available, 12237 and 5832 people depend on them. This is way above then the set of 250 people per tap water. The average distance from the household is 350 meters and 410 meters for Qambar and Khairpur, respectively.

<sup>2</sup><https://github.com/SpaceNetChallenge/SpaceNet8>

<sup>3</sup><https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps>

<sup>4</sup><https://bit.ly/3rrMecs>

<sup>5</sup><https://washdata.org/data/household>

## 5 Conclusion

This paper presents the gap between the Sphere handbook’s international guidelines and current geospatial AI HADR research. We briefly describe the Sphere handbook and how it can help us design better geospatial AI solutions. Using existing methods and derived information to calculate water safety (WASH) key indicators for two cities in Pakistan for 2022 floods. We hope this study will steer the HADR research to align with international guidelines leading to highly focused solutions for individuals affected by disasters.

## References

- [1] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. Murphy, “Floodnet: A high resolution aerial imagery dataset for post flood scene understanding,” *IEEE*, 2020.
- [2] R. Gupta, B. Goodman, N. Patel, R. Hosfelt, S. Sajeev, E. Heim, J. Doshi, K. Lucas, H. Choset, and M. Gaston, “Creating xbd: A dataset for assessing building damage from satellite imagery,” in *CVPRW*, 2019.
- [3] N. Dejito, R. A. Flores, R. de Guzman, I. Tingzon, L. Carvajal, A. J. R. Aroca, and C. Delgado, “Mapping access to water and sanitation in colombia using publicly accessible satellite imagery, crowd-sourced geospatial information and randomforests,” *HADR*, 2021.
- [4] J. Z. Xu, W. Lu, Z. Li, P. Khaitan, and V. Zaytseva, “Building damage detection in satellite imagery using convolutional neural networks,” *HADR*, 2019.
- [5] L. Cheng, K. R. Varshney, and H. Liu, “Socially responsible ai algorithms: Issues, purposes, and challenges,” *JAIR*, 2021.
- [6] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in *IGARSS*, 2018.
- [7] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzalos, “Detecting urban changes with recurrent neural networks from multitemporal sentinel-2 data,” in *IGARSS*, 2019.
- [8] S. Verma, A. Panigrahi, and S. Gupta, “Qfabric: Multi-task change detection dataset,” in *CVPRW*, 2021.
- [9] S. Verma, S. Gupta, H. Shin, A. Panigrahi, S. Goswami, S. Pardeshi, N. Exe, U. Dutta, T. R. Joshi, and N. Bhojwani, “Geoengine: A platform for production-ready geospatial research,” in *CVPR Demo*, 2022.
- [10] H. Shin, N. Exe, U. Dutta, T. R. Joshi, S. Verma, and S. Gupta, “Europa: Increasing accessibility of geospatial datasets,” in *IGARSS*, 2022.
- [11] Y. Shen, S. Zhu, T. Yang, and C. Chen, “Cross-directional feature fusion network for building damage assessment from satellite imagery,” *HADR*, 2020.
- [12] V. Benson and A. S. Ecker, “Assessing out-of-domain generalization for robust building damage detection,” *HADR*, 2020.
- [13] E. Khvedchenya and T. Gabruseva, “Fully convolutional siamese neural networks for buildings damage assessment from satellite images,” *HADR*, 2021.
- [14] J. Xia, N. Yokoya, and B. Adriano, “Building damage mapping with self-positive unlabeled learning,” *HADR*, 2021.
- [15] D. Shin, S. Grover, K. Holstein, and A. Perer, “Characterizing human explanation strategies to inform the design of explainable ai for building damage assessment,” *HADR*, 2021.
- [16] D. Kuzin, O. Isupova, B. D. Simmons, and S. Reece, “Disaster mapping from satellites: damage detection with crowdsourced point labels,” *HADR*, 2021.

- [17] R. Hänsch, J. Arndt, D. Lunga, M. Gibb, T. Pedelose, A. Boedihardjo, D. Petrie, and T. M. Bacastow, “Spacenet 8 - the detection of flooded roads and buildings,” in *CVPRW*, 2022.
- [18] R. G. Franceschini, J. Liu, and S. Amin, “Damage estimation and localization from sparse aerial imagery,” *ICMLA*, 2021.
- [19] K. S. Ochoa, “Creating a coefficient of change in the built environment after a natural disaster,” *HADR*, 2021.
- [20] J. Seo, S. Lee, B. Kim, and T. Jeon, “Revisiting classical bagging with modern transfer learning for on-the-fly disaster damage detector,” *HADR*, 2019.
- [21] Q. Yu, C. Wang, B. Cetiner, S. X. Yu, F. McKenna, E. Taciroglu, and K. H. Law, “Building information modeling and classification by visual learning at a city scale,” *HADR*, 2019.
- [22] J.-B. Boin, N. Roth, J. Doshi, P. Lluca, and N. Borensztein, “Multi-class segmentation under severe class imbalance: A case study in roof damage assessment,” *HADR*, 2020.
- [23] J. Lee, J. Z. Xu, K. Sohn, W. Lu, D. Berthelot, I. Gur, P. Khaitan, K. Huang, K. M. Koupparis, and B. Kowatsch, “Assessing post-disaster damage from satellite imagery using semi-supervised learning techniques,” *HADR*, 2020.
- [24] J. Doshi, D. Garcia, C. Massey, P. Lluca, N. Borensztein, M. Baird, M. Cook, and D. Raj, “Firenet: Real-time segmentation of fire perimeter from aerial video,” *HADR*, 2019.
- [25] F. Huot, R. L. Hu, M. Ihme, Q. Wang, J. Burge, T. Lu, J. Hickey, Y.-F. Chen, and J. R. Anderson, “Deep learning models for predicting wildfires from historical remote-sensing data,” *HADR*, 2020.