British Geological Survey

ImageCat

# METEOR: Exposure Data Classification, Metadata Population and Confidence Assessment. Report M3.2/P

UKSA IPP2 Grant Programme

Open File Report OR/22/024

BRITISH GEOLOGICAL SURVEY

UKSA IPP2 GRANT PROGRAMME

OPEN FILE REPORT OR/22/024

# METEOR: Exposure Data Classification, Metadata Population and Confidence Assessment. Report M3.2/P

C. Huyck, Z. Hu, P. Amyx, G. Esquivias, M. Huyck, M. Eguchi

*Contributor/editor*

K. Smith, C. Jordan, P. Henshaw, S. Ghosh, R. Eguchi, V. Silva, Y.J. Lee

*Keywords*

Report; METEOR; Exposure; Metadata; Confidence.

*Front cover*

METEOR (Modelling Exposure Through Earth Observation Routines) Project logo

*Bibliographical reference*

HUYCK, C., HU, Z., AMYX, P., ESQUIVIAS, G., HUYCK, M., EGUCHI, M. 2019. METEOR: Exposure Data Classification, Metadata Population and Confidence Assessment. Report M3.2/P. *British Geological Survey Open File Report*, OR/22/024. 77pp.



Keyworth, Nottingham British Geological Survey 2019

## BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

*The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.*

*The British Geological Survey is a component body of UK Research and Innovation.*

*British Geological Survey offices*


**Nicker Hill, Keyworth,**
**Nottingham NG12 5GG**

Tel 0115 936 3100

**BGS Central Enquiries Desk**

Tel 0115 936 3143
email enquiries@bgs.ac.uk

**BGS Sales**

Tel 0115 936 3241
email sales@bgs.ac.uk

**The Lyell Centre, Research Avenue South,**
**Edinburgh EH14 4AP**

Tel 0131 667 1000
email scotsales@bgs.ac.uk

**Natural History Museum, Cromwell Road,**
**London SW7 5BD**

Tel 020 7589 4090
Tel 020 7942 5344/45
email bgslondon@bgs.ac.uk

**Cardiff University, Main Building, Park Place,**
**Cardiff CF10 3AT**

Tel 029 2167 4280

**Maclean Building, Crowmarsh Gifford,**
**Wallingford OX10 8BB**

Tel 01491 838800

**Geological Survey of Northern Ireland, Department of Enterprise, Trade & Investment, Dundonald House, Upper Newtownards Road, Ballymiscaw,**
**Belfast, BT4 3SB**

Tel 01232 666595
www.bgs.ac.uk/gsni/

**Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU**

Tel 01793 411500 Fax 01793 411501
www.nerc.ac.uk

**UK Research and Innovation, Polaris House,**
**Swindon SN2 1FL**

Tel 01793 444000
www.ukri.org

Website www.bgs.ac.uk
Shop online at www.geologyshop.com

# Document Verification

| | |
|---|---|
| Project | METEOR: Modelling Exposure Through Earth Observation Routines |
| Report Title | METEOR: Exposure Data Classification, Metadata Population and Confidence Assessment |
| Related Milestone | M3.2 |
| Reference as | Huyck, C., Hu, Z., Amyx, P., Esquivias, G., Huyck, M. & Eguchi, M. (2019) METEOR: Exposure Data Classification, Metadata Population, and Confidence Assessment. Report M3.2/P. *British Geological Survey Open File Report,* OR/22/024. 77pp. |
| Release Type | Public / ~~Confidential~~ / ~~Confidential with Embargo Period~~ |

| Prepared by: Contributors | | |
|---|---|---|
| Name(s):<br>C. Huyck<br>Z. Hu<br>P. Amyx<br>G. Esquivias<br>M. Huyck<br>M. Eguchi | Signature(s):<br>C. Huyck (ImageCat)<br>Z. Hu (ImageCat)<br>P. Amyx (ImageCat)<br>G. Esquivias (ImageCat)<br>M. Huyck (ImageCat)<br>M. Eguchi (ImageCat) | Date(s):<br>06/02/2019 |
| Approved by: Project Manager | | |
| Name:<br>Kay Smith | Signature:<br>K. Smith (BGS) | Date:<br>11/02/2019 |
| Approved by: UKSA IPP Project Officer | | |
| Name:<br>Connor McSharry | Signature:<br>C. McSharry (UKSA) | Date:<br>15/03/2019 |

| Date | Version | Alterations | Editor |
|---|---|---|---|
| | | | |
| | | | |

# Glossary

| | |
|---|---|
| ABS | American Bureau of Shipping |
| ARA | Applied Research Associates |
| ASCE | American Society of Civil Engineers |
| ASTM | American Society for Testing and Materials |
| ATC | Applied Technology Council |
| BGS | British Geological Survey: An organisation providing expert advice in all areas of geosciences to the UK government and internationally |
| CAPR | California Action Plan for Reintegration |
| CAT | Catastrophe |
| CATS | Consequences Assessment Tool Set |
| CBD | Commercial Business District |
| CCI | Climate Change Initiative |
| CIESIN | Centre for International Earth Science Information Network |
| CLC | CORINE Land Cover |
| CODA | Code-Oriented Damage Assessment |
| Cs | Design-level base shear |
| CORINE | Coordination of information on the environment |
| DCR | Demand-to-capacity ratio |
| DCW | Digital charts of the world's populated places |
| DF | Damage Factor- An index of earthquake damage to a building structure, defined as the cost repair to pre-earthquake condition, divided by the building replacement value. |
| DHS | Department of Homeland Security |
| DLR | Deutsches Zentrum für Luft- und Raumfahrt e.V. (German Space Agency) |
| DMD | Disaster Management Department, Prime Minister's Office of Tanzania, focused on disaster risk |
| DNB | Day-Night Band |
| DRM | Disaster Risk Management; the application of disaster risk reduction policies and/or strategies |
| DRR | Disaster Risk Reduction; disaster risk reduction is aimed at preventing new and reducing existing disaster risk and managing residual risk, all of which contribute to strengthening resilience and therefore to the achievement of sustainable development |
| EC | European Commission |
| EEA | European Environment Agency |
| EERI | Earthquake Engineering Research Institute |
| EO | Earth Observation; the gathering of information about Earth's physical, chemical and biological systems via remote sensing technologies, usually involving satellites carrying imaging devices |
| EOG | Earth Observation Group |
| EPEDAT | Early Post-Earthquake Damage Assessment Tool |
| ESA | European Space Agency |
| ESRI | Environmental Systems Research Institute |
| FATHOM | Provides innovative flood modelling and analytics, based on extensive flood risk research |
| FEMA | Federal Emergency Management Agency |
| FGDC | Federal Geographic Data Committee |
| GADM | Database of Global Administrative Areas |
| GAUL | Global Administrative Unit Layers |
| GCRF | Global Challenges Research Fund |
| GED4GEM | Global Exposure development for Global Earthquake Model |
| GEF format | GEM exposure file format |

| | |
|---|---|
| GEM | Global Earthquake Model; a non-profit organisation focused on the pursuit of earthquake resilience worldwide |
| GIGO | Garbage-in, garbage-out |
| GIO | GMES Initial Operations |
| GIS | Geographic Information System; a conceptualised framework that provides the ability to capture and analyse spatial and geographic data |
| GHS | Global human settlement |
| GHSL | Global human settlement layer |
| GMES | Global Monitoring for Environment and Security |
| GMM | Ground-motion models |
| GMPE | Ground motion prediction equations |
| GPS | Global Positioning System; a satellite-based radio navigation system owned by the United States government and operated by the United States Space Force |
| GPW | Gridded Population of the World |
| GRUMP | Global Rural Urban Mapping Project products |
| GUF | Global Urban Footprint |
| HAZUS | Hazard U.S. is a GIS-based natural hazard analysis tool |
| HDC | High density clusters |
| HOT | Humanitarian OpenStreetMap Team; a global non-profit organisation that uses collaborative technology to create OSM maps for areas affected by disasters |
| HRSL | High resolution settlement layer |
| HydroSHED | Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales |
| IBC | International Building Code |
| IDCT | Inventory Data Capture Toolkit |
| ImageCat | International risk management innovation company supporting the global risk and catastrophe management needs of the insurance industry, governments and NGOs |
| IMPSA | Global Impervious Surface Area |
| IPP | International Partnership Programme; the UK Space Agency's International Partnership Programme (IPP) is a £30M per year programme, which uses expertise in space-based solutions, applications and capability to provide a sustainable economic or societal benefit to emerging nations and developing economies |
| IRS | Indian Remote Sensing programme |
| ISA | Impervious surface area |
| ISBN | International Standard Book Number |
| ISO | International Standards Organization |
| ISSN | International Standard Serial Number |
| ITV | Insurance to Value ratio |
| JRC | Joint Research Centre |
| LC | Land Cover |
| LDC | Low density clusters |
| LDC | Least Developed Country on the Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee (DAC) list |
| ISRO | Indian Space Research Organisation |
| LRFD | Load-and-resistance-factor design |
| LFRS | Lateral force resisting systems |
| LIDAR | Light detection and ranging |
| MERIS | Medium Resolution Imaging Spectrometer |
| M | Milestone, related to work package deliverable |
| MD | Metadata |
| MDA | MacDonald Dettwiler and Associates |

| | |
|---|---|
| METEOR | Modelling Exposure Through Earth Observation Routines; a three-year project funded by the UK Space Agency to develop innovative application of Earth Observation (EO) technologies to improve understanding of exposure and multihazards impact with a specific focus on the countries of Nepal and Tanzania |
| MH | Multi-hazard |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSB | Microsoft Build Footprints |
| NASA | National Aeronautics and Space Administration |
| NEHRP | National Earthquake Hazards Reduction Program |
| NGA[1] | National Geospatial-Intelligence Agency |
| NGA[2] | Next Generation Attenuation |
| NGDC | National Geophysical Data Centre |
| NGO | Non-Governmental Organisation; organisations which are independent of government involvement |
| NI | National Aeronautics and Space Administration - Indian Space Research Organisation |
| NOAA | National Oceanic and Atmospheric Administration |
| NPP | National Polar-orbiting Partnership |
| NSET | National Society for Earthquake Technology, non-governmental organisation working on reducing earthquake risk in Nepal and abroad |
| NTL | Night-Time Lights |
| ODA | Official Development Assistance; government aid that promotes and specifically targets the economic development and welfare of developing countries |
| OES | Office of Emergency Services |
| OPM | Oxford Policy Management, organisation focused on sustainable project design and implementation for reducing social and economic disadvantage in low-income countries |
| OSM | OpenStreetMap |
| ORNL | Oak Ridge National Laboratory |
| PAGER | Prompt Assessment of Global Earthquakes for Response |
| PEER | Pacific Earthquake Engineering Research Center |
| REGIO-OECD | Regional and Urban Policy Organisation for Economic Co-operation and Development |
| R | Response modification factor for seismic design |
| RC | Reinforced concrete frames |
| RMS | Risk Management Solutions |
| RP | Return Period |
| SAIC | Science Applications International Corporation |
| SAR | Synthetic Aperture Radar |
| SFR | Single family residential |
| SDGs | Sustainable Development Goals; these goals were set up in 2015 by the United Nations General Assembly and are intended to be achieved by the year 2030 |
| SIDD | Spatial Inventory and Damage Data |
| SME | Subject Matter Experts |
| SPOT | Satellite Pour l'Observation de la Terre (Satellite for observation of Earth) |
| SQL | Structured Query Language |
| SRTM | Shuttle Radar Topography Mission |
| T | Building Period |
| UID | Unique Identification code |
| UKSA | United Kingdom Space Agency; an executive agency of the Government of the United Kingdom, responsible for the United Kingdom's civil space programme |
| UN | United Nations |

| | |
|---|---|
| UNWPP | United Nations World Population Prospects |
| URL | Uniform Resource Locator |
| URM | Unreinforced Masonry |
| USD | United States Dollars (currency unit) |
| USGS | United States Geological Survey |
| VIIRS | Visible Infrared Imaging Radiometer Suite |
| VMAP | Vector Map |
| V/W | Shear force over weight. Calculates base shear. |
| WB | World Bank |
| WHE | World Housing Encyclopaedia |
| WP | Work Package; discrete sets of activities within the METEOR Project, each work package is led by a different partner and has specific objectives |
| WSF | World Settlement Footprint |

# Foreword

This report is the published product of a study by ImageCat as part of the Modelling Exposure Through Earth Observation Routines (METEOR) project led by British Geological Survey (BGS).

METEOR is grant-funded by the UK Space Agency's International Partnership Programme (IPP), a >£150 million programme which is committed to using the UK's space sector research and innovation strengths to deliver sustainable economic, societal, and environmental benefit to those living in emerging and developing economies. IPP is funded from the Department for Business, Energy and Industrial Strategy's (BEIS) Global Challenges Research Fund (GCRF). This £1.5 billion Official Development Assistance (ODA) fund supports cutting-edge research and innovation on global issues affecting developing countries. ODA-funded activity focuses on outcomes that promote long-term sustainable development and growth in countries on the OECD Development Assistance Committee (DAC) list. IPP is ODA compliant, being delivered in alignment with UK Aid Strategy and the United Nations' (UN) Sustainable Development Goals (SDGs).

The objective of this report is to summarise the use of data for exposure development, development patterns for Tanzania, data validation, levels of data collection and impact on loss estimates as well as a preliminary assessment of confidence.

# Contents



**FIGURES**

**TABLES**

# Summary

This report describes a specific piece of work conducted by ImageCat as part of the METEOR (Modelling Exposure Through Earth Observation Routines) project, led by British Geological Survey (BGS) with collaborative partners Oxford Policy Management Limited (OPM), SSBN Limited, The Disaster Management Department, Office of the Prime Minister – Tanzania (DMD), The Global Earthquake Model Foundation (GEM), The Humanitarian OpenStreetMap Team (HOT), ImageCat and the National Society for Earthquake Technology (NSET) – Nepal.

The 3-year project was funded by UK Space Agency through their International Partnership Programme, details of which can be located in the Foreword, and was completed in 2021.

The project aimed to provide an innovative solution to disaster risk reduction, through development of an innovative methodology of creating exposure data from Earth Observation (EO) imagery to identify development patterns throughout a country and provide detailed information when combined with population information. Level 1 exposure was developed for all 47 least developed countries on the OECD DAC list, referred to as ODA least-developed countries in the METEOR documentation, with open access to data and protocols for their development. New national detailed exposure and hazard datasets were also generated for the focus countries of Nepal and Tanzania and the impact of multiple hazards assessed for the countries. Training on product development and potential use for Disaster Risk Reduction was performed within these countries with all data made openly available on data platforms for wider use both within country and worldwide.

This report (M3.2/P) is the second report generated by ImageCat for the work package EO data for exposure development (WP3) led by ImageCat. The other 7 METEOR work packages included, Project Management (WP1 – led by BGS), Monitoring and Evaluation (WP2 – led by OPM), Inputs and Validation (WP4 – led by HOT), Vulnerability and Uncertainty (WP5 - led by GEM), Multiple hazard impact (WP6 – led by BGS), Knowledge sharing (WP7 – led by GEM) and Sustainability and capacity building (WP8 – led by ImageCat).

# 1 METEOR Project

## 1.1 PROJECT SUMMARY

| | |
|---|---|
| Project Title | Modelling Exposure Through Earth Observation Routines (METEOR): EO-based Exposure, Nepal and Tanzania |
| Starting Date | 08/02/2018 |
| Duration | 36 months |
| Partners | UK Partners: The British Geological Survey (BGS) (Lead), Oxford Policy Management Limited (OPM), SSBN Limited<br><br>International Partners: The Disaster Management Department, Office of the Prime Minister – Tanzania, The Global Earthquake Model (GEM) Foundation, The Humanitarian OpenStreetMap Team (HOT), ImageCat, National Society for Earthquake Technology (NSET) – Nepal |
| Target Countries | Nepal and Tanzania for “level 2” results and all 47 Least Developed ODA countries for “level 1” data |
| IPP Project | IPPC2_07_BGS_METEOR |

*Table 1: METEOR Project Summary*

## 1.2 PROJECT OVERVIEW

At present, there is a poor understanding of population exposure in some Official Development Assistance (ODA) countries, which causes major challenges when making Disaster Risk Management decisions. Modelling Exposure Through Earth Observation Routines (METEOR) takes a step-change in the application of Earth Observation exposure data by developing and delivering more accurate levels of population exposure to natural hazards. METEOR is delivering calibrated exposure data for Nepal and Tanzania, plus ‘Level-1’ exposure for the remaining Least developed Countries (LDCs) ODA countries. Moreover, we are: (i) developing and delivering national hazard footprints for Nepal and Tanzania; (ii) producing new vulnerability data for the impacts of hazards on exposure; and (iii) characterising how multi-hazards interact and impact upon exposure. The provision of METEOR’s consistent data to governments, town planners and insurance providers will promote welfare and economic development and better enable them to respond to the hazards when they do occur.

METEOR is co-funded through the second iteration of the UK Space Agency’s (UKSA) International Partnership Programme (IPP), which uses space expertise to develop and deliver innovative solutions to real world problems across the globe. The funding helps to build sustainable development while building effective partnerships that can lead to growth opportunities for British companies.

## 1.3 PROJECT OBJECTIVES

METEOR aims to formulate an innovative methodology of creating exposure data through the use of EO-based imagery to identify development patterns throughout a country. Stratified sampling technique harnessing traditional land use interpretation methods modified to characterise building patterns can be combined with EO and in-field building characteristics to capture the distribution of building types. These protocols and standards will be developed for broad application to ODA countries and will be tested and validated for both Nepal and Tanzania to assure they are fit-for-purpose.

Detailed building data collected on the ground for the cities of Kathmandu (Nepal) and Dar es Salaam (Tanzania) will be used to compare and validate the EO generated exposure datasets. Objectives of the project look to: deliver exposure data for 47 of the least developed ODA countries, including Nepal and Tanzania; create hazard footprints for the specific countries; create open protocol; to develop critical exposure information from EO data; and capacity-building of local decision makers to apply data and assess hazard exposure. The eight work packages (WP) that make up the METEOR project are outlined below in section 1.4.

## 1.4 WORK PACKAGES

Outlined below are the eight work packages that make up the METEOR project, which are led by various partners. Table 2 provides an overview of the work packages together with a brief description of what each of the work packages cover.

| Work Package | Title | Lead | Overview |
|---|---|---|---|
| WP.1 | Project Management | BGS | Project management, meetings with UKSA, quarterly reporting and the provision of feedback on project deliverables and direction across primary stakeholders. |
| WP.2 | Monitoring and Evaluation | OPM | Monitoring and evaluation of the project and its impact, using a theory of change approach to assess whether the associated activities are leading to the desired outcome. |
| WP.3 | EO Data for Exposure Development | ImageCat | EO-based data for exposure development, methods and protocols of segmenting/classifying building patterns for stratified sampling of building characteristics. |
| WP.4 | Inputs and Validation | HOT | Collect exposure data in Kathmandu and Dar es Salaam to help validate and calibrate the data derived from the classification of building patterns from EO-based imagery. |
| WP.5 | Vulnerability and Uncertainty | GEM | Investigate how assumptions, limitations, scale and accuracy of exposure data, as well as decisions in data development process lead to modelled uncertainty. |
| WP.6 | Multiple Hazard Impact | BGS | Multiple hazard impacts on exposure and how they may be addressed in disaster risk management by a range of stakeholders. |
| WP.7 | Knowledge Sharing | GEM | Disseminate to the wider space and development sectors through dedicated web-portals and use of the Challenge Fund open databases. |
| WP.8 | Sustainability and Capacity-Building | ImageCat | Sustainability and capacity-building, with the launch of the databases for Nepal and Tanzania while working with in-country experts. |

*Table 2: Overview of METEOR Work Packages*

# 2 Introduction

Advances in computing and the widespread emergence of desktop GIS software began to make loss estimation possible for governments in the mid-1990s. Engineering companies specializing in estimating potential losses for the insurance industry began to court the U.S. federal and state governments as customers following catastrophic events such as Hurricane Hugo and the Northridge Earthquake. Tools such as Consequences Assessment Tool Set (CATS), originally developed by Science Applications International Corporation (SAIC) for military applications, were repurposed for disaster response. EQE® launched EPEDAT shortly before the Northridge Earthquake through the CAPR division, providing California Office of Emergency Services (OES) with a rapid loss estimate of 40 billion dollars. Shortly after, RMS received funding from FEMA to begin programming the earthquake module of HAZUS, followed by the hurricane module from ARA and the flood module provided by EQE. There have been countless applications of these software packages at the state and local level.

Insurers use “CAT models”, which estimate losses probabilistically, usually for tens of thousands of hypothetical events, whereas governments tend to evaluate a hand-full of likely events in detail using “loss estimation” tools. These basic differences in application result in very different building exposure data requirements. For insurance, CAT models are used primarily for pricing insurance or reinsurance for a limited number of properties. The locations of these properties are typically known, as is the occupancy and the amount of insurance purchased for the property. Other factors may or may not be known- such as detailed structural information, the size of the building or portion of the building insured, the number of stories, or the potential maximum replacement cost. Information collected can be quite detailed with “secondary characteristics”, characterizing attributes such as the distance between buildings, the surrounding vegetation, or even the density of nails in the roof. For governments, the focus is the welfare of society as a whole, although potential impacts to essential facilities such as schools and hospitals may be a key concern. Common uses include: 1) exploring the potential magnitude and distribution of fatalities, injuries, building damage, and financial impact for hypothetical scenario events, approaching events ( in the case of hurricanes ), or recent and unfolding events; 2) calculation of average annualised losses or probable losses in a given time frame (50 year event, 100 year event ); 3) benefits of extending or enforcing building codes; 4) benefits of mitigating losses at the building or infrastructure level; and 5) evolving hazards due to climate change or step-changes in knowledge concerning hazards. These analyses need a representation of the “general building stock” to reflect the impact on the entire region.

There are three principle components of risk analysis: exposure, hazard, and vulnerability. In most risk applications hazard and vulnerability subject matter experts (SMEs) help develop the software. Although exposure data drastically impacts the accuracy of the risk assessment, there has not been an equal rigorous assessment of the exposure development process as there has been with hazard or vulnerability. The three-level classification used to describe the level of detail for a loss estimation study are geared towards the incorporation of structural engineering and local hazards, rather than reflecting the exposure data itself. The complex process of representing exposure data in a spatial database through a data fusion process requires careful balancing and simple gaps in the process skew results. For example, for over 10 years the HAZUS block-level exposure data in the flood model represented the hazard in a given block as evenly spread throughout the block- evenly distributing the exposure to the centre of water bodies as delineated in census maps. This spread of exposure, even in an analysis where the “inundation” extent represents a river bank under normal conditions, will lead to a loss- and the average annual losses would reflect this over-exposure. In another example, although the square footage density for a given census tract is a simple calculation, HAZUS applies default mapping schemes to an entire region instead of adjusting the mapping scheme for the urban landscape. Even in intensely urban environments where the square footage of buildings far exceeds the square footage of a census tract (i.e., downtown Los Angeles, Manhattan, coastal Miami, and the San Francisco financial district), HAZUS predicts primarily low-rise buildings. Across the exposure development landscape, simple factors such as an average number of persons per household, the average size of buildings, or the "model building type" used to evaluate replacement cost vary significantly by region. Although the results of a loss estimation study are sensitive to these parameters, they are rarely given as much attention as the vulnerability and hazard in a given study.

Despite this lack of visibility, the last ten years have seen an expansion of tools available for exposure development. Loss estimation has begun to be applied globally and as researchers have expanded their reach into data-poor environments, they have incorporated creative methods to collect or infer important parameters. This has been facilitated in part by global availability of very high-resolution satellite imagery supported by visualisation platforms such as Google maps and Bing maps, interpreted products based on moderate resolution data such as GUF, GHSL, LandScan, and WorldPop, geo-registered images and video such as Google Streetview or Panoramio, and widespread crowdsourced GIS databases- notably OSM building footprints (Huyck et al., 2011, Huyck & Eguchi, 2017). Several global databases can easily be adapted for loss estimation and are readily available, including ShakeCast and GED4GEM. It is not unusual for NGO projects in developing countries to result in nationwide exposure databases that rival commercial databases or the HAZUS database in quality (Jaiswal and Wald, 2008; Earle et al., 2009; Huyck et al., 2011). That said, the quality of an exposure database is difficult to express, particularly with respect to replacement cost (Huyck and Eguchi, 2017). Data fusion methods for exposure development are not well understood, nor is the impact of various methods on uncertainty. This is in part because the impact on results is going to be a complex relationship between diversity of the building stock, spatial dispersion of the hazard itself, size of the underlying tracts or units, and quality of the data in general.

The basic principle that bad input data results in bad modelling results ("Garbage-in, garbage-out" (GIGO)) warrants that exposure data development be taken seriously. A discipline and language of exposure development are needed to guide the community so that there is a common understanding of uncertainty, error, accuracy, and process. Assumptions need to be laid bare.

To that end, this report proposes a taxonomy of exposure data classification. We have borrowed from existing classification schemes, used primarily to reflect the level of building structural classification. We then explore the primary methods that can be used to develop an exposure database for each level. Although the relative levels explored below impact the accuracy of risk analysis, the vintage, quality, and availability of all databases sourced when making a data set have an impact on the accuracy.  It may be more accurate, for example, to use census data for the primary source of building count data, rather than building specific data, if the building specific data is old and there has been significant building development. A more thorough discussion on ranking the accuracy of the data set is explored in subsequent sections.

# 3 Use of EO Data for Exposure Development: An Overview

EO data primarily refers to the use of satellite imagery or remote sensing technology for gathering and monitoring physical and atmospheric information. These data are used for resource management, monitoring land-use change, tracking development, weather forecasting, and monitoring the impacts of natural hazards. In the context of exposure, EO data is useful for understanding the extent or severity at which the human population is exposed to natural hazards or catastrophic events; as well the distribution of building exposure.

The built environment is comprised of buildings, lifelines, and infrastructure. Effective DRR and DRM requires an assessment of probable losses and impacts. Over the past several decades, CAT models have allowed the insurance industry to minimise the fiscal impact of natural hazards by predicting losses to actual and probable events. The insurance industry uses CAT models to assess risk for pricing, to redistribute risk through reinsurance, and to assess the probable economic impacts when an event actually occurs. Assessing risk with CAT models is so pervasive and integral to the decision-making process, that lack of adequate data can be a serious issue in emerging markets, such as developing countries, where exposure is increasing due to foreign or domestic investments and increasing population density. Governments have begun to build much of the same capability through loss estimation models, which are CAT models without insurance calculations (Cutter et al., 2007). Before an event, governments and NGOs run loss estimation models to assess the cost-effectiveness of mitigation strategies and to develop realistic scenarios for emergency planning purposes. For hazards with a slow onset, such as hurricanes and volcanic eruptions, loss estimation tools also provide hypothetical outcomes to probable scenarios. After an event, models estimate the severity and distribution of damage which facilitates critical decisions such as how to allocate resources throughout the affected areas, and in the case of international disasters, how much aid to send to the affected region.

CAT models and loss estimation tools often depend upon GIS exposure databases to characterise the spatial distribution of assets. These datasets are often a source of great uncertainty in the loss estimates, particularly in global disasters. When data are available, it generally is available only for targeted areas, which spatially skews loss estimates and often results in unexpected losses. An independent and spatially consistent EO-based solution that results in more accurate and up-to-date information on building inventories will have a significant impact on characterizing the “true” exposure of risks, especially on a global basis (Eguchi et al., 1999). The decision-making activities that these enhanced building exposure databases support include: a) better understanding and pricing of the risk transfer mechanism for insurers in new and existing markets, b) better risk mitigation, communication, and reduction strategies for government agencies, and c) enhanced capabilities of government and non-profit agencies to perform post-disaster needs assessments. By harnessing EO data the DRR community can characterise the built environment for worldwide for CAT modelling and loss estimation. This section presents an overview of some of the tools and methods available to create EO-derived building exposure databases for use in CAT models and loss estimation tools.

## 3.1 OVERVIEW OF EO DATA USE IN CAT MODELLING

For regional applications, it rarely feasible to directly detect building attributes with EO data. Although there has been considerable applied research into detecting building attributes (such as soft stories or building material) from remotely sensed data (see for example: (Antos et al., 2016), this research will not be reviewed here). The focus of this chapter will be large-scale regional building exposure development, and how this can be facilitated with EO data. The basic premise is that although it is not generally feasible to directly detect building attributes for large regions from EO data, it is possible to identify patterns of building development that correspond to building attributes. These classes can be used to refine statistical distributions of building types by use, and to more effectively distribute the estimated number of buildings to a higher resolution than the administrative level for which they have been provided (or, more often, estimated). For example, if a given county is 50% mountainous terrain and 50% urban environment, instead of distributing the estimated number of buildings evenly throughout the area, buildings might be distributed in the urban area with only a slight remainder spread in the mountainous terrain- consistent with the estimated rural population. Furthermore, if there is a CBD (Commercial Business District) in the middle of the urban area, this region could be delineated, and the number, density, and structural classifications of buildings assumed for this region would be adjusted accordingly.

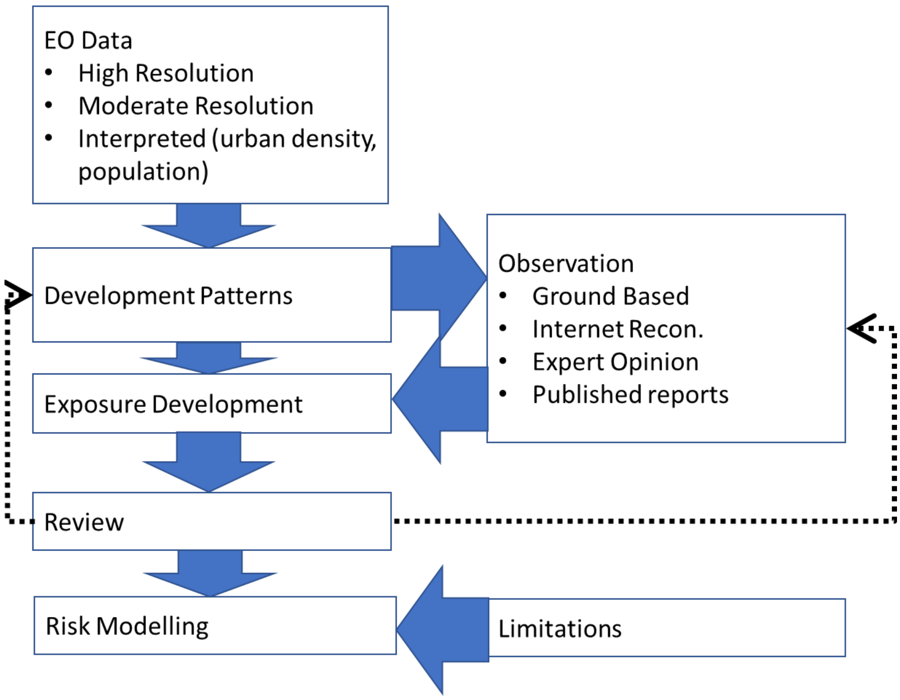


*Figure 3.1: Developing exposure databases with remotely sensed data*

The Figure 3.1 provides a general description of the approach of using EO data in the development of exposure. Using the example above, EO data is first used to distinguish the mountainous, urbanised, and CBD regions. Then statistical assessments of these regions are gathered through any combination of onsite visits, local expertise, research and review of high-resolution images, or internet reconnaissance- depending on time and budget. These statistics are applied to the remote sensing data to extrapolate a cell-based building exposure throughout the entire country. It is then straight forward to map the results, review then in Google Earth, and adjust as needed. The spatial distribution and structural assumptions inherent in the exposure data product developed with this approach are much more accurate than the initial estimate at the county level. This chapter provides a review of this process with regards to the role of EO data.

## 3.2 MODERATE RESOLUTION IMAGERY AND DATA PRODUCTS

In this chapter a range of EO data such as direct multispectral satellite imagery and interpreted or fused data products are discussed. An example of an interpreted data product is LandScan from the Oakridge National Laboratory (ORNL) which provides a global gridded ambient-population data product. Although this population distribution data set is not fully a remote sensing product, LandScan utilises high resolution satellite imagery in combination with demographic information, see Table 3 below for more information, to product a very valuable moderate resolution product that is used in classifying development patterns. The data sources below can be divided into the following categories: 1) Infrared or day-night light band data, 2) urban coverage or settlement density estimates, 3) population estimates, 4) land use, 5) Synthetic Aperture Radar (SAR) and 6) Optical satellite data.

Infrared, or day-night light band data, like the Visible Infrared Imaging Radiometer Suite (VIIRS), are invaluable for distinguishing development patterns in urban environments and can be used in conjunction with the CIESIN Global Rural Urban Mapping Project products (GRUMP), and other sources, to identify rural settlements. Infrared data generally correlates with economic activity, but can vary substantially given weather patterns, aurora borealis, economic conditions, energy conservation norms, and the prevalence of street lights. The infrared radiance signature of some development patterns appears very similar, such as high density residential development and industrial zones, thus the use of various data product combination for supervised classifications and validation. In most areas, however, infrared products are a very powerful data source.

Urban coverage or building density estimates data products provide an excellent source of data for delineating development patterns. If upon reviewing estimates of building density from the Global Human Settlement (GHS) layers or the Global Impervious Surface Area (IMPSA) data appear accurate, for the study region, these data sets can be used in conjunction with infrared or population data for development patterns classification. Extreme care must be taken to assure that these products are applicable to study area as algorithms written to maximise global accuracy are often inaccurate at the local level. At the very least, that can usually be used to assure that urban areas are not placed in low population areas which can occur when the analyst delineates urban areas by only using infrared data. This occurs due to the high infrared detection from highly vegetated or forested land.

Population estimates are typically derived through a combination of the methods as briefly described above. Population distribution data products use information collected by the census bureaus and can also be derived from extracted building estimates. It is not unusual to review population data and find that the distribution bears very little correlation with observed population data in Google Earth. Data quality varies drastically with the political interest of the supplying agency or agencies, i.e. data obtained from a humanitarian or military group. Users are advised to review the data carefully and proceed with caution.

If the study area has “land use or land cover” data available by an international or federal agency, this can provide a head start on the process of identifying development patterns. Users will need to review the land use- land cover data product and map the classifications to development patterns. For example, the land use data product may provide several land classifications such as forest, water, and/or desert, which can be used to delineate an ‘uninhabited’ development pattern sample. Typically, there will be a simple “developed” or “urban” class, which will need to be further delineated with some of the other moderate resolution data sources discussed above. Users are advised, however, that land use data sets may be out of date and not capture recent development.

*Table 3: Specific sources of EO data and data products*

| Source/dataset | Description/Use |
|---|---|
| NASA - NPP- Visible Infrared Imaging Radiometer Suite (VIIRS) | The Visible Infrared Imaging Radiometer Suites (VIIRS) was launched in 2011 on-board the Suomi National Polar-Orbiting Partnership spacecraft. VIIRS empowers operational environmental monitoring and numerical weather forecasting with more than twenty environmental data records including clouds, sea surface temperature, ocean colour, polar wind, vegetation fraction, aerosol, fire, snow and ice, vegetation, , and other applications. The 2015 VIIRS Night-time Light Time Series composite is the latest global annual data set and is key in the remote sensing classification processes of human development and settlement. The Earth Observations Group (EOG) at NOAA/NGDC produces average radiance composite images using night-time data from the VIIRS Day/Night Band (DNB). VIIRS is produced as 15 arc-second (~500m) geographic grids and are made available in geotiff format as a set of 6 tiles. |
| ESA – <br> Sentinel-1 | Sentinel-1 is a C-band SAR constellation that consist of two satellites Senital-1A launched April 3rd 2014 and Sentinel-1B launched April 25th, 2016. It is a collaborative work between the ESA, the European Commission, industry, service providers and data users. Sentinel-1 was designed and built by a consortium of around 60 companies led by Thales Alenia Space and Airbus Defence and Space. There are 4 operational modes: Strip Map has a 5m resolution with 80km swath, Interferometric Wide Swath is 5x20m resolution at 250km swath, Extra Wide Swath is 25x100m resolution with a 400km swath, and Wave Mode is 5x20m resolution and mostly for open ocean. |
| NASA/USGS - LandSat-8 | Landsat 8 launch in 2013 is a landmark of moderate resolution with 15 meters panchromatic, 30 meters multispectral, and 100 meters thermal band resolution capability. The increased resolution from previous satellites makes Landsat-8 possible to detect human-scale processes such as urban growth. As well enables scientists to evaluate environmental change over time. Landsat 8 is key in the remote sensing classification process to improve our understanding of the Earth's land surfaces and the impact of humans on the environment, particularly in land surface change detection, i.e. urban areas. |
| NI (NASA ISRO) SAR (2020) | Using advanced radar imaging that will provide an unprecedented, detailed view of Earth, the NASA-ISRO Synthetic Aperture Radar, or NISAR, satellite is designed to observe and take measurements of some of the planet's most complex processes, including ecosystem disturbances, ice-sheet collapse, and natural hazards such as earthquakes, tsunamis, volcanoes and landslides. The two bands (L-band and S-band) both use Polarimetric Synthetic Aperture Radar. The L-band has a 24-centimeter wavelength and the S-band has a 12-centimeter wavelength. |
| NASA & USGS - Landsat-9 (2020) | Landsat 9, like Landsat 8, will have a higher imaging capacity than past Landsats, allowing more valuable data to be added to the Landsat's global land archive. Landsat 9 will carry two science instruments. Both instruments have sensors with moderate spatial resolution—15 m (49 ft.), 30 m (98 ft.), and 100 m (328 ft.) depending on spectral band—and the ability to detect the same range in intensity as Landsat 8, or better. Landsat 9 will be placed in an orbit that it is eight days out of phase with Landsat 8 to increase temporal coverage of observations. |

| Source/dataset | Description/Use |
|---|---|
| Oak Ridge National Laboratory (ORNL)<br>- LandScan | ORNL receives high-resolution images, census data, vector and raster layers from multiple sources (Digital Globe, GEOEYE, National Geospatial-Intelligence Agency (NGA[1]), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), US Geological Survey (USGS) and MacDonald Dettwiler and Associates (MDA) Federal) allowing ORNL to produce a comprehensive global population distribution database, LandScan. LandScan is produced using GeoCover land cover, MODIS 1Km, VMAP Level 1 and above, national and sub-national census data, Landsat Thematic Mapper data and panchromatic multispectral imagery (1-5m) (Bright et al., 2016). LandScan is a unique global population database given that LandScan measures the ambient population unlike the previously mentioned global population databases, which are based on static population. LandScan 2012 the latest version and is key in the remote sensing classification process. |
| Global Human Settlement (GHS) – Built-Up Grid | The GHS is developed and maintained by the Joint Research Centre, the European Commission's in house science service. The Global Human Settlement (GHS) framework produces global spatial information about the human presence on the planet over time. This in the form of built up maps, population density maps, and settlement maps. The newest 2015 GHS Built-Up grid is now available at approximately 38m, 250m, and 1km resolution. These data contain a multi-temporal information layer on built-up presence as derived from Landsat image collections (GLS1975, GLS1990, GLS2000, and ad-hoc Landsat 8 collection 2013/2014). The built-up surface area percentages indicates intensity of development. |
| Global Human Settlement (GHS) – Settlement Grid | The GHS is developed and maintained by the Joint Research Centre, the European Commission's in house science service. The Settlement Grid contains an assessment of the REGIO-OECD "degree of urbanization" model using the GHS Population grid cells for the four epochs (2015, 2000, 1990, and 1975). Each grid uses built-up areas detected through Landsat satellite imagery and integrated with the CIESIN GPW v4 population. The REGIO-OECD has 3 classes: high density clusters (HDC), low density cluster (LDC), and rural areas. |
| Global Human Settlement (GHS) – Population Grid | The GHS is developed and maintained by the Joint Research Centre (JRC), the European Commission's (EC) in-house science service. This dataset depicts the distribution and density of population as the number of people per cell.<br>The residential population estimates for 1975, 1990, 2000 and 2015 were provided by CIESIN GPWv4. These values were disaggregated from census or administrative units to grid cells, and combined with the Global Human Settlement Layer (GHSL) Built-Up grid to assist in the distribution and density of the global layer per corresponding epoch. |
| WorldPop | The WorldPop project was initiated in 2013 to unite the continent-focused AfriPop, AsiaPop and AmeriPop projects, with an aim of producing detailed and freely-available population distribution and composition maps for the whole of Central and South America, Africa and Asia. The population grids are available at a high resolution of 100m grids for estimated number of people per pixel for years 2010, 2015 and predicted 2020; these are available as adjusted grids to match the UN national totals and unadjusted. Population maps are updated to new versions when improved census or other input data become available. WorldPop is a great source for identifying rural populations that are typically undetected. |
| DLR- World Settlement Footprint (WSF; formally Global Urban Footprint (GUF) | The GUF project is the worldwide mapping of human settlements with a spatial resolution of 0.4 arcsec (~12 m). GUF is created from 180,000 TerraSAR-X and TanDEM-X scenes. The result is a three colours global map: black for "urban areas", white for "land surface" and grey for "water". This mapping style allows for the analysis of urban structures, settled areas, regional population distribution, and the arrangement of rural and urban areas. |

| Source/dataset | Description/Use |
|---|---|
| CIESIN<br><br>High-Resolution Settlement Layer 2015 | The High-Resolution Settlement Layer (HRSL) is a human population distribution estimate grid for 33 countries. The population estimates are based on census data and 0.5m high resolution DigitalGlobe satellite imagery. The population grids provide detailed delineation of settlements in urban and rural areas which are useful for research in disaster response or development of communication infrastructure. Settlement extent data were developed by the Connectivity Lab at Facebook using computer vision techniques to classify blocks of optical satellite data as either containing buildings or not. CIESIN used proportional allocation to distribute population data from subnational census data to the settlement extents. |
| CIESIN<br><br>Global Rural-Urban Mapping Project (GRUMP) –Population | CIESIN developed GRUMP global population distribution to provide a global population estimate that minimizes the impact of administrative boundaries. Prior to GRUMP, the Gridded Population of the World (GPW) was the first global population dataset. However, the first generation GPW relies primarily on administrative boundaries and census data resulting in an evenly distributed population grid within a given zone. GRUMP- Population Grid improved GPW results by incorporating additional datasets which yield higher population concentrations surrounding observable developed areas. The additional datasets includes census administrative units and CIESIN's Urban Extent Mask developed from NOAA's Night-time Light satellite images and the Digital Chart of the World's Populated Places (DCW) (CIESIN, Population Grid, 2004). Using an interpolation with urban areas as a weighting function populations were reallocated from rural areas to known urban areas. This data set is made available as a raster calculating population density per square kilometre and estimated population count by grid cell. |
| CIESIN<br><br>Global Rural-Urban Mapping Project- Human (GRUMP) -Settlement Points | GRUMP contains a geo-referenced framework of urban and rural areas by combining census data with satellite data. GRUMPv1 is comprised of three data products. First, GRUMPv1 provides a higher resolution gridded population data product at 30 arc-seconds, or ~1km at the equator, for 1990, 1995, and 2000. Second, GRUMPv1's urban extents data set delineates urban areas based on NOAA's night-time lights data set and buffered settlement centroids (where night lights are not sufficiently bright). Third, GRUMPv1 provides a point data set of all urban areas with populations of greater than 1,000 persons. This will be key in identifying rural populations. The Settlement Points v1.01 is now available with improved geospatial locations. |
| CIESIN<br><br>Global Rural-Urban Mapping Project (GRUMP) – Urban Extent | The Global Rural-Urban Mapping Project (GRUMP) Urban Extent layer is a moderate resolution binary (presence/absence) map delineates the global rural/urban extent. The Urban Extent layer utilizes NOAA's Night-time Light (NTL) 1994-1995 data to detect stable human settlement and ESRI's Digital Chart of the World's Populated Places (DCW) version 3 at the 1:1,000,000 scale for initial settlement points (CIESIN, Urban Extent, 2004). For areas of inadequate to low electrical power sources the urban extents were extrapolated using a population-area ratio; Tactical Pilotage Charts were incorporated for urban delineation in Africa and Latin America for such reasons. The Urban Extent grid remains as v1 however, the Urban Extent Polygon v1.01 reflects the changes made to the updated settlement points. |
| CIESIN<br><br>Gridded Population of the World (GPW) | The Gridded Population of the World (GPW) data set models the distribution of human population (counts and densities) on a continuous global raster surface. GPW provides a spatially disaggregated population layer that is compatible with data sets from social, economic, and Earth science disciplines, and remote sensing. It provides globally consistent and spatially explicit data for use in research, policy-making, and communications. The newest GPW, version 4.10, uses the most detailed spatial resolution available with an output resolution of 30 arc-seconds (approximately 1 km at the equator). Significant improvement from the previous versions 2.5 arc-minute resolution. The input data are extrapolated to produce population density and count estimates for the years 2000, 2005, 2010, 2015, and 2020. United Nations World Population Prospects (UNWPP) adjusted products are also available for those years to match the 2015 UNWPP revision. |

| Source/dataset | Description/Use |
|---|---|
| Impervious Surface (IMPSA) | The impervious surface layer was developed using Night-time Lights Time Series 2000-2001 and LandScan 2004 population count, at a coarse resolution, along with USGS 30m resolution impervious surface area (ISA) estimates of the for calibration in the US. During ISA estimate modelling, USGS 30m Landsat was used to derive ISA as reference data then aggregated to 1km equal area grid in an Albers projection; the night-time light series (2000-2001) and LandScan (2004) were re-projected to 1km. Impervious surfaces of the world is measured in percentage derived using the below equation (Elvidge et al., 2007). |
| Copernicus/EEA (European Environment Agency)- CORINE | The Coordination of information on the environment (CORINE) Land Cover (CLC) inventory was initiated in 1985 (reference year 1990). The 4th CLC inventory for the reference year of 2012 was produced under the Copernicus GMES Initial Operations (GIO). It has the shortest production time in history of CLC. Two high-resolution satellite image coverages (IRS Resourcesat-1/2, SPOT-4/5, RapidEye constellation) taken in 2011-2012 provided multi-temporal information to support the update. CORINE consists of 44 land cover classes spanning over 39 European countries with a 100-meter pixel resolution. |
| ESA CCI Land Cover 2010 | The Climate Change Initiative – Land Cover (CCI-LC) team has produced 3-epoch series of global land cover maps at 300m spatial resolution, where each epoch covers a 5-year period (2008-2012, 2003-2007, and 1998-2002). These maps were produced using multiple sensors and various time frame in order to maximize the product consistency. The 2003-2012 MERIS data serves as a baseline to derive the 2010, 2005 and 2000 maps. |
| HydroSHED 2000 | Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS) is an elevation data derived from Shuttle Radar Topography Mission (SRTM) at 3-arcsecond resolution and a variety of auxiliary datasets used for reference and quality control. HydroSHED provides regional and global-scale hydrologic information such as river networks, watershed boundaries, drainage direction and flow. New algorithms were applied for data improvement (void filling, sink filling, filtering, etc.), iterative manual corrections were applied, and finished with to a variety of coarser resolutions. This data set is useful for evaluating false positive of high population settlement due to high infrared readings. |
| ESRI – World Population Estimates | ESRI has developed a World Population Estimate and Density grid at 150m resolution, Settlement Score at 75m resolution, and a Confidence raster layer at 450m resolution. The settlement likelihood and population values were created by using the most recent census data available. |

## 3.3 HIGH AND VERY HIGH RESOLUTION

High resolution and very high resolution imagery typically ranges from a few meters to as fine as a foot (30cm). This include very high satellite imagery and aerial photography. Often high resolution imagery is available for a given area from a variety of sources, including the local government, the federal government, commercial satellite data providers, and mapping platforms like Google Maps or Bing Map. These data sources are typically too high resolution for the types of textural analysis useful for identifying development patterns. They can be used to extract building specific footprints which can feed into the building count and building area estimates, if the study region is small. These methods can work very well in environments with uniform building patterns such as suburbs, but do not work consistently across roof types and will not penetrate tree cover. In dense urban environments, textural analysis would likely not be able to distinguish between individual buildings but total building area estimates calculated do fairly match the known building areas value for a given administrative area. Although high resolution remote sensing data is frequently affirmed as a source for estimating building counts, we recommend the use of high resolution images for a variety of other purposes which yield much more dependable results, such as: 1) establishing what development pattern categories to use in the study area, 2) creating mapping schemes, 3) manually digitizing development patterns in key areas, 4) establishing key parameters, such as average number of buildings per grid or average building size by development pattern category, and 5) validating a given exposure database.

High resolution data, as seen through GoogleEarth, is useful for delineating development patterns and classification training areas. Users appreciate that optical data is easy to understand, as it is essentially a photograph from space. Obtaining and processing optical satellite data for a given study area can be a significant undertaking. This is due to the fact that pixel resolution of optical imagery typically is higher than an interpreted data products; i.e. LandSat-8 panchromatic band has a pixel resolution of 15-meters while the Suomi NPP VIIRS data product is approximately at ~500-meters. Users often make the mistake of assuming that higher resolution data is always better, but in practice this is exactly the opposite for many exposure development activities. Higher resolution imagery captures a fuller range of phenomena that can be viewed from the space including trees, shadows, and so forth. For exposure development, the analysis requires an extensive site specific analysis to process the high resolution data in a manner that takes into account the diversity of all the observed development patterns. The analyst must have a strong understanding in order to decide which of the imagery, products, or individual spectral bands are most useful. As well as what combination of textural analysis and patterns signatures are suitable to “bridge” across the “noise” inherent in the imagery to capture the desired phenomenon. Typically an estimation of the density of buildings is desired in which case an interpreted data product may be best data source to use. However, in many cases upon review the interpreted data product may be inadequate for a given area which may justify extensive use of optical resolution remote sensing imagery. For example, in many African countries there is inadequate amount of identified rural communities in interpreted products, presumably because the methods used depend upon night-time light data. Simple edge detection may identify these settlements but there needs to be a significant effort to identify and remove false positives. In addition, given the volume of data, custom programming scripts need to be written for running the process.

## 3.4 RESEARCHING DEVELOPMENT PATTERNS

Development patterns are regions of similar construction types. These development patterns establish the categories used for sampling structural attributes. Identifying the type of development patterns categories allow researchers to harness EO data to rapidly classify vast amount of areas, nationally or globally, at a fraction of the time, and expense, which would have been spent if the surveys were conducted in the field for the entire cover area. Development patterns can be established on a very large scale using global datasets, for example by using population density of VIIRS night light data as a proxy for building height and establishing regions of development patterns, or they can be established on a regional scale by running image processing segmentation (classification) algorithms and linking the results of the classification to development patterns. Digitizing development pattern increases the accuracy of segmentation. These samples serve as training for supervised classification algorithms that evaluate the spectral signature of the sample development pattern areas with the EO data sets. Either way, whether the focus is global, national, or local, the first step is to establish the development patterns that are likely to be relevant for the area of interest. This is an iterative process that involves coordination between structural engineers and remote sensing experts.

Internet reconnaissance is a very useful process that aids in the correct delineation of development patterns. Internet searches of the study region can greatly help researchers and analysts by providing specific information regarding the area that is not available by merely looking at the satellite image. This information can take any number of forms, from general information about a population or the region's architectural style to very specific georeferenced clues. As well, the type of information a researcher seeks will also depend greatly on the quality of the available satellite imagery for the region.

**General:** General information is any information that gives an understanding of the study area. This might include governmental or international organisation reports on demographics, census recordings, population statistics, descriptions of geographical features or a region's layout, and economic infrastructure. The most helpful general information that can be found is information that can be or is georeferenced. This information can therefore be directly used to create a boundary or delineate and classify a development pattern and is thus actionable.

For example, many travel websites or blogs include descriptions of the specific region. A traveller that is writing about their experiences will often mention where hotels or restaurants are located. An example might be "Our hotel was next to the temple on Main Street" or "We avoided the commercial centre that runs along the north side of the airport." These phrases can be directly actionable. In the delineation process, the analyst now knows that the building next to the temple on Main Street is likely a hotel and the buildings running along the north of the airport are probably commercial. Much of this general information is not specifically actionable, but can be very helpful in the delineation process. A travel website might say, "Villagers live in grass huts," this prompts the analyst to associate grass roof structures as an indicator of residential development for the study region.

**Geographic descriptions:** Many websites will offer a written description of geographical features of an area. These are especially helpful because they are often georeferenced. An example of this would be a description of a river being a city boundary.

**Economic information:** An understanding of the region's economy can help the delineator understand what they are seeing in the satellite image; a city with a large industrial base will be expected to have many large industrial buildings. Similarly, a region with an agricultural base will tend to have residential structures in a more rural setting, and industrial zones in a denser urban setting.

**Population Information:** A search will often bring up population statistics. Population information can be very helpful when satellite imagery is unclear. A large population in a moderate size administrative unit means more densely populated cities or villages and a smaller population over a larger administrative region will usually be either one small compact village or a thinly spread agricultural/residential style of living. Knowing this can help clarify what the delineator is seeing in the image. In unclear imagery, a village can appear as just a slightly different texture than the environment around it. The delineator can feel more confident in declaring this textural difference as a village if they understand what they are looking for.

**Images:** As mentioned above, the use of imagery of the study region is very beneficial. There are two manner as to how this can be accomplished. First, in a very general sense, the analyst can use the imagery as a guide to better understand what the study looks like at bird's eye view or nadir. Photographs of buildings inform the delineator what the architecture of a region looks like. Images taken from airplanes provide another view by which to understand the study area and how it is laid out, again, informing the viewer how the land and structures on it look and are being used. The second and most valuable way in which the images aid delineation is by providing georeferenced information. Such clues can be street signs, signs on buildings, or even text that belongs with a photo that tells specifically where the photo is taken.

## 3.5 ILLUSTRATIVE EXAMPLES FOR TANZANIA

After the initial web reconnaissance of visual, statistical, or demographic information search a set of primary development patterns can be created. Using the information available the engineer creates development pattern categories that the analyst will use when delineating sample areas. The analyst gives a set of key descriptors or metrics that define each development pattern type. Below are examples of the different types of development patterns observed in Tanzania. The engineer used GoogleEarth and available street level images to add context along with a description.

Each development pattern category should be defined in the least ambiguous manner. Thus, the researcher should review the study areas before developing and establishing a development pattern classification system. Observing the study area before delineating sample polygons will help create a classification system that reflect the area and will smoothly apply to a mapping scheme. This section discusses source data for establishing development patterns and provides examples in Tanzania where available.

### 3.5.1 Development Pattern 1

This type of rural development can be found outside of city boundaries and is typically associated with agricultural development. The regions typically consist of small, remote villages with single roads in and out. Buildings are typically spaced far apart and are almost exclusively 1 to 2 stories. Local materials and construction practices are generally used and performed in these areas.

*Figure 3.2: Development Pattern 1 in Tanzania. Left: GoogleEarth[1], accessed in 2018. Right: available street level images.*

### 3.5.2 Development Pattern 2

This development pattern reflects areas typically dominated by single family residential structures (SFR). Commercial properties, such as local markets, are present, however residential structures are the primary occupancy. The built-up area is dense, however open land (yards, vacant lots, etc.) are present and can be observed via satellite imagery. All structures are low-rise, with most in the 1 to 2 story range.

*Figure 3.3: Development Pattern 2 in Tanzania. Left: GoogleEarth[1], accessed in 2018. Right: available street level images.*

### 3.5.3 Development Pattern 3

This development pattern can be characterised by structures where the majority of the population lives in multi-family residential housing. The built-up area is typically comprised of long narrow apartment blocks in the 3 to 5 story range. Large open spaces (courtyards or parking lots) are present between buildings, therefore building density is typically not high. Smaller (<300 sq. m.) 1 to 2 story buildings can be observed, but they are typically limited to small offices or commercial structures within the complexes.

The images below are aerial view of development pattern 3. Roofs are constructed of the same material, and therefore should be the same colour throughout. They are easy to identify using satellite imagery, as they are uniform throughout the development; buildings are spaced out (large open spaces) and are much larger than SFR. Heights range from five-six stories.

---

[1] Map data: Google, Maxar Technologies

*Figure 3.4: Development Pattern 3 in Tanzania. Left: GoogleEarth[2], accessed in 2018. Right: available street level images.*

*Figure 3.5: 3-story RC frame homes, from available street level images*

*Figure 3.6: Typical buildings in high density residential areas, from available street level images*

[2] Map data: Google, Maxar Technologies


*Figure 3.7: Aerial view of high-density residential development, from GoogleEarth[3], accessed in 2018.*

### 3.5.4 Development Pattern 4

This development pattern is typically associated with extremely dense, informal settlements. They are usually found within boundaries of large cities, and are typically comprised of very small (<100 sq. m.) standalone structures with little to no space between adjacent buildings. The settlement is unplanned, therefore there is no organisation to the configuration of building layouts. Almost all structures are single-story and are typically erected using cheap and accessible local materials. Roof colours will vary from building to building, and even on the same structure. There is very little (to no) space between neighbouring buildings. Although most building footprints appear fairly regular (square, rectangular), closer inspection shows that corners are typically not 90 degrees, suggesting buildings were constructed without formal planning. Building footprints are small (<500 sq. feet) and most are one-story. This development pattern is easy to detect in a satellite image and will be identified manually in this study. Floor area is anticipated to be almost 100% of the built area.


*Figure 3.8: Development Pattern 4 in Tanzania. Left: GoogleEarth[3], accessed in 2018. Right: available street level images.*

### 3.5.5 Development Pattern 5

This development pattern is characterised by urban areas predominately occupied by low to mid-rise residential and commercial structures. An occasional high-rise apartment or office building may be present. These developments are typically found near or around major city centres. Buildings are tightly spaced and are fairly regular in shape.

[3] Map data: Google, Maxar Technologies


*Figure 3.9: Development Pattern 5 in Tanzania. Left: GoogleEarth[4], accessed in 2018. Right: available street level images.*

### 3.5.6 Development Pattern 6

This development pattern is similar to the central business district of any major African city. Mid- to high-rise apartments and commercial offices occupy most of the area, but low-rise commercial and residential structures can be situated in between. Typical of an urban area, buildings are spaced relatively close and building layouts of both building and city blocks are structured.


*Figure 3.10: Development Pattern 6 in Tanzania. Left: GoogleEarth[4], accessed in 2018. Right: available street level images.*

### 3.5.7 Development Pattern 7

This development pattern is characterised by areas dominated by ports, mining or industrial activities. Structures are typically closely spaced and regular in shape. A majority of buildings within these regions are warehouses, rectangular shape, and single story. Smaller low-rise, office, and commercial structures can also be found on site.

[4] Map data: Google, Maxar Technologies


*Figure 3.11: Development Pattern 7 in Tanzania, from available street level images and GoogleEarth[5], accessed in 2018 (lower left).*

### 3.5.8 Development Pattern 8

This development pattern is typically associated with dense, low–rise, commercial and residential structures. The pattern may appear similar to the informal developments of classification 4, but multi-storied buildings are present, building footprints are typically larger and informal construction techniques are not as prevalent. It is not found in Tanzania.

*Figure 3.12: Development Pattern 8. Left: GoogleEarth[5], accessed in 2018. Right: available street level images.*

## 3.6 DEVELOPMENT PATTERN CLASSIFICATION USING EO

Classification, or segmentation of EO data is done in an image processing environment. Segmentation can be a supervised or unsupervised process; the best methods and data to use will depend on the specifics of the environment. There are out of the box classification

[5] Map data: Google, Maxar Technologies

algorithm as well as more advanced machine learn and decision tree options. Deciding which algorithm to use will vary depending on the availability of funding, time, and computing space. For example, a certain algorithm can use more rigorous statistics, iteration cycles, and pixel constraints which results in a longer processing time and less inclusive pixel determinations, possibly more accurate. However, not all project or task in the overall process require such rigor thus an algorithm with lower constraints would result in a more inclusive pixel determination, meaning more pixels are defined but maybe less accurately labelled, and shorter run-time.

## 3.7 MAPPING STRUCTURAL TYPES TO DEVELOPMENT PATTERNS

To create an exposure database from the six identified development patterns, “mapping schemes,” or building height and structure type distributions are linked to each zone type on an individual country basis. Mapping schemes must be established with structural engineers either specializing in, or guided by those familiar with CAT modelling. Engineers review a variety of sources to carefully characterise each development pattern.  This may include any available field surveys, a literature review of typical building practices, including a review of the Earthquake Engineering Research Institute (EERI), World Housing Encyclopaedia (WHE) and historic reports. A full review of geotagged photographs and other online sources, frequently enlists sampling through crowdsourced technologies, additional review of aerial photos, satellite imagery or street-embedded video available through Google Earth, Microsoft Bing, or similar tools operating on a regional basis.

### 3.7.1 Mapping Scheme data sources

For the Tanzania exposure database, online sources, scholarly journals, and imagery (ground/satellite/aerial) were originally sourced to identify typical construction materials and lateral force resisting systems (LFRS) within the region. Sources such as the WHE and Prompt Assessment of Global Earthquakes for Response (PAGER) offer both site specific construction and LRFS distributions across the region. Ground photos were found either online, through surveys or geotagged on host sites such as Google Earth. An engineer used these to validate any structural type assumptions found through the research, and to identify other structural types overlooked. These photos, alongside high resolution satellite/aerial imagery found in Google Earth, were used to identify common traits, such as building size, roof colour, roof slope, etc. available only via birds-eye view. This is important as it allows the analyst to develop a relationship between structural types (identified via ground survey/photo) and traits visible via satellite/aerial imagery, since ground imagery is not available for all buildings. Online educational and vacation videos were found on video hosting sites such as YouTube and Vimeo. Not only do these provide a wide coverage of the area, they also provide an interior view of structures, revealing clues about the LFRS typically hidden by the façade. Ground photos/videos however will not be available for every structure, therefore it is important to identify a relationship between structural types (identifying via ground survey/photos) and traits visible via satellite/aerial imagery.

The following example, is a sample of data and resources used for the Tanzania exposure database.

World Housing Encyclopaedia – GEM (Global Earthquake Model) taxonomy reports are available for specific sites within the country. Included for Tanzania, is a rural home which identifies the structural type, plan shape and number of stories. This specific structure is a rammed earth house with either a grass thatch or corrugated iron sheet on timber poles.

#### 3.7.1.1 PAGER

This database identifies a distribution of structural types throughout the country of Tanzania using UN-Habitat survey data (UN-Habitat, 2007; Jaiswal and Wald, 2008). Breakdowns include 66% unreinforced masonry (rubble stone and brick), 15% mud walls, 0.5% wood framed, 19% adobe and 0.5% informal. The inclusion of known structural types is beneficial to the project team, however the distribution is reflective of the building stock for the entire country and not the land use classifications mentioned previously. Additionally, one should note the exclusion of structural types commonly found in urban regions, particularly those structures identified as mid to high rise (4+ stories) construction. These reinforced concrete frames and reinforced concrete frames with masonry infill are not included in the PAGER database, however, were observed in ground and aerial imagery. For these urban (financial) development patterns, adobe and mud wall structures will not be a majority, as RC frames and URM buildings populate most of the area. The mapping schemes developed in this project accurately identify the correct building distribution for each land use classification.

| Source/dataset | Description/Use |
|---|---|
| World Housing Encyclopaedia<br>Earthquake Engineering Research Institute (EERI) | The World Housing Encyclopaedia is developed by a consortium of EERI members throughout the world. Structural characteristics are characterized on a nationwide basis specifically with earthquake vulnerability in mind. |
| Prompt Assessment of Global Earthquakes for Response (PAGER)<br>USGS | Building on data from the previous two entries, USGS has expanded the general framework of HAZUS to the entire world. Occupancies are collapsed to simply urban and rural. Given that the objective of PAGER is casualty estimation, the mapping schemes within PAGER represent an estimation of the percentage of population residing in any given structural type, not the percentage of buildings or square footage associated with a given structure type. The database draws on and harmonizes numerous sources: (1) UN statistics, (2) UN Habitat's Demographic and Health Survey (DHS) database, (3) national housing censuses, (4) the World Housing Encyclopaedia and (5) other literature (Jaiswal et al., 2008). The mapping schemes from PAGER form a powerful basis for inferring structural types globally. The database is freely available for public use, subject to peer review, scrutiny, and open enhancement. For more details see both: *Creating a Global building Inventory for Earthquake Loss Assessment and Risk Management*, and Jaiswal and Wald (2008). |
| UN Data<br>UN Habitat- Global Urban Observatory | Raw survey data that is collected for the national housing census, such as those currently being assessed by UN-Habitat provides detailed building counts and square footage data by both occupancy and structure type. Estimating the number of building units given population requires careful assessment of this data, as well as remote sensing assessment. Estimating the number of building units given population requires careful assessment of the raw survey data. If a user is developing a new study region with the Inventory Data Capture Toolkit (IDCT) toolkit, they will be able to update these assumptions for their specific dataset. |
| Valuation data | Valuation data typically provide an estimate of cost per square foot by building type. Commercial companies such as MSB, Bluebook, RS Means, and Spons provide services to estimate the value of facilities in a database based on square footage (primarily US-based residential and commercial). This is key for going from estimates of damage to estimates of loss. |

*Table 4: Sources Facilitating the Development of Mapping Schemes*

3.7.1.2 ONLINE SOURCES

Online technical, scholarly, and humanitarian reports are a great source for understanding the local construction practices. These reports contain detailed descriptions of existing construction methodologies and materials used, as well as give insight into new project developments and advanced construction methods. In the Tanzania pilot, engineers were able to obtain a technical report from the Department of Structural and Construction Engineering University of Dar es Salaam. This Tanzanian report identifies the application of local construction technologies, common materials used, current construction practices, and new projects are detailed. Figure 3.13 is an example of the types of information the reports provide. By adding images with the text engineers or analyst get a clear understanding of how and of what material residential homes are made from in Tanzania.

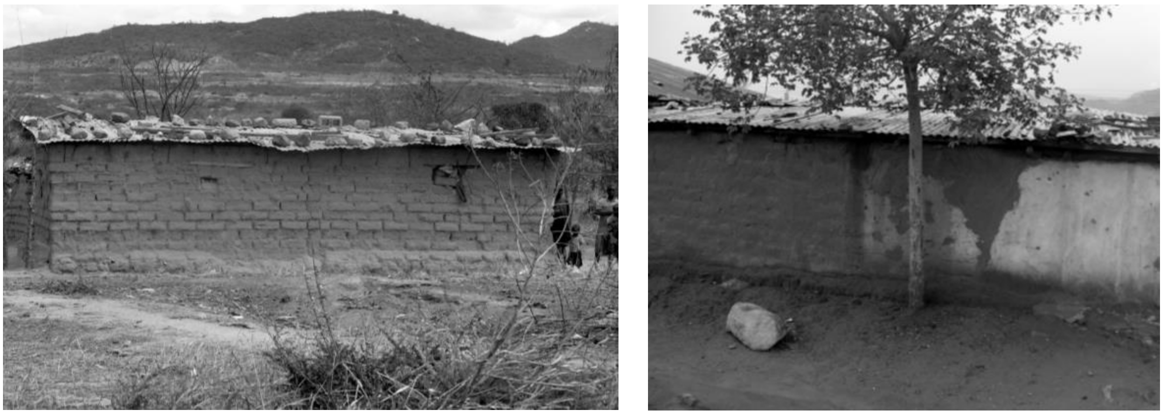

*Figure 3.13: Example if common construction practices in Tanzania (technical report from the Department of Structural and Construction Engineering University of Dar es Salaam)*

3.7.1.3 GROUND PHOTOS/VIDEOS/SATELLITE IMAGERY

A combination of ground images, aerial photography, and videos can be used for developing a mapping scheme. In Tanzania, these sources were used to validate structural types and allocate the appropriate distribution to each land use classification. Geotagged photos are particularly important as they not only allow the engineer to identify the particular structural type but also assess the surrounding areas. Figure 3.14 below is a geotagged photo in downtown Dar es Salaam. From the photograph, ImageCat engineers were able to infer the expected structural type, and simultaneously observe the surrounding structures to conclude that those structures are all of similar height and material; this based solely on the satellite image.

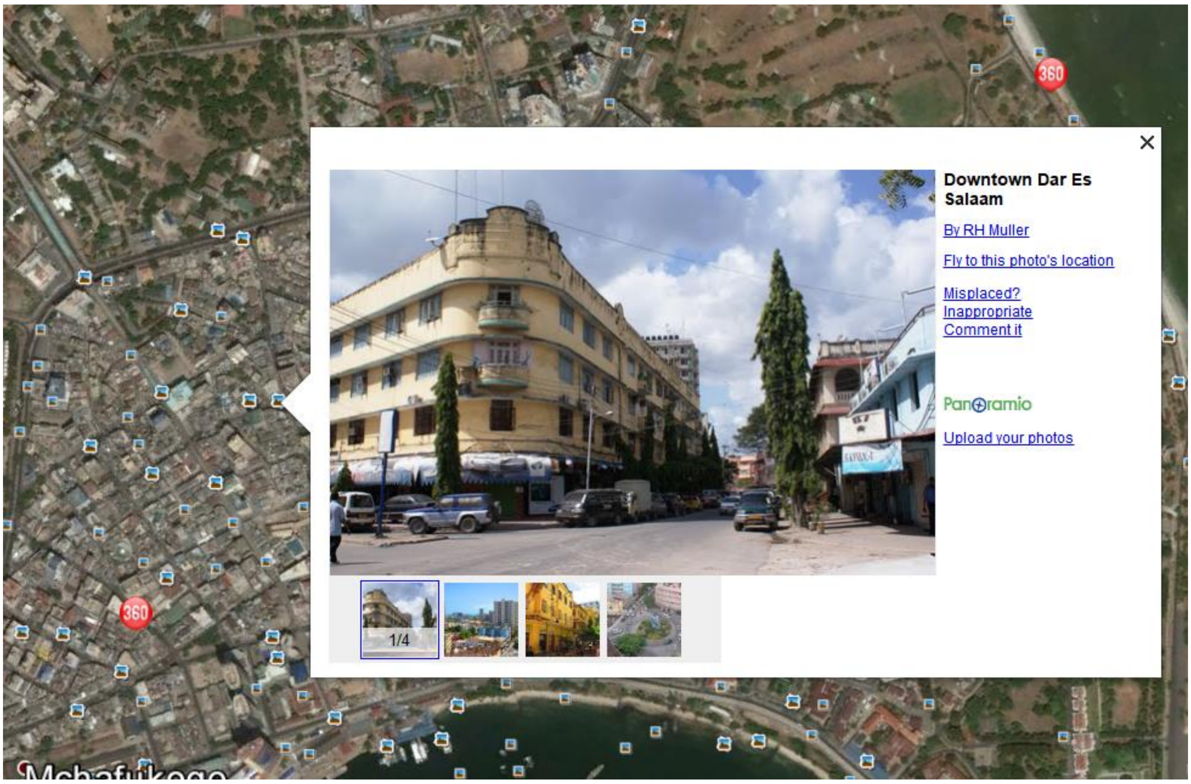


*Figure 3.14: Example of Photos available from Panoramio Online[6].*

[6] Panoramio Online has since been discontinued with many of the photos incorporated into Google Earth and Google Maps (see https://www.panoramio.com/)

Once these common construction practices are verified, the engineer establishes mapping schemes for the various development patterns (rural, residential, urban, etc.). Most of these patterns are fairly homogenous (e.g. dense residential), so construction types, number of stories and building footprints will remain relatively the same. Engineers identify the anomalies within the land use class, and assess these structures and frequency of reoccurrence. A distribution of building properties is created and checked with similar land use grids to verify the consistency. In city centres and urban areas, such as those grids classified as urban or high urban, both aerial and ground photos are necessary to establish both height and structural type distributions. Online sources such as Emporis and Skyscraper page identify well-known high-rise structures, and are used to establish a conservative ceiling for building heights. With this data, along with knowledge of local construction practices, the engineer establishes height and structural distribution of the specified urban land use classification.

### 3.7.2 Sampling Buildings for the Development of Mapping Schemes

In order to get more detail on construction patterns for mapping schemes, it is often necessary to collect data from the field. The method of deciding which buildings to survey should be based on random sampling, and be robust from a statistical perspective. The method recommended by the project team is summarised in the GEM document, "User guide: Field Sampling Strategies for Estimating Building Inventories" (Porter et al., 2014). It is highly recommended that those interested in developing exposure databases refer to this document directly. An overview is provided from the abstract:

> *"Four sampling strategies are offered for sampling a zone depending on whether (a) the zone has homogeneous building heights, (b) all important building features are visible, (c) expert advice is available on recognizing hidden features, and (d) a prior estimate from experts on building-type distribution is available. The strategies are:*
>
> *1. Simple sampling without prior expert judgment. This procedure applies where the survey team can identify building type from visible features. It is useful for zones with homogeneous (fairly uniform) heights. That is, one cannot easily pick out clusters of buildings that are on average shorter, typical, or taller in terms of number of stories. It uses simple weighted averages to extrapolate field observations to the entire zone.*
>
> *2. Stratified sampling without prior expert judgment. Like 1, except that the zone has a heterogeneous (not so uniform) mix of heights, e.g., a central business district with some generally low-, mid- and high-rise blocks. It uses a procedure called moment matching to select the sample and extrapolate to the zone.*
>
> *3. Use local expertise to infer types from visible features. Like 1 or 2, but a least one attribute used to define building type is not visible. Prior expert advice is needed to infer building type from visible features. Field data collectors then employ either simple or stratified sampling depending on height homogeneity. If the expert is willing to estimate the distribution of building types, this procedure shows how to do that.*
>
> *4. Field survey to enhance prior expert opinion of type distribution. Like 3, except that it requires the expert to provide a prior distribution of building type by area, which is then combined with data from the field observations using Bayes' theorem. It is estimated that one team of two people can survey enough buildings to represent the distribution of one zone in approximately one day.*
>
> *A separate procedure is offered for special buildings, such as buildings that serve some important function or are known to be particularly seismically vulnerable or particularly seismically resilient. Examples include hospitals, tall masonry towers, or base-isolated buildings. An appendix provides the mathematical basis of these procedures, including stratified sampling, Bayesian updating, and a discussion of the trade-off between sample size and sensitivity to rare building types."*

### 3.7.3 Field Data Collection with IDCT

The Inventory Data Capture Toolkit (IDCT) from Global Earthquake Model (GEM) group is an open source in-situ data collection system. The application facilitates the collection of building information that is key is catastrophic loss modelling. The tool allows users to take infield photographs, GPS location, and make key observations in accordance to the GEM taxonomy. This direct data collection method eliminates the need of paper form building evaluations. Reducing the time spent to create a digital exposure database, ensures uniform mapping schemes, and reduces human error during copying data into a digital tables. In addition to optimizing direct in-field observations the IDCT can be pre-loaded with remote sensing footprints, offline base maps tiles, and vector sampling zones. The offline base map tiles are particularly useful in post-catastrophic events and enables users to input high resolution imagery to better pin the building location. The data collected by engineers before a natural disaster can be used to create a building exposure model. As well, post-disaster data takes stock of building conditions and assists in building damage mitigation. Since the IDCT is open source the code can be adapted to meet the needs of the user if the GEM taxonomy is not sufficient. The IDCT works on Android or Microsoft operating systems and is best for tablet usage. The sampling data collect can be combines with the Spatial Inventory and Damage Data (SIDD) tool to develop a statistical approach to creating a mapping schemes and exposure data. The Android version is widely preferred for the ease of use in the field and capability for use on Android smartphones (Foulser-Piggott et al., 2014).

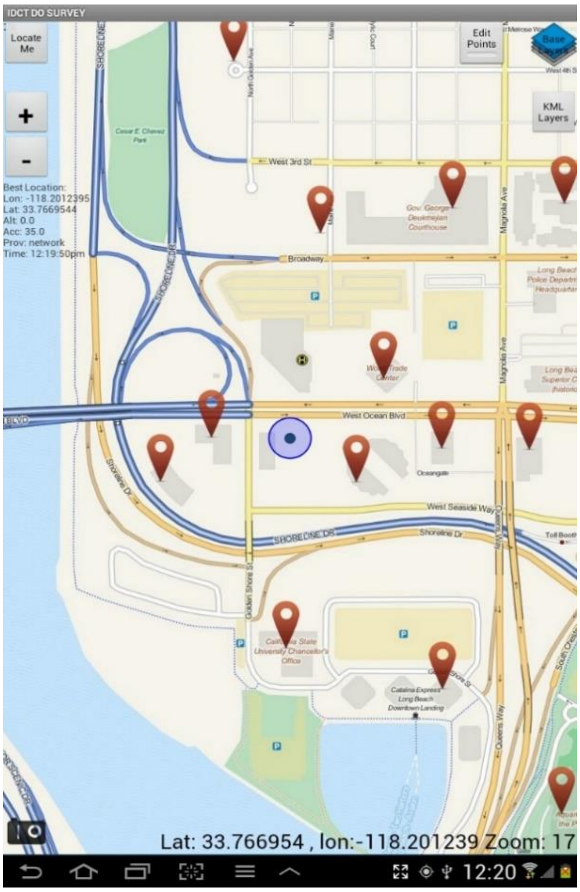


*Figure 3.15: IDCT User Interface zoomed to user location using the "Locate Me" icon. Inventory Data Capture Toolkit (IDCT) from Global Earthquake Model (GEM)[7].*

[7] https://storage.globalquakemodel.org/what/physical-integrated-risk/inventory-capture-tools/

### 3.7.4 Extrapolate Exposure

With mapping schemes, development patterns, population data or building counts, and cost per square foot, users can use GIS processing and programming to distribute the assets. This process is simplified by SIDD (spatial inventory data development) (Hu et al., 2013), another product distributed by GEM (Huyck et al., 2014). SIDD is part of the GEM IDCT (Inventory Data Capture Tools) software for developing building exposure data in a GEF format (GEM Exposure File), suitable for analysis with the OpenQuake engine. SIDD serves as a critical intermediary between raw sample data collected in the field, building footprint data extracted through remote sensing, and the final estimate of regional exposure contained within a GIS data set that is loaded into OpenQuake. Exposure data can be created from a very comprehensive study or cursory adjustments of mapping schemes for the purposes of sensitivity analysis. This flexibility was incorporated as a design requirement to accommodate as many users as possible. As with the sampling protocol, the reader is directed to the GEM website for further details.

## 3.8 VALIDATION

Development of exposure data with EO saves time, but it is easy to make mistakes. Common sources include misinterpreted units, inclusion of erroneous data, geocoding or other data resolution problems, and so on. This section presents some general guidelines to identify obvious sources of error.

Mapping essential attributes can help identify anomalies. These includes the number of sites distributed at the admin II and III level, the number of point facilities at the admin II and III level, the population at the admin II and III level, the density of all facilities per person at the admin II and III level, and scatterplots of number of facilities and population at the national scale. Outliers can help identify errors in base tables or GIS layers, or human errors in processing. It is particularly important to review the distribution of building counts. Once buildings are distributed, the number of buildings can be aggregated to the Admin 3 level and compared with the population. Data. Where these numbers are not consistent with the expected rates of people per household and households per building for that administrative area, key parameters of the analysis should adjusted and the process re-run. In addition to the ratio of people per building, the ratio of buildings in settlements to rural areas is a key parameter. Unfortunately, there are usually no “settlement boundaries” that can be used to develop hard ratios to compare with published statistics. However, because buildings are distributed first to settlements, and then distributed to rural areas, there are clear visual cues when the distribution between urban and rural areas is inadequate. If too many buildings are assigned to urban areas, there is a void in rural areas. Checking this against EO data visually can help confirm whether these areas are populated. Typically, this would require the adjustment of residential building density for a specific region.

In addition, it is a good idea to compare the final distribution of buildings against the population or urban areas interpreted EO layers outlined in Table 3. Although these datasets post population or urban/area density and do not represent the same output, a visual comparison can reveal potential issues with building stock distribution. A user should consider choosing several locations to reflect a variety of physical environments and settlement conditions, then compare the results qualitatively against the interpreted EO data on higher resolution imagery. Bing and Google Earth work very well for these assessments. In many cases, the building stock data will match the high-resolution imagery better than the interpreted EO data sets. In other cases, adjustments may need to be made in certain parameters.

### 3.8.1 Working with local community members

With regard to the overall exposure development process, when developing a Level 2 or higher exposure dataset, it is essential to work with local community members and/or government agencies where possible. Engineers from a given area will have intrinsic knowledge of their buildings, construction, and governments in a way that cannot be observed with EO data. For example, local engineers might be able to provide context as to how a region developed over the last several decades- including migration from rural to urban areas, development of multi-family housing, and the construction of high-rise districts. They may be able to further provide information as to how these patterns coincide with the observed development patterns, and provide data that can't be seen from space- such as the type masonry, level of reinforcement, adherence to code, and so on. It is important to recognise, however, that the combined judgment of only a small number of experts will likely vary significantly. That is, results will show sensitivity to the assumptions of one or two experts.  A target number of 5 or more experts per country can mediate outlying opinions.

It is often difficult to distil such conversations with local engineers into the process of exposure management. However, with the stratified sampling technique, particularly as made available through the SIDD toolset, it is fairly straightforward to encapsulate these observations in development-pattern specific mapping schemes. Observations need to be broken down into estimated percentages in order to be applied quantitatively. This can take a little additional research, and is bolstered by local reports and data where available. Ultimately, involving the community in the development of exposure data helps to bolster acceptance of the final product, as well as establishes a dialog that is essential to communicating the limitations of the data and clarifying appropriate use. In addition, it is also preferable to have some of the spatial data development or validation activities other than replacement costs and building-type distributions carried out by local specialists. Local experts have the benefit of being able to source and deliver official data from the public sector in a way that international teams are not able to easily achieve.  Partnering with local experts also has the benefit of sharing of international knowledge on exposure development practices to help build capacity.

## 3.9 LIMITATIONS

An important part of developing exposure data with EO data sources is to be transparent about the process and the anticipated impact on results. Developing full metadata as described in this report and highlighting assumptions and limitations will help guard against inappropriate use of the data, particularly outside of the charge of risk assessment. In the end, most of the methods to develop exposure data with EO data are various methods of spreading an estimate of the number of buildings derived from population statistics throughout an enumeration district depending on a host of assumptions, including persons per household, households per building type, construction pattern distributions for various development patterns, the homogeneity of various construction patterns, building densities at the grid level as derived through image processing algorithm, and typical replacement costs. Although process may have been carefully tailored through expert opinion, sampling, research, and manual review, the exposure data for Levels 1-3 are not a count or census, and usually the data for Level 4 and 5 are not either. At the cell or small city level, they may not be representative. While mapping schemes and replacement costs may be representative of typical building infrastructure and replacement costs for the entire country, regional variations in costs and building distributions (due to cost of materials and labour) will vary.

# 4 Levels of Data Collection and the Impact on Loss Estimates

Understanding the impact that the building exposure has on loss estimates is a substantial challenge for several reasons. The exposure data is arguably not a source of modelling uncertainty- given that in most cases the exposure data is an “input” to the model that may be done by a separate team. The error is often considered “user error” for not allocated proper resources to collect data.  This even though key factors in exposure data collection such as valuation and shifts in exchange rates can easily exceed the uncertainty due to hazard and vulnerability combined. Another factor complicating the quantification of uncertainty due to building exposure is that a generalised exposure database used for government or NGO purposes is frequently run on a deterministic basis. The error, in this case, is highly dependent on what event is chosen. A probabilistic approach provides a much richer look at the impacts but is often not feasible for government applications given the level of effort and processing requirements.  Finally, an adequate examination of the exposure should consider exposure databases that are largely compiled independently and at different scales to consider epistemic uncertainty. It might be quite simple for a given project team to adjust certain assumptions while collecting exposure data and then conduct a sensitivity analysis as to the impact of these decisions. This is a good start but limits the assessment to the imagination and capabilities of a given team. It also, in most cases, would involve adjusting assumptions with the same processing steps- which can substantially contribute to error. Following a robust simulation approach (Taylor, 2015; Lee et al., 2018; Lee, Y. 2018) the quantification of a given source of uncertainty should include multiple versions of component tested, whether vulnerability functions, GMPEs, flood models, or exposure databases.

## 4.1 LOS ANGLES CASE STUDY

To this end, the project team set out to develop multiple exposure databases for a single location, ideally largely compiled by different teams or at least different methods and run the results through a probabilistic analysis. Los Angeles was selected for a variety of reasons. The project team has access to SeismiCat, a full probabilistic model for this region (ImageCat, 2019[b]). There have been several studies in Los Angeles County California, US that have contributed to the development of various open source exposure databases over the last several decades. These were available for the project team to use directly or heavily draw upon, allowing the project team to develop exposure databases for all 5 levels of exposure- a process that would not have been feasible without using existing datasets. The data sets were collected, analysed, and metadata was developed using the standards put forth in Section 2. This section presents a description of the 5 databases, including key differences and level of effort. Maps of the resulting exposure that were collected and observations about key differences in assumptions and the impacts on quantification, and finally the results of the probabilistic analysis.

### 4.1.1 Level 1 Data for Los Angeles

Level 1 data was downloaded from GEM’s OpenQuake site. For the US, OpenQuake provides data from FEMA P-366 (Jaiswal et al., 2017) and largely taken from HAZUS. It is less crude than it may be for many other countries, for the US, and may be more like a level 2 in most countries that said, the project team found there were significant differences between this dataset and the one developed for level 2, as described below. Most of the effort required was mapping the structural types to those used by the Seismicat software. There are no meaningful indications of building height or era of construction in this data set.

### 4.1.2 Level 2 Data for Los Angeles

Level 2 data was developed using the exposure data provided with HAZUS 4.2. HAZUS includes data from a variety of sources, but there are only a few key datasets and parameters used to create the default general building stock. The number of structures for residential development is based on the US Census, in this case 2010. The non-residential building stock is largely developed using data provided by a private company specializing in data provision- Dun and Bradstreet. Extending from building count to building size and replacement cost, HAZUS assumes a single model building type and size for each occupancy classification from the data collected by the census and Dun and Bradstreet. These estimates of building size do not have a distribution like the structural values, but only a single value. The replacement cost developed through the per square foot for each of these model building types provided by RS Means, in this case 2018. Unlike the level 1 data, the era of construction is provided through the census data. A height distribution is provided though the structural type in broad ranges but in practice is set to 100% low rise construction.

The level of effort to develop the level 2 exposure database was less than a week. The data tables are provided in a SQL Server database provided with the software and are largely undocumented, requiring an expert to preform and confirm complex joins, with key pieces of information such as units not provided. This required a significant about of sanity checking to confirm the analysis, even though the project team has performed much of this work before. The data was aggregated to 15 arc seconds to make the results easily comparable between levels, and to reduce computation time. This product, though significant processing was required, represent a HAZUS default product.

#### 4.1.3 Level 3 Data for Los Angeles

Level 3 harnesses the capabilities of EO and applies them to the default Level 2 database. The process for developing this dataset was compiled by the project team on a previous project for NASA. As described in Section 2 above, development patterns were extracted from EO data and used to adjust the mapping schemes provided by HAZUS that are used to assign structural types and height. The general classes correspond to areas that are primarily industrial, rural, suburban, multi-family, or commercial business districts of various levels of density of building area. This allows for a more accurate assessment of the likely building type for a given retail structure that is identified in downtown Los Angeles as opposed to an unpopulated area, for example, or an industrial building identified by Dun and Bradstreet in a commercial business district. The distribution of model building types is also used as a key component in the valuation, using the Inhance "ITV", or Insurance to value module (ImageCat, 2019[a]). This represents a simple EO-based enhancement of the default parameters provided in a national dataset to reflect local exposure. The era was also considered, using the HAZUS data from Level 2. The data was aggregated to 15 arc-seconds to make the results easily comparable between levels, and to reduce computation time.

#### 4.1.4 Level 4 Data for Los Angeles

Level 4 takes advantage of many of the key EO-based tools discussed above for Level 3, but supplements the process using EO-based building extraction. For this dataset, building footprints provided by Microsoft were used to develop the estimated number of buildings and the estimated square footage of building stock. Through visual inspection, 2,000 square feet of building footprint was determined to be an adequate delineation between residential and non-residential construction. These data were then aggregated into Based on the height distributions for the various development patterns discussed in level 3, the buildings were "extruded" to reflect the total square footage of building space rather than just the footprint. The non-residential square footage was then distributed into occupancy classes based on the aggregated HAZUS data from Level 2, and these occupancies were used to assign structural classes for each development pattern. As with Level 3, the Inhance ITV module was used to assign the replacement cost, and the HAZUS data was used to assign the era of construction. With the Level 4 data, only the era of construction is derived from the

HAZUS data. The level 4 exposure represents the best solution possible with "empty footprint" data. That is, footprints extracted through EO building extraction with no attributes.

#### 4.1.5 Level 5 Data for Los Angeles

For Level 5, the Los Angeles tax assessor data was acquired by the Los Angeles tax assessor and processed in a manner discussed above. This source of data is completely independent of levels 1 to 4. Although it is more detailed, it may not always be more accurate and was compiled for a purpose other than tax assessment. For each record, a fire code is used to indicate flammability or material type, and given this structure type, height, the use code, and the era of construction a more detailed structural assessment is assumed. This process was modelled on an internal memo completed by Hope Seligson during the preparation of *Data Standardization Guidelines for Loss Estimation Population Inventory Databases for HAZUS MR-1* (ImageCat and ABS, 2006) but was tailored to meet the vulnerability codes of Seismicat. The tax assessor provides the building height in ranges. Late in the development of the exposure data, a building footprint database was discovered that was derived from Lidar data and included height in feet. This was used to assign a height in stories. All other parameters were obtained directly from the dataset. There were no efforts made to adjust the data to account for missing buildings that may not be inventoried by the tax collector or adjust the assessments to reflect a more accurate replacement cost. The result is that key facilities such as ports and airports appear to be missing.

### 4.2 COMPARISONS BETWEEN DATASETS

Table 5 below and the following maps and text provide some initial feedback on the process of collecting the data and how these processes may ultimately have impacted the results. The purpose of collecting this data was to assess the relative difference between levels of effort, and data collection processes. Although discoveries from developing Level 5 data could have been used to augment and improve Level 4 or Level 1 data, this was not the intended purpose. Similarly, EO data might be used to augment the Level 5 data to make certain assumptions more accurate. However, none of these lessons learned were applied to the other exposure datasets in order to avoid artificially decreasing the amount of uncertainty between data sets. In addition, none of these datasets is considered the master data set.

| | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|---|---|---|---|---|---|
| Population | 2010 Census | 2010 Census | 2010 Census | Not Applicable | Not Applicable |
| Number of Buildings | Residential:<br>2010 Census<br><br>Non-Residential:<br>Dun and Bradstreet, 2006 | Residential:<br>2010 Census<br><br>Non-Residential:<br>Dun and Bradstreet, 2010 | Residential:<br>2010 Census<br><br>Non-Residential:<br>Dun and Bradstreet, 2010 | Count: Microsoft Building Database<br><br>Distribution by occupancy for non-res: Dun and Bradstreet, 2010 | Los Angeles Tax Assessor-number of records and Use code |
| Building area | HAZUS estimates per Building | HAZUS estimates per Building | HAZUS estimates per Building, adjusted for height profile with EO data | Microsoft Building Database, digitised area | Tax-assessed area |
| Building Height | HAZUS default (None) | HAZUS default (None) | Height profile by development pattern with EO data | Height profile by development pattern with EO data | LIDAR-derived footprints |
| Era | NA | 2010 Census | 2010 Census | 2010 Census | Year developed field of tax assessor |
| Replacement Cost | 2010 RS Means | 2018 RS Means | Inhance EO-based algorithm | Inhance EO-based algorithm | Improved Value Assessed by Tax Assessor |

*Table 5: Summary of sources of key data parameters for each of the pilot datasets*

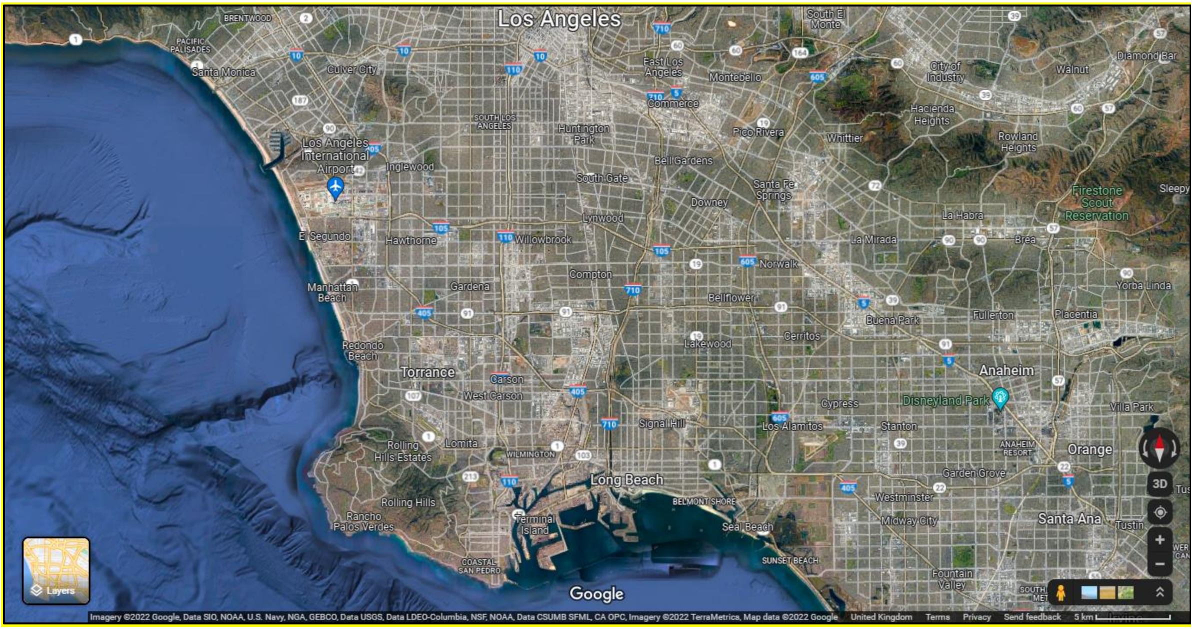


*Figure 4.1: Los Angeles reference map. Source from Google Maps: Imagery@2022*[8]

[8] Imagery @ 2022 Google, Data SIO, NOAA, U.S. Navy, NGA, GEBCO, Data USGS, Data LDEO-Columbia, NSF, NOAA, Data CSUMB SFML, CA OPC; Imagery @ 2022 TerraMetrics; Map data @2022 Google

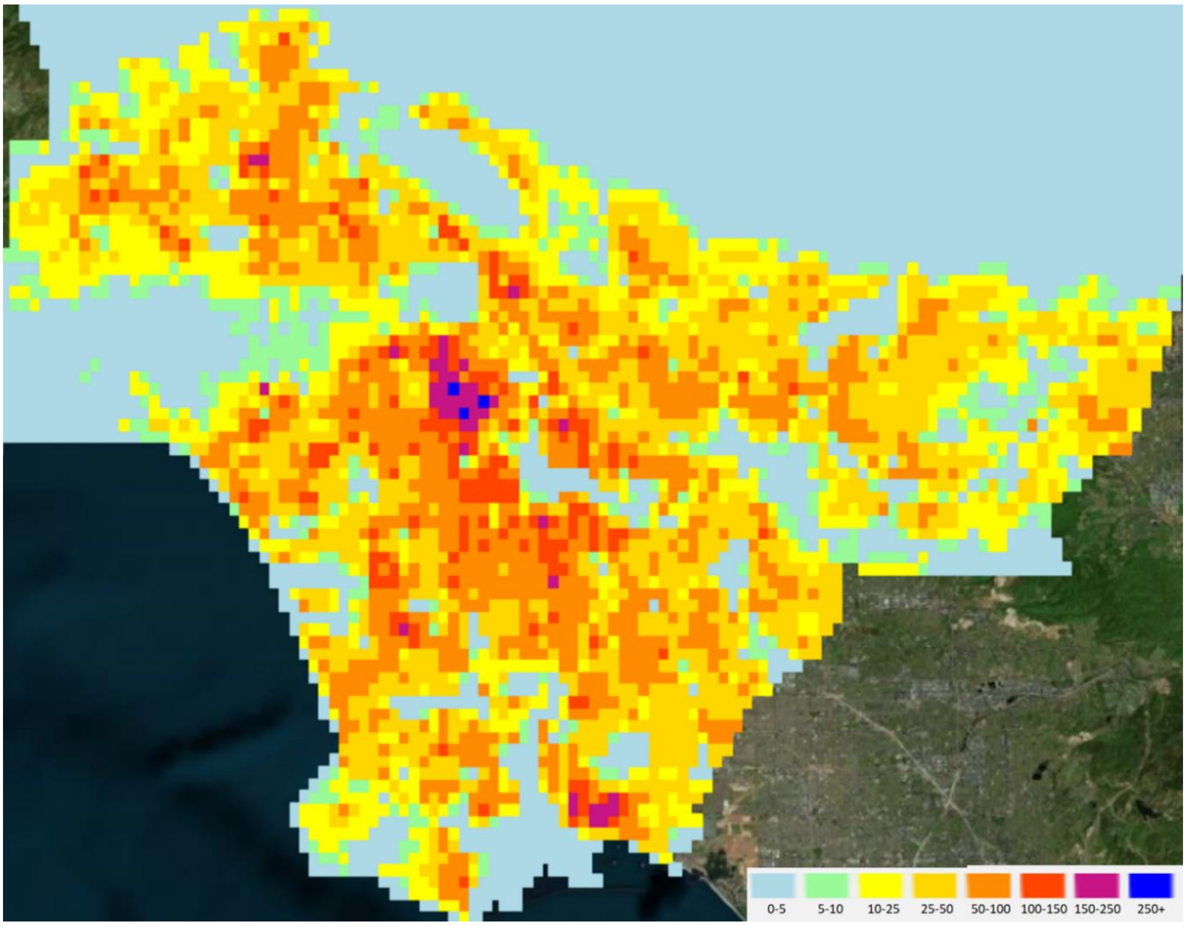


*Figure 4.2: Level 1 Pilot data, Millions of USD of Exposure per 15 arc second (~500 meters squared). Basemap source: ESRI Basemap World Image layer*[9]

The Level 1 data appears to provide a very good quality Level 1 dataset for Los Angeles with key areas of massive exposure clearly visible. Values are significantly lower than many of the other datasets, perhaps due to an earlier provision of Dun and Bradstreet data and RS Means estimate. In addition, there is a low-level exposure throughout the entire county, with the 0-5 Million USD in many unpopulated areas (such as to the North). This also includes the ports and airports.

[9] ESRI Basemap World Imagery - Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

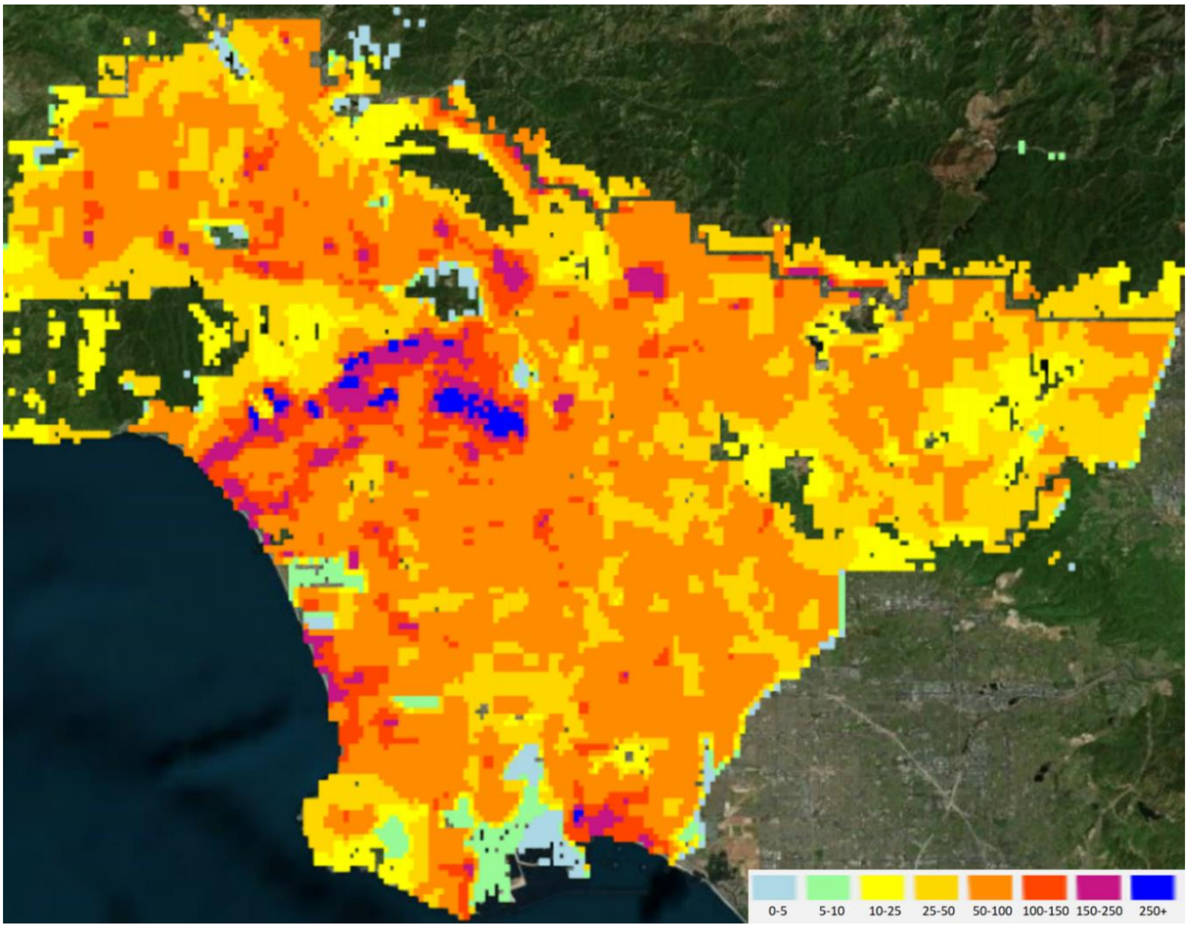


*Figure 4.3: Level 2 Pilot data Millions of USD of Exposure per 15 arc second (¬500 meters squared). Basemap source: ESRI Basemap World Image layer[10]*

For Level 2 Data, the results appear to be higher than level 1, with a significantly more homogenous distribution of exposure. This may be because the differences in methods used to aggregate the data to grid cell (note, the Level 1 data, provided by 30 are seconds, was divided by 4 to yield 15 arc second grids for direct comparison). In the Level 2 data, key commercial districts with high value exposure are more prominent, including a much wider region surrounding the city centre in the centre of the map, and the "Wilshire Corridor" leading through Beverly Hills, West Hollywood down to Santa Monica. Exposure in Long Beach is more prominent as well.

[10] ESRI Basemap World Imagery - Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

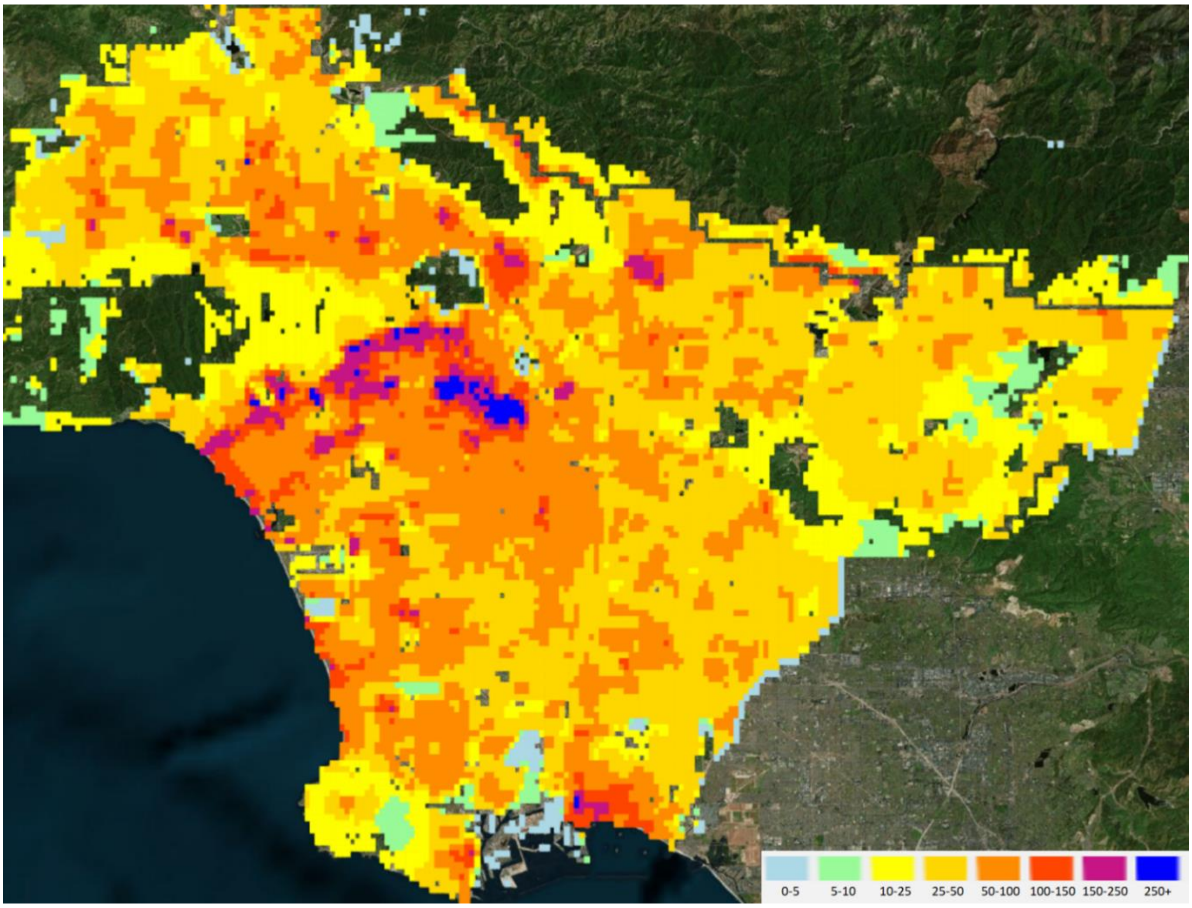


*Figure 4.4: Level 3 Pilot data, Millions of USD of Exposure per 15 arc second (¬500 meters squared). Basemap source: ESRI Basemap World Image layer*[11]

Level 3 data looks very much like Level 2, which is not surprising given the same volume of square footage and distribution of exposure by occupancy. Use of an independent replacement cost generated significantly lower results in the residential building stock. The key inferred attribute- structure type and height, is not evident in the maps and will only be reflected in the results.

[11] ESRI Basemap World Imagery - Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

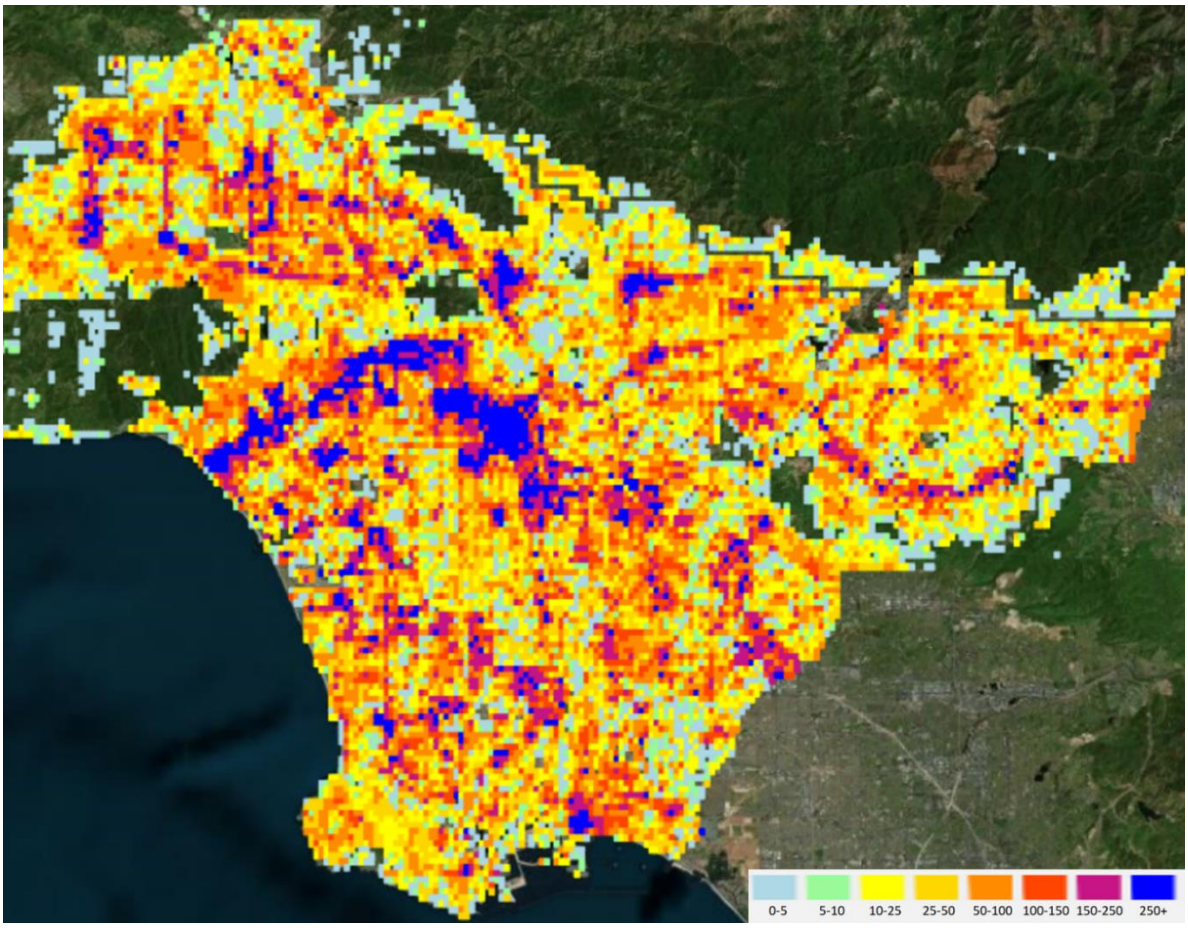


*Figure 4.5: Level 4 Pilot data, Millions of USD of Exposure per 15 arc second (¬500 meters squared). Basemap source: ESRI Basemap World Image layer*[12]

Level 4 data indicates significantly more spatial diversity in the exposure, with many areas that are not included in levels 1 to 3 evident. There are many more areas with the designation of greater than 250 Million USD per 15 arc-second cell, including industrial areas throughout the city. The port is more accurately represented as well, and the entire Wilshire corridor is in the 250+ bin. This is despite the indication that the Inhance ITV replacement cost module appears to be consistent with the HAZUS replacement costs outside of residential areas. The preliminary conclusion here is that there are many types of buildings that may not be reflected in the HAZUS, non-residential dataset, or at least are not reflected accurately spatially. This is a valuable lesson for the key role extraction of building footprints can play in the EO-based development of exposure data sets.

[12] ESRI Basemap World Imagery - Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

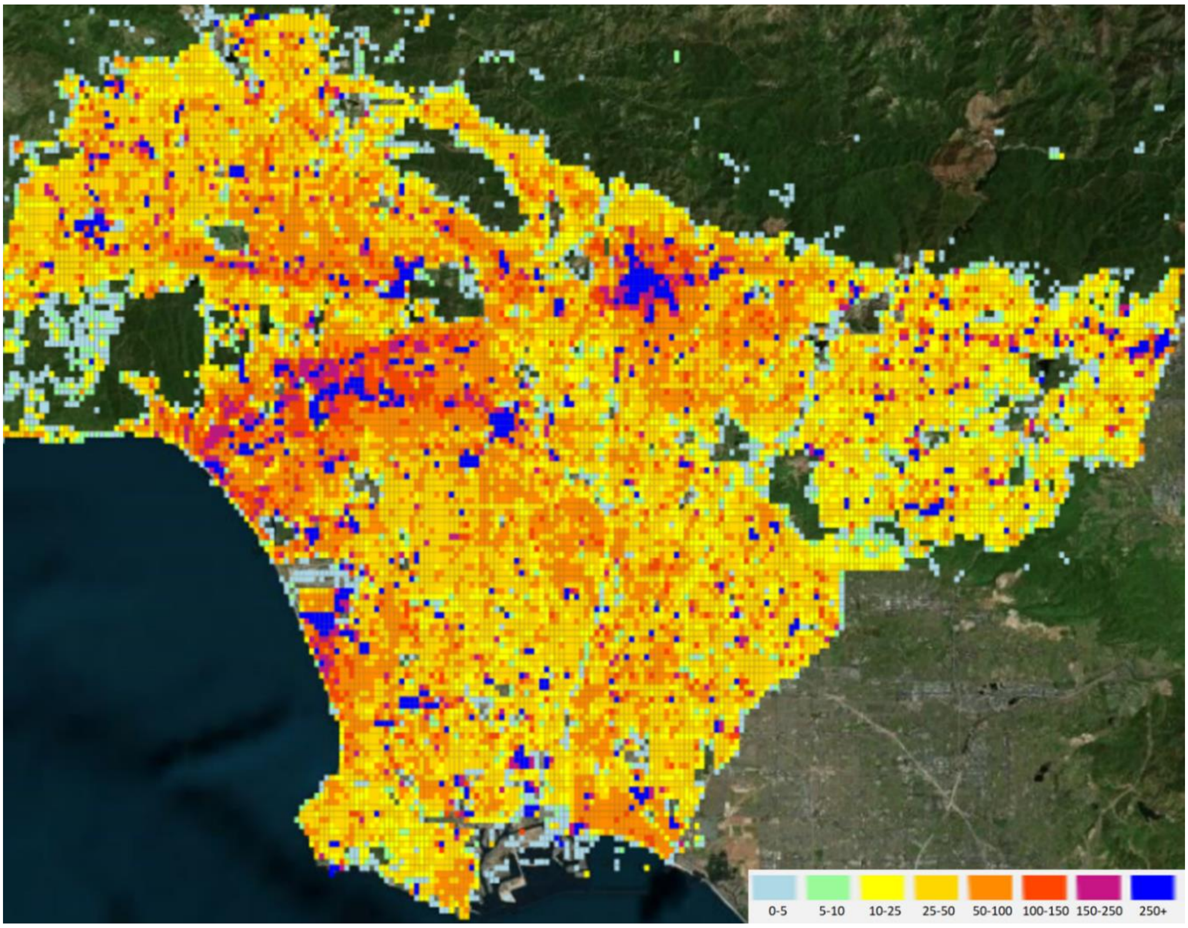


*Figure 4.6: Level 5 Pilot data, Millions of USD of Exposure per 15 arc second (¬500 meters squared). Basemap source: ESRI Basemap World Image layer*[13]

Level 5 appears similar to an average of the values displayed in Level 3 and Level 4. As with Level 4, there are many more pockets of very high value, but unlike level 4, the residential areas appear smooth and consistent. The level 5 data does not appear to capture industrial areas as accurately, with areas like the port (West of Long Beach) or Diamond Bar poorly represented. These are likely due to artefacts given the purpose of data collection, with certain entities except from taxation. Tax Assessor's data. In addition, there may be significant differences between the assessed "improved" value and the replacement cost, including the frequency of assessment and depreciation. These issues are noted for reference but determining the exact reasons for the discrepancies is outside the scope of the pilot study.

[13] ESRI Basemap World Imagery - Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

## 4.3 AGGREGATE DIFFERENCES IN EXPOSURE

Figure 4.7 and Figure 4.8 below present the aggregate square footage and total replacement cost for each of the pilot exposure databases. There are several significant differences between them that are worth exploring. Though it is important to note that the goal is not to identify the correct answer, an examination of how the development of the data may have contributed to these differences is warranted. The replacement cost itself varies by a factor of 2. Given this, with all other factors being equal, one would expect the losses calculated from these datasets to also vary by a factor of 2. Levels 2 and 4 are higher than 1, 3, and 5. For level 2, this appears to largely be due to a higher per unit replacement cost applied within the HAZUS program. To verify this value, the project team checked the results in the original data. HAZUS tables' hzExposureOccupT and hzExposureSBldgTypeT confirmed these values. For level 4, the higher value is due to a significant increase in the square footage. Level 4 depends on area extracted from building footprints, which exceeds the inventoried assets by either HAZUS or the Los Angeles Tax Assessor data by approximately a factor of two. In general, building data extraction with optical imagery would be expected to have errors associated with parking lots or other features with similar spatial characteristics as rooftops, but given a constrained area this is unlikely to be a factor of 2. The HAZUS data may underestimate assets, given the area for all exposure is estimated by assigning a single value given the type of occupancy, and that the none residential construction types are provided through Dun and Bradstreet, where the primary purpose is aggregating business data. Also, the census data for these calculations is dated 2010. Although Los Angeles is largely urbanised, infill development has been a priority in the last several years. The tax assessor, too, will under estimate the exposed square footage. The tax assessor only tracks properties that pay taxes. Properties that are not taxed will not be represented either by square footage or replacement cost.

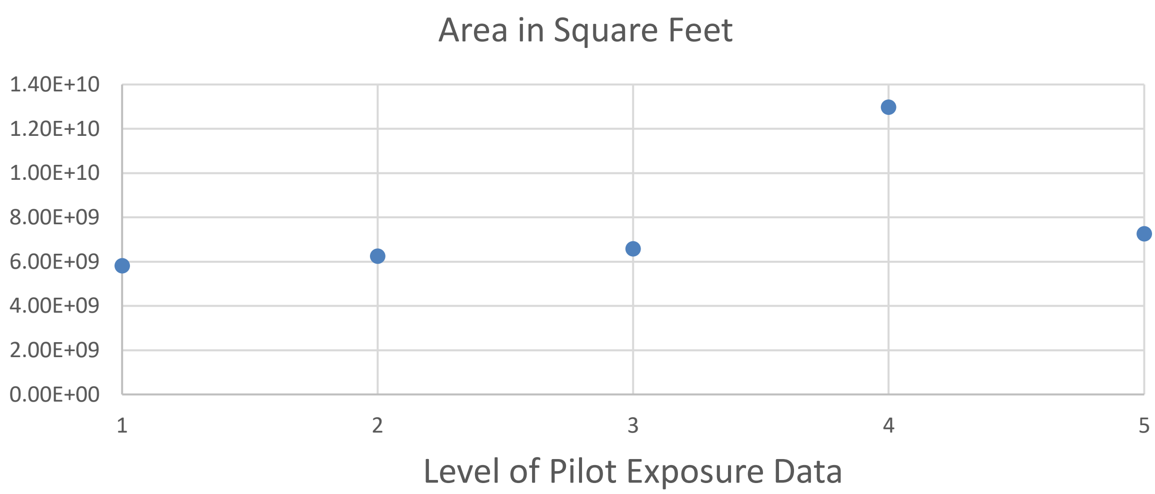


*Figure 4.7: Total square footage of building stock for all 5 levels*

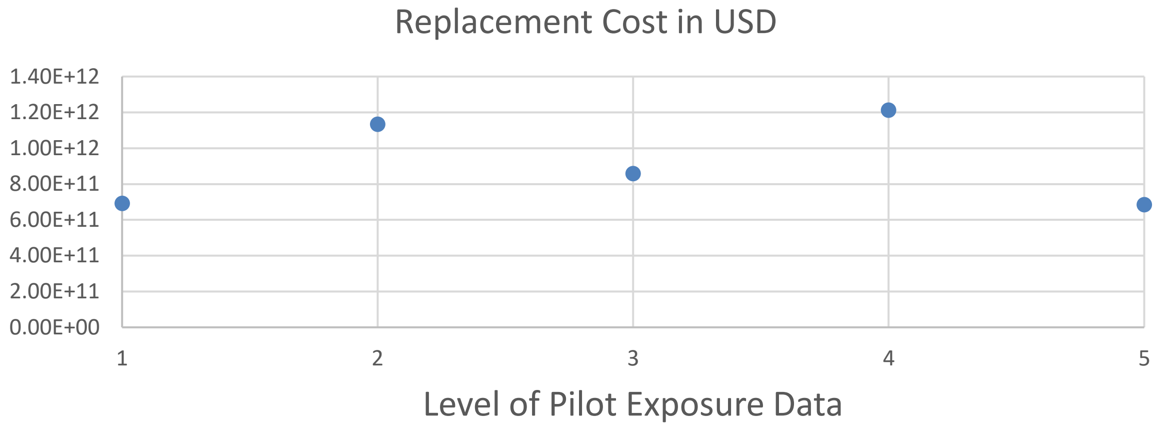


*Figure 4.8: Total replacement cost of building stock for all 5 levels*

To investigate what the square footage of exposed building stock might be, the project team analysed the LIDAR and optical imagery-based building footprints (also used to estimate stories height for the tax assessor data records). Using a simple scalar of 15 feet per story, the LIDAR-based data indicated 17% more building area than the level 4 data, or 1.5 billion square feet of building stock. This is more than twice the area inventoried by the tax assessor. Further analysis of this data found that only 89% of the square footage represented in the database was tracked by the tax assessor. This would indicate that the square footage and replacement cost estimated in the level 4 data is not outside of the realm of possibility and may even be the most accurate estimate. In addition, sampling and additional investigation into the source of these variations based on publicly available data would help to resolve ambiguities for the use of optical and LIDAR based building footprint extraction technologies for the purposes of risk assessment.

## 4.4 PROBABILISTIC EARTHQUAKE MODELLING

The databases discussed above were all run though the Seismicat program. This section provides high level description of the Seismicat software developed by project team partner ImageCat.

The Seismicat portfolio seismic risk tool uses a comprehensive set of individual earthquake simulations, called an 'event set.' Each simulation has a geographic distribution of ground shaking calculated from current ground motion prediction equations (Ancheta et al., 2014) with adjustment for local site conditions. By computing portfolio-wide losses for each simulation, we directly account for the site-to-site correlation of loss within each earthquake event.

The Seismicat event-set follows the methods deployed in the 2014 USGS National Seismic Hazard Mapping (Petersen et al., 2014). The same earthquake source (fault) modelling and ground motion models are used, and the USGS methods are adapted to produce discrete earthquake events. The event-set was produced as part of a Robust Simulation technology for hazard and catastrophe loss analysis with a more complete accounting and disclosure of modelling uncertainties. The Seismicat event set systematically simulates earthquakes on known faults as modelled by the USGS, in each possible fault rupture location, over the full range of magnitudes causing damage, and including background seismicity. Each event simulation provides the spatial distribution of shaking and other hazards. Hazard uncertainties relating to simulations, such as maximum magnitude, fault rupture area versus magnitude, attenuation uncertainties, etc., are carefully accounted for in event set construction and usage.

The vulnerability of buildings, equipment and contents follow methods developed by the authors (Graf and Lee, 2009) called Code-Oriented Damage Assessment (CODA), as well as ATC-13 (ATC, 1985). Probabilistic models based on HAZUS technology are also available, producing "expected loss" results rather than a full statistical distribution. The vulnerability models relate earthquake damage repair costs to earthquake ground shaking intensity as measured by Spectral Acceleration (Sa). For CODA and ATC-13, the variability of damage for a defined hazard state is modelled as a function of the quality of the data, based on the level of engineering investigation (ASCE, 2003; ASTM, 2007; Graf and Lee, 2009).

Geographic correlation of damage and loss is of primary concern in the seismic risk assessment of a geographically distributed portfolio, so the physical size of the source fault rupture must be properly modelled, and the spatial distribution of shaking modelled with appropriate ground motion attenuation relationships.

## 4.5 SEISMIC HAZARD

The Seismicat event set follows the methods deployed in the 2014 USGS National Seismic Hazard Mapping (Petersen et al., 2014). The USGS model has been periodically updated (1996, 2002, 2008 and 2014) and serves as the authoritative national basis for seismic building codes and national risk models (e.g. HAZUS). The same earthquake source (fault) modelling and attenuation relationships are used, and the USGS methods are adapted to produce discrete earthquake events.

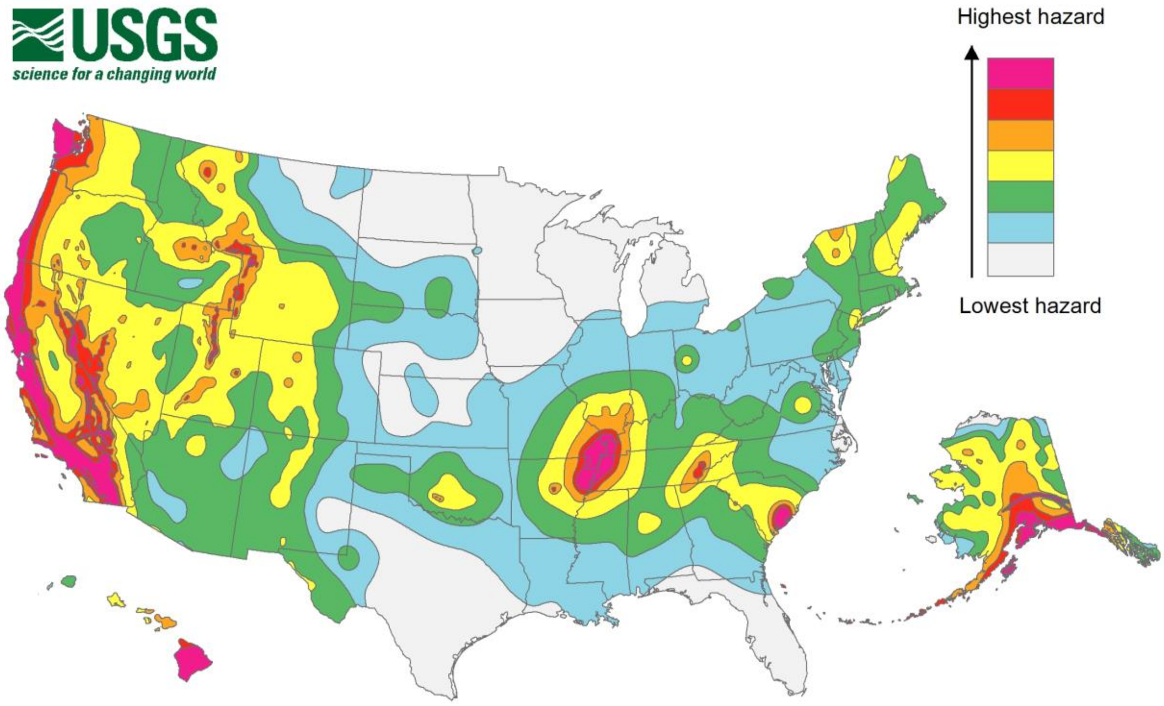


*Figure 4.9: USGS 2014 seismic hazard map of the United States (https://www.usgs.gov/programs/earthquake-hazards/science/2014-united-states-lower-48-seismic-hazard-long-term-model)*

### 4.5.1 Ground Shaking Map from USGS

For a more accurate propagation of uncertainty, The Seismicat event set takes a simulation-based approach, where a 500,000-year time window is used to randomly sample the events diachronically based on the rate of occurrence. The 'event set' systematically exercises the full range of earthquake magnitudes and rupture locations for each seismic source, including known faults and background seismicity. The set of scenarios is carefully constructed so that the ensemble accurately reproduces the severity and frequency of ground shaking for the region of interest, as modelled by the USGS. These simulations usually involve many tens of thousands of scenarios in each complex tectonic region such as southern California, where numerous known and unknown faults exist.

### 4.5.2 Ground Shaking Uncertainty

Empirical ground motion prediction relationships (e.g., PEER's NGA[2] West 2 relationships, Ancheta et al., 2014) are subject to uncertainty, modelled as a lognormal standard error.

The uncertainty in predicting ground motion amplitudes from the future earthquakes is one of the major sources of uncertainty with significantly impact on earthquake risk analysis. The random variability in ground motion prediction from attenuation relations can be partitioned into two parts: the inter-event term *(σ)* and intra-event term *(τ)* (Joyner and Boore, 1981; Abrahamson and Youngs, 1992; Abrahamson and Silva, 1997). The inter-event term accounts for the discrepancy in mean ground motions recorded from earthquake to earthquake. For instance, an earthquake that has a higher than average stress drop is expected to generate ground motions systematically higher than average. The intra-event term, on the other hand, measures the randomness of ground motions across a geographic region from a single earthquake.  The two uncertainty terms are typically treated as independent variables, and the total variance at a single site is the combination of the two terms $(\tau^2 + \sigma^2)$.

### 4.5.3 Site-Specific Hazards

Site ground conditions affect the intensity and duration of ground shaking, as well as the shape of the ground motion response spectrum. In comparison to rock sites, soft soils amplify moderate ground motions, extending the duration of ground shaking, and shifting seismic energy to longer periods.  At very high levels of shaking, soft soils may reduce peak ground motions, compared to rock.  In the Seismicat event set, ground shaking is computed for actual site conditions as determined from regional maps when the soil amplification models are available in the original GMMs (such as the NGA[2] West 2 GMPEs). When such correction models are not available or outdated in the related GMMs, ground shaking is computed using methods consistent with building codes and national standards (IBC, NEHRP, etc.). Where detailed site-specific information is obtained, as from a geotechnical investigation report, the actual ground condition is input and used rather than a mapped condition.

### 4.5.4 Seismic Vulnerability

*Seismicat* adapted the published CODA model (Graf and Lee, 2009) for use in probabilistic seismic risk modelling.  In CODA, shaking-induced damage is a function of a demand-to-capacity ratio (DCR), where the demand is the 5% damped spectral acceleration at the building's fundamental period, and capacity is the product of design strength (Cs, or V/W, at LRFD level) and the Response Modification Factor, R.

$$\text{Mean Damage Factor } DF = f\,(DCR)$$

$$\text{DCR} = \text{Sa (T)} / [\text{Cs x R}]$$

*Equation 1: Mean damage factor*

Because the CODA models utilise these parameters from seismic building codes, CODA models are easily adapted to year of construction and location.  The evolution of building code through time and by location (seismic zones) is straightforward to trace, so the CODA models can make a good initial estimate of seismic resistance if the structural type is known. With engineering investigation, the specific features of the building in question are easily accommodated, using the same engineering parameters found in building codes.

## 4.6 RESULTS OF THE PILOT STUDY FOR LOS ANGELES

The results of the probabilistic risk assessment indicate that the various methods of collecting exposure data contribute significantly to the magnitude of losses. In addition, in this study, even though the exposure data was often collected and assigned vulnerability classifications using very different methods, the volume of square footage and replacement cost were ultimately a more important factor. Finally, there is no clear “winner” in terms of accuracy. Each of the data sets appears to have short comings, but these shortcomings may ultimately counter-balance another short-coming and make the results more accurate. Without further study, the project team is not prepared to indorse one exposure data set over another.

| Losses in USD for 3 Return Periods | | | | | |
|---|---|---|---|---|---|
| RP | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
| 100 | 5.52E+10 | 9.64E+10 | 7.27E+10 | 1.15E+11 | 3.34E+10 |
| 250 | 7.98E+10 | 1.39E+11 | 1.04E+11 | 1.64E+11 | 5.79E+10 |
| 475 | 9.79E+10 | 1.70E+11 | 1.26E+11 | 1.99E+11 | 7.62E+10 |

*Table 6: Losses for all exposure data levels for the Los Angeles pilot study.*

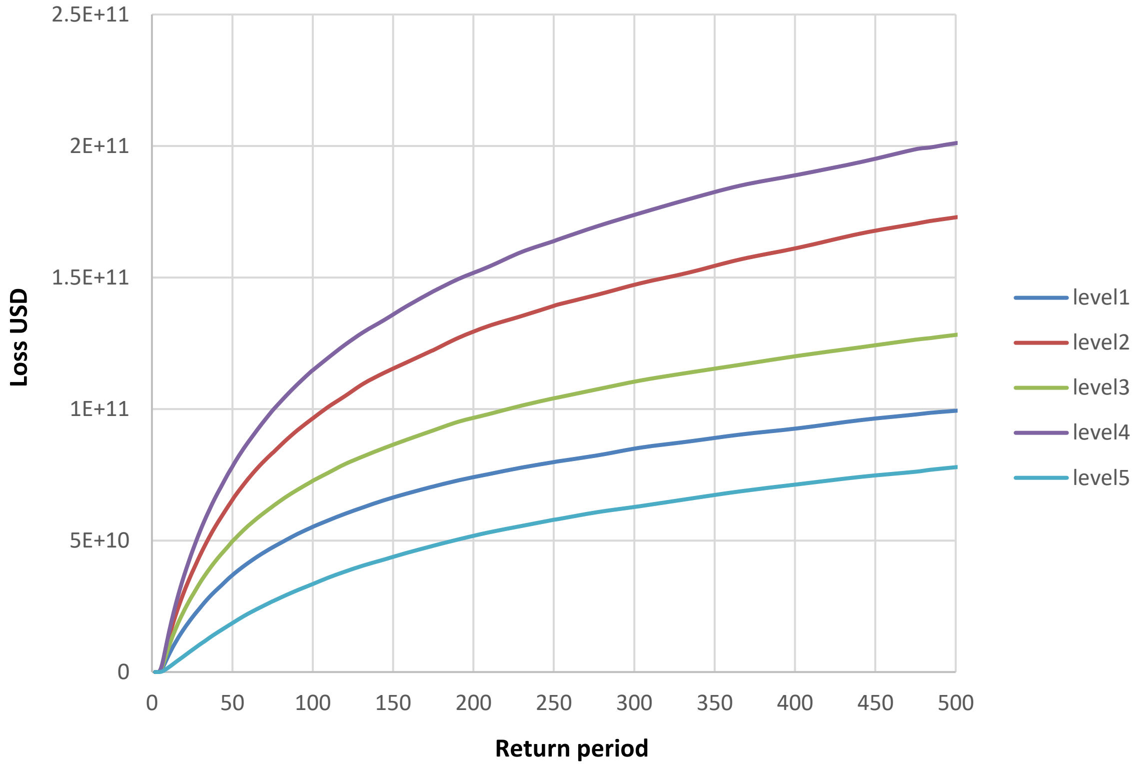


*Figure 4.10: Loss (USD) by Return Period*

Table 6 presents the results for all 5 levels for 3 return intervals. At the 475-year level, the losses vary by a factor of approximately 2.5, ranging from approximately 75 to 200 Billion dollars. The 100-year event varies from approximately 30 to 115 billion dollars, varying by a factor of approximately 3.5. The variation is significant. The volume of data required for this assessment made further sensitivity do to vulnerability and hazard problematic with the context of this pilot test, but given other large portfolio analysis conducted for the Los Angeles basin, these components would be expected to contribute to a similar amount of uncertainty.

### 4.6.1 No clear winner for accuracy

An important finding from this research is that there is no clear winner in terms of accuracy. Precision and accuracy are not always correlated. Table 7 examines some of the issues with each data set. Based on an initial review, Level 4 would appear to be the most dependable source of exposure data in this instance, but further research would be required to confirm. Level 4 is a clear outlier in terms of the calculated results, and decision makers would be well advised to investigate the source of the higher values- the greater volume of building inventory- before proceeding.

| | Advantages | Disadvantages |
|---|---|---|
| Level 1 | More detailed than one would expect for Level 1 Data<br><br>Based on detailed census data | 2010 Census<br><br>Non-residential appears underrepresented<br><br>Primitive mapping scheme |
| Level 2 | Based on detailed census data | Non-residential appears underrepresented<br><br>Significant jump in per-square footage replacement cost from previous data. Not in line with other estimates |
| Level 3 | More accurate mapping schemes<br><br>Spatial variation of assets appears to coincide with exposure<br><br>More detailed assessment of replacement cost (multiple building types, context dependent). | Replacement cost appears to under estimate building stock, perhaps due to division of commercial vs residential structures based on building size<br><br>Dependents upon estimates of number of buildings and square footage from HAZUS data, which encapsulates under-representation of non-residential building stock and 2010 data |
| Level 4 | Building count and square footage depending on data extracted from EO- presumably more accurate than 2010 census data with an assumed area per building type. Appears to agree with highly accurate Lidar/stereo-optical based assessment. | Square footage is a clear outlier from other data sets.<br><br>Likely to overestimate certain building formations. |
| Level 5 | Based on data collected at every location.<br><br>More accurate year of construction.<br><br>More accurate height in stories. | Many buildings not included in tax-assessed records. Much of the square footage appears to be under-estimated or not inventoried.<br><br>Price per square foot for “improvements” does not correlate well with replacement cost figures. Likely due to the assessment cycle and the incorporation of depreciation. |

*Table 7: Advantages and disadvantages of different data sets*

Another key finding is that Level 1 and Level 2, though both based on HAZUS inventories, vary significantly in the average per square footage rebuilding cost. These data sets were processed only a few years apart. If one were to depend solely on the most convenient default data that was available, there would still be significant variation in results. Updating these values alone results in a 60% increase in the value of the building stock.

Lastly, it is important to note most of the difference in results are directly attributable to two factors: square footage and replacement cost. Figure 4.11 below plots 2 ratios for each 5 levels of data- the 475-year loss to the mean 475-year loss of all five levels, and the total exposure in dollars to the mean exposure in dollars of all 5 levels. Thus, the deviation from a 1:1 ratio indicates the relative contribution of the vulnerability assessments to the loss. Level 5 demonstrates a deviation, which is likely due to an independent and more accurate estimate of the year of construction, a more accurate building height, and an independent method of assigning the vulnerability through the fire class code and a detailed occupancy code. Regardless, in most cases the magnitude of probabilistic losses scale with the total exposed value- which is a function of the replacement cost per square foot and the total square footage.

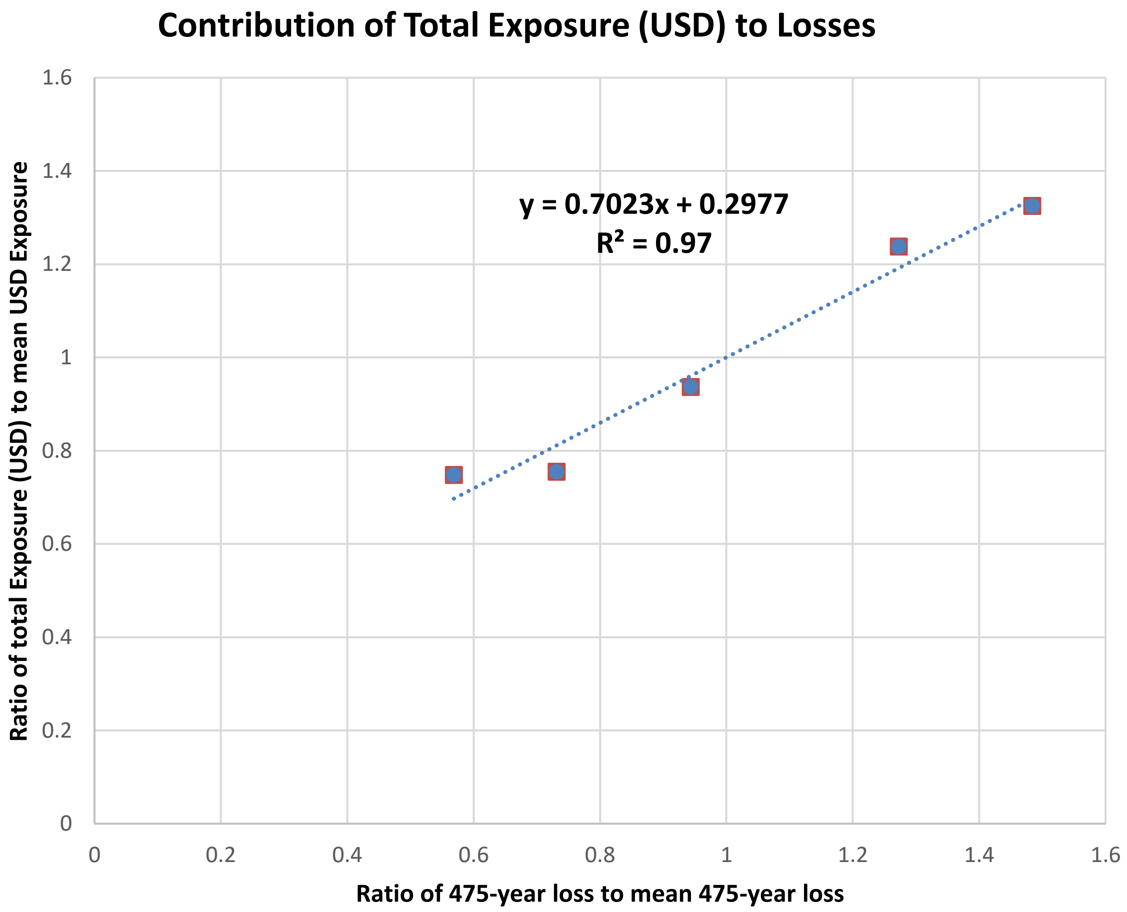


*Figure 4.11: Relative contribution of exposure value to loss*

### 4.6.2 Concluding observations

Precision and accuracy are not the same. For Los Angeles County all 5 data levels were fairly each to assemble, for more detailed data does not mean the results are more accurate. One might assume that the losses would progress to the top, bottom, or middle of the distribution. There was no clear pattern, however, and the conclusion reached is that each data set should be used to sanity check the others to identify the ideal key parameters.

Losses are highly dependent on the accuracy of the exposure data, especially two key values- building area and value per square foot. These two values may overwhelm other factors, so it is important to be confident in the methods to obtain these figures. These may be more important than fine tuning the structural distribution. Sources of exposure area to be clearly defined- with or without garages? What types of buildings are included? Where did the data come from? Are there reasons large swaths would be missing? Is it appropriate to scale a given value? Likewise, with replacement cost- is there a depreciation to the value? Does it include demolition, debris removal, and permitting? What is assumed about the price of materials and the variability? The price of labour- unionised? Was community labour employed? What conversion factors are used, if any, and can the results be scaled based on assumptions with regard to these values that linearly impact the results? All of these questions need to be considered and clarified. Given the relative impact, an exhaustive assessment of these two indicators deserves considerable attention.

# 5 Preliminary assessment of confidence

Given the above discussion of the levels of analysis and key attributes, the table below presents the primary sources of uncertainty given typical development practices and illustrates how important that concern is, typically. However, the text above illustrates the many decision points and limitations encountered when developing an exposure database. Table 8 is simply designed as a preliminary guide.

| Source of uncertainty | L1 | L2 | L3 | L4 | L5 |
|---|---|---|---|---|---|
| Mapping schemes- accuracy of structural breakout, capturing regional differences and local differences. | | | | | |
| Spatial dispersion of assets throughout a tract or administrative region. | | | | | |
| Number and total estimated area of buildings | | | | | |
| Replacement cost per unit area | | | | | |
| Accuracy of administrative units | | | | | |
| Accuracy of census data | | | | | |
| Geocoding Resolution | | | | | |
| Data completeness and accuracy of site-specific data. | | | | | |
| Inference of occupancy | | | | | |

*Table 8: A preliminary guide of the source of uncertainty for each exposure level.*

# Appendix A: ISO-19139 Metadata Mapping

## ISO-19139 Metadata: MD Section

The MD Section of 19139 has a straightforward and direct one to one input where there is a single space for each of the data set metadata information (FGDC, 2018). This section provides general metadata information for the final data set. Information regarding intermediate data sets, processing steps, or field summaries are contained in the data quality section.

The standard metadata information for a data set is found in the identification information section [/MD_Metadata/identificationInfo/] of the 19139 and captures all of the following elements:

1. Data set Description – Abstract
2. Data set Summary - Purpose (or intended use)
3. Format (shapefile, geodatabase, ascii)
4. Data set topology level
5. Spatial Representation Type
6. Geometry (Point, Polygon)
7. Data set object count
8. Tags (Key words- one for each)
9. Data set language code
10. Data set character set
11. Data set environment description page
12. Linking information to ISO-19110 feature catalogue from ISO-19139
13. <u>Credits:</u> Free text space where user is recommended to insert information of responsible parties. The following are the most noted credit types and serve as a guide but are not all required or limited to these options.
    i. Commissioned by
    ii. Distributed by
    iii. Provided/Created by
14. Data Set Spatial Information:
    i. Spatial Resolution Measure
    ii. Spatial Resolution Unit
    iii. Spatial Reference System Code
    iv. Spatial Reference System Code Space
    v. Spatial Reference System Version
    vi. Spatial Extent Type
    vii. Spatial Extent - Bounding Box West
    viii. Spatial Extent - Bounding Box East
    ix. Spatial Extent - Bounding Box South
    x. Spatial Extent - Bounding Box North

15. Data Set Limitation and Disclaimers:
    i. Data set limitations (i.e. not suitable for site-specific hazard analysis such as floods)
    ii. Data set disclaimers (i.e. recommended for screening only, data represent single story buildings)
    iii. Usability Constraints
    iv. Usability License (i.e. Creative Common, Classified, Public, Proprietary)
16. Data Set Citation:
    i. Data set Citation Title (Project Name)
    ii. Data set Citation Date (One for each: Created, finalized, and Published)
    iii. Data set Citation Date Type (One for each: Created and Published)
    iv. Data set Citation Edition or Version
    v. Data set Citation Additional Details (URL)
17. Contact Information: The contact information allows for multiple inputs. The user can add one for each needed point of contact type (commissioner, provider, and owner)
    i. Point of Contact Name
    ii. Point of Contact Organization/Department/Company
    iii. Point of Contact Position
    iv. Point of Contact Role (needs to specify Commissioner, Provider, and Owner)
    v. Point of Contact Email
    vi. Point of Contact Address - Address Number
    vii. Point of Contact Address – City
    viii. Point of Contact Address – State
    ix. Point of Contact Address –
    x. Point of Contact Address – Country
    xi. Point of Contact Phone Number
    xii. Point of Contact Fax Number
    xiii. Point of Contact contacting instructions
    xiv. Point of Contact Office Hours
18. Distribution Information: Note that there is a separate location for distributor information and should not be added to the general contact information as seen above.
    i. Distribution Format Name
    ii. Distributor Name
    iii. Distributor Organization/Department/Company
    iv. Distributor Position
    v. Distributor Phone Number
    vi. Distributor Fax Number
    vii. Distributor Address - Address Number
    viii. Distributor Address – City

ix. Distributor Address – State
x. Distributor Address - Postal Code
xi. Distributor Address – Country
xii. Distributor Email
xiii. Distributor Office Hours
xiv. Distributor contacting instructions
xv. Distributor Role
xvi. Data transfer size

## ISO-19139 Metadata: MD-Data Quality

There are two sections under the ISO-19139 MD- Data Quality section that are used to capture the processing steps, input data or intermediate data, and attribute summaries. These two sections are the lineage section (MD_Metadata/dataQualityInfo/DQ_DataQuality/lineage) and the report section (MD_Metadata/dataQualityInfo/DQ_DataQuality/report). Unlike the metadata identification section (see above) the Data Quality section does not have a direct, one to one, information input location (or slot). There will be multiple 'Process Step' and 'Report' indexes for each and all notable exercises undertaken to develop the final data set.

### Data Quality - Lineage Section

The lineage section is designated for tracking the input databases, data manipulation procedures or methodologies, and data processing. The data input types may include EO-based data products, in-situ observations, open source data, government reports and surveys, and expert opinions. Data processing can give insight as to how a new survey was conducted, how various input data sets are fused or improved upon, and how various modelling or stratification methods are performed.

The lineage structure goes as follows:

1. **Lineage Statement:** Overview statement describing the processes involved in creating the final data product.
2. **Data Source:** There are two 'data source' reference areas where the bibliographic information for the input data can be inserted. The first is a standalone 'data source' section and the second data source location is found within the 'process step' section.
   - 2.1. It is recommended that a complete list of all input data sets is created within this standalone 'data source' section. This complete list should be used as the bibliographic page of the final data set and should not include any description of how individual data sets are implemented in the final data set.
3. **Process Step:** This section provides the space for a process description, description of the rationale behind the process, and data source citations information.
   - 3.1. The processing description and rationale description sections are a free text format that provide ample space for giving a full and detailed description of each.
   - 3.2. Here the secondary data source location is found and should be used for providing information such as the source type and a description or explanation of how the individual source was used or implemented within the processing step. Only a quick citation, author and year, should be added to refer back to the bibliography.

Below is an example of the data source information structure (item A) and examples of the type of processing steps that can be tracked for an exposure database (item B-E).

**A Input data source reference dialogue boxes:**

1. Source Description: The source description should only contain the bibliographic reference or citation. Any additional description of the data set should be directed to the 'Other Details' text section.
2. Medium Name (data type)
3. Scale Denominator (if applicable)
4. Reference System:
   - i. Code
   - ii. Code Space

iii. Version

5. <u>Source Citation</u>
   i. Titles
      a) Title of data
      b) Alternative Title
      c) Collective Title
   ii. Online Resources
      a) Linkage
      b) Protocol
      c) Profile
      d) Name
      e) Description
      f) Function
   iii. Presentation Form
   iv. FGDC Geospatial Data Presentation Form
   v. ISBN
   vi. ISSN
   vii. Dates:
      a) Created
      b) Published
      c) Revised
   viii. Edition
   ix. Edition Date
   x. Series:
      a) Name
      b) Issue
      c) Page
6. <u>Data set Contact:</u>
   i. Name
   ii. Organization
   iii. Position
   iv. Role
   v. Contact Email
   vi. Contact Address: Address Number, City, State, Postal Code
   vii. Contact Phone Number
   viii. Contact Fax Number
   ix. Contacting Instructions

x. Contact Hours

7. Other Details: This text box should be filled with additional data source description information such as:
   i. Data abstract provided by originator (for Data source bibliographic list only)
   ii. The original purpose of the data (for Data source bibliographic list only)
   iii. What or how this source is used (for processing step section only)
   iv. If data is a supplementary aerial or satellite imagery and/or GIS data: What is the download and data-acquisition date (for Data source bibliographic list only)
8. Source Extent:
9. Description: text box
10. Bounding Box: West, East, South, North
11. Geographic Description: Alphanumeric code
12. Temporal Period Extent: Beginning to End Date (for in-situ observations or surveys)
13. Temporal Instant Extent: Instant Date (for in-situ observations or surveys)
14. Vertical Extent: Minimum-Max

**B Data source examples:**

1. Administrative boundary source (WB, GADM, GAUL)
2. Grid Source (National Grid, Custom)
3. Gridded Population (LandScan, WorldPop, GPW)
4. Gridded Urbanity (GRUMP, GHSL, GUF, Facebook)
5. Replacement cost literature or studies
6. Census Data (i.e. Number of people per household, number of schools)
7. Land use (LC2000, CORINE, USGS)
8. Common construction practices (WHE, GEM, PAGER)
9. Site-specific Data (OSM, Tax Assessor, DigitalGlobe)
10. Governmental supplementary or supporting data sources (i.e. Ministry of Health)
11. EO-based imagery (VIIRS, Landsat, TerraSAR-X, Sentinel)
12. Average building size information (OSM, Tax Assessor, Reports, EO)
13. Currency Values (Native country, transfer country, date, days rate)

**C Description of data modelling or stratification procedure**

1. Unit of aggregation (gridded, admin boundary, postal code, census block) explanation
2. Building count distribution methods
3. Population Distribution methods
4. Structural Distribution
5. Development Pattern Schema (General building stock) - Level 1 – 3
6. Occupancy Taxonomy (optional)

7. Structure Taxonomy (How does this differ from distribution?)

**D Building information sampling or surveying procedure**

1. Number of buildings sampled (sample size)
2. Fulfilled Survey parameters (building counts by types, footprint delineations, number of stories, story height, occupancy)
3. Sampling coverage area (randomly selected locations in city, full block coverage, major cities of country)
4. Sampling method: remote sensing using aerial imagery, in-country street survey, GoogleEarth/Bing, combination, statistical decision tree
5. Who did the sampling (number of individuals, group effort, census bureau, supervised classifications)
6. Level of Expertise: Qualifications, Education/training level, collected by untrained surveyor, collected by students, collected by training engineer, combination
7. How were the samples conducted (software/devices [IDCT app, Trimble, GPS-unit], paper survey)
8. How much time was spent on survey (hours, days)

**E Building data set and attribute fusion procedures**

1. Building name assignment
2. Building type assignment
3. Building age/era assignment
4. Building height assignment
5. Building occupancy assignment
6. Building structure assignment
7. Building area assignment
8. Average building area calculation
9. Replacement cost assignment
10. Average replacement cost calculation
11. Data description completeness (% fill by XYZ)

### Data Quality - Report Section

The data quality 'Report' section contains the evaluation of the quality of the input data, processing methods, final data set quality review, and results summaries. The data evaluation consists of a quantitative or qualitative assessment. A statistical quantitative review is not always possible. In that case, a qualitative assessment is conducted. Understanding the steps taken to ensure accuracy (sanity check, peer reviews, and outlier review) may give users a better understanding of the data processing procedure and increase the confidence of the data quality. This section is purely for review; reference information is found under the data source bibliography list in the lineage section.

The Report section has four dialogue boxes:

1. **Measure:** The data evaluator provides a name for the measure (i.e. Building Stratification Method Evaluation) and full description of how the evaluation is conducted.
2. **Evaluation Method:** An evaluation type is designated and a full description of the evaluation results or findings are documented.
3. **Conformance Result:** If applicable, a conformance section is available where the data evaluator will indicate if the data set pass or does pass the conformance test. As well there is a free text space available for providing an explanation of the conformance test.
4. **Quantitative Result:** If applicable, the evaluator is provided space to input statistical findings.

Below is an example of the report metadata structure (item A) and examples of topics or summaries that can be captured through the 'report' section (items B-F).

**A Report dialogue boxes:**

1. Report type options:
   i. Empty
   ii. Completeness Commission
   iii. Completeness Omission
   iv. Conceptual Consistency
   v. Domain Consistency
   vi. Format Consistency
   vii. Topological Consistency
   viii. Absolute External Positional Accuracy
   ix. Gridded Data Positional Accuracy
   x. Relative External Positional Accuracy
   xi. Thematic Classification Correctness
   xii. Non-Qualitative Attribute Accuracy
   xiii. Quantitative Attribute Accuracy
   xiv. Accuracy of a Time Measurement
   xv. Temporal Consistency
   xvi. Temporal Validity
2. Dimension: Horizontal or vertical. This identifies the axe(s) of the spatial quality that is being reviewed.
3. Measure information:
   i. Name of the measure: Short title or statement of the evaluation process that is being conducted.
   ii. Description of the measure: This is a free text structure that givens space to fully describe the purpose of the procedure or summary.
4. Evaluation Method:
   i. Evaluation type options: Empty, direct internal, direct external, and indirect

ii. Evaluation description: The procedure evaluation or field summary results are described in full detail

iii. Evaluation procedure citation (optional, if applicable)

5. Conformance Results:

i. Conformance Results Explanation: Text space available for describing the type of conformance test performed, results, and findings.

ii. Conformance Test Pass (pass, not pass)

6. Quantitative Results:

i. Value Results

ii. Value

iii. Value unit

iv. Error Statistic

**B Evaluate the quality of the input data, methods, or fields; examples:**

1. Average building size method evaluation
2. Replacement cost value evaluation
3. Total number of building count evaluation
4. Building stratification method evaluation
5. Development Pattern execution evaluation
6. Geocoding or geo-referencing evaluation
7. Accuracy of field values

**C Sampling method evaluations:**

1. Where parameter requirements fulfilled
2. Sufficient number of buildings surveyed
3. Sufficient survey coverage
4. How many people reviewed data
5. Review iterations
6. Reviewers qualifications
7. Result sanity check
8. Sampling method medium reliability
9. Surveyor expertise levels
10. Surveyor qualifications
11. Surveyor reliability

**D Building attribute description quality and completeness:**

1. Structure types used evaluations
2. Building height used evaluation

3. Building age or era used evaluation
4. Building distribution method evaluation

**<u>E Summary statements:</u>**

1. Replacement Cost Summary
   i. By administrative level
   ii. By structural type
   iii. By occupancy
   iv. Total replacement cost
2. Number of Buildings Summary
   i. By administrative level
   ii. By structural type
   iii. By occupancy
   iv. Total number of buildings
3. Development Pattern Schema counts by type
4. Geocoding percentage at certain admin units (95% at admin2)
5. Percentage provided by specific source (specific fields with multiple sources)

**<u>F Validation process:</u>**

1. Statistical significances
2. Country census comparison
3. Outlier analysis

## Appendix B: ISO-19110 Metadata: Feature Catalogue Section

The ISO-19139 Metadata standard alone does not have a metadata structure that encompasses feature catalogue metadata. This form of metadata is stored by ISO standard 19110. The feature catalogue metadata is metadata for the individual fields, or attributes, of a database. Below is a list of the type of information that can be captured through the feature catalogue metadata.

1. Data (Feature) set name
2. Data (Feature) set scope
3. Data (Feature) set version
4. Data (Feature) set version date
5. Data (Feature) set language
6. Data (Feature) set character set
7. Data (Feature) set producer information:
   i. Responsible party name
   ii. Responsible party organization name
   iii. Responsible party position
   iv. Responsible party contact information
      a) Telephone number
      b) Fax number
      c) Physical Address (delivery point number, city, administrative area, postal code, country)
      d) Electronic mail address
      e) Hours of service
      f) Contact instructions
   v. Responsible party role
8. Data (Feature) set type (points, line, grid, polygons)
9. Data (Feature) set type name
10. Feature attribute (field) names (field name, label, and alias; UID)
11. Feature attribute (field) definition: Brief description of the purpose of the field. If a field has a summary, such as building count by administrative unit, this description space can be used to direct the user to the data quality 'reports' section where a full summary description can be found. Only short field summaries are recommended for this space.
12. Feature attribute (field) source
13. Feature attribute (field) source reference
14. Feature attribute (field) catalogue listed value: A definition may be assigned to a notable value within the field. This helps users better understand the meaning or context of the given value. Below are the types of definition that can be assigned:
   i. Enumerated Domain: For numerical or text values that are repeating within the field. Here the value is given a definition (i.e. SFR or 3 = single family residential housing).

a) Value label (name)

b) Value definition

c) Value definition source

ii. Range Domain: The range can be established are a numerical value field.

a) Minimum

b) Maximum

c) Mean

d) Standard deviation

e) Unit

f) Measurement resolution

iii. Code set Domain: If the field value uses a code set from an established Value that is specified by a given authority

iv. Unrepresentable Domain: Characterizes values stored in the field that are not enumerated, used to add notes to field.

v. Feature Catalogue listed value definition

vi. Feature Catalogue listed value definition Source (i.e. ImageCat analyst, OSM)

### Relational Tables: Optional linkable tables

The final data set will only contain 6 attribute fields (unique ID, longitude coordinate, latitude coordinate, measured unit [i.e. number of buildings, average square footage, megawatt capacity], unit replacement cost value, and total replacement cost value). The final data set may include a 7$^{th}$ attribute for facility type whenever pertinent to replacement cost values. However, data providers may wish to include additional information that is valuable for a particular study. Below are examples of different types of information that can be provided through a relational table. Please note that much of the type can be captured through the data quality lineage section or the feature catalogue attribute metadata.

Per-record metadata

1. Unique ID*
2. Longitude (centroid of polygon)*
3. Latitude (centroid of polygon)*
4. Megawatt output (energy facilities)**
5. Facility area**
6. Number of facilities**
7. Data location sources
8. Data coordinate resolution type (i.e. XY specific or administrative level 1)
9. Data coordinate resolution type percentage (95% at administrative level 2)
10. Data coordinate resolution processing description
11. Data coordinate resolution disclaimers (i.e. XY- as provided by...; accuracy unknown)

12. Data coordinate resolution validation *(limited – corroborating sources*)
13. Data coordinate resolution accuracy improvement description
14. Data coordinate resolution sources
15. Source URL
16. Source cross-validation
17. Facility name
18. Facility address
19. Facility type***
20. Road/Railroad length in kilometres**
21. Railroad surface condition
22. Railroad surface preparation
23. Railroad surface roughness factor
24. Road national inventory road class
25. Road functional class
26. Number of bridge spans
27. Road/Railroad Hazus-MH class weights
28. Administrative boundary information (code and name)
29. Replacement cost currency exchange rate
30. Description of how facility area is obtained (i.e. area based on a sampled average, or digitized footprint area)
31. Description of how the unit or total replacement cost is applied (i.e. base on expert source, or known country value standard)

* Part of the original 6 attributes for the final data set

** A type of measured unit

*** Used as a 7th attribute field when applicable.

# References

ABRAHAMSON, N.A. & SILVA, W.J. (1997) Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismological Research Letters*, 68 (1), 94-127

ABRAHAMSON, N.A. & YOUNGS, R.R. (1992) A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America*, 82 (1), 505-510.

APPLIED TECHNOLOGY COUNCIL (1985) ATC-13. *earthquake damage evaluation data for California*.

ANCHETA, T.D., DARRAGH, R.B., STEWART, J.P.,SEYHAN, E., SILVA, W.J., CHIOU, B.S.J., WOODDELL, K.E., GRAVES, R.W., KOTTKE, A.R., BOORE, D.M. & KISHIDA, T. (2014) NGA-West2 database. *Earthquake Spectra*, *30* (3), 989-1005.

ANTOS, S. E., LALL, S. V., & LOZANO-GRACIA, N. (2016). The Morphology of African Cities. *World Bank*.

ASCE 31-03 (2003) Seismic Evaluation of Existing Buildings, 2003.

ASTM E 2026-07 (2007) Standard Guide for the Estimation of Building Damageability in Earthquakes. *American Society of Testing and Materials*, 2007.

BRIGHT, E.A., A.N. ROSE, M.L. URBAN & J.J.J. MCKEE. LandScan (2016) High-Resolution Global Population Data Set. No. LandScan 2016 High-Res Global Pop; 005589MLTPL00. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States), 2017.

CENSUS BUREAU, U. S. (2010). *Decennial Census Datasets*. Retrieved from United Sates Census Bureau: https://www.census.gov/programs-surveys/decennial-census/data/datasets.2010.html

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4R20Z93.

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Urban Extents Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4GH9FVG.

CUTTER, S. L., C. T. EMRICH, B. J. ADAMS, C.K. HUYCK & R. T. EGUCHI (2007) New information technologies in emergency management. *In*: Emergency Management: Principles and Practice for Local Government, 2nd edition.

DEPARTMENT OF HOMELAND SECURITY. (2012). Hazus MH 2.1 Technical Manual - Multi-Hazard Loss Estimation Methodology: Earthquake Model. Federal Emergency Management Agency, Mitigation Division, Washington, D.C. *Retrieved from https://www.fema.gov/media-library-data/20130726-1820-25045-6286/hzmh2_1_eq_tm.pdf*

DUN & BRADSTREET (2006) Market Analysis Profile of non-residential properties aggregated by Standard Industrial Classification (SIC) code categories for the purposes of characterizing building exposure for use within the HAZUS modelling program.

DUN & BRADSTREET (2010) Market Analysis Profile of non-residential properties aggregated by Standard Industrial Classification (SIC) code categories for the purposes of characterizing building exposure for use within the HAZUS modelling program. Published 2006 and projected for 2010.

EARLE, P. S., WALD, D. J., JAISWAL, K. S., ALLEN, T. I., HEARNE, M. G., MARANO, K. D., HOTOVEC, A.J., & FEE, J. M. (2009) Prompt Assessment of Global Earthquakes for Response (PAGER): A system for rapidly determining the impact of earthquakes worldwide. *US Geological Survey Open-File Report*, *1131*, 15.p

EGUCHI, R., HOUSHMAND, B., HUYCK, C., SHINOZUKA, M and TRALLI, D., (1999), A New Application for Remotely Sensed Data: Construction of Building Inventories Using Synthetic Aperture Radar Technology. *MCEER Re- search and Accomplishments 1997-1999*, MCEER: Buffalo

ELVIDGE, C.D., B.T. TUTTLE, P.C. SUTTON, K.E. BAUGH, A.T. HOWARD, C. MILESI, B. BHADURI, & R. NEMANI (2007) Global distribution and density of constructed impervious surfaces. *Sensors,* 7, 9, 1962-1979.

FEDERAL EMERGENCY MANAGEMENT AGENCY (FEMA). (2012). Hazus-MH 2.1 Technical Manual: Earthquake Model.

FEDERAL EMERGENCY MANAGEMENT AGENCY (FEMA) (2015) HAZUS-MH Data Inventories: Dasymetric vs. Homogeneous.

FEDERAL GEOGRAPHIC DATA COMMITTEE (2018) ISO Geospatial Metadata Standards. Accessed March 2018 https://www.fgdc.gov/metadata/iso-standards

FOULSER-PIGGOTT, R., BEVINGTON, J., & VICINI, A. (2014). *End-to-end demonstration of the Inventory Data Capture.* Pavia, Italy: GEM Foundation. doi:10.13117/GEM.DATA-CAPTURE.TR2014.06

FRYE, C. (2017) *ESRI's 2016 World Population Estimate Methodology*. An ESRI white paper. Accessed March 2018 https://esri.app.box.com/s/ghgbyvy86jkx90uijmo9j7bmxosbt10f/

FRYE, C. (2008) The ArcGIS citation data model and method. In Meeting of the Social Science History Association, Miami, FL.

GRAF, W.P., & Y. LEE (2009) Code-Oriented Damage Assessment. *EERI Earthquake Spectra*, 25, 1, February, 2009.

HENSHAW, P., SILVA, V., & O'HARA, M. (2018) GED4ALL Global Exposure Database for Multi-Hazard Risk Analysis. D1-Exposure Database Schema and Complementary Tools 2017-10, X pp., *GEM Foundation*, Pavia, Italy.

HOMER, C.G., DEWITZ, J., YANG, L., JIN, S., DANIELSON, P., XIAN, COULSTON, J., HEROLD, N., WICKHAM, J. & K. MEGOWN (2015) Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81, 345-353.

HUIZINGA, J., DE MOEL, H., & SZEWCZYK, W. (2017) Global flood depth-damage functions: Methodology and the database with guidelines (No. JRC105688). *Joint Research Centre* (Seville site).

HU, Z., HUYCK, C., EGUCHI, M., & BEVINGTON, J. (2013) SIDD User Manual, Version 1.0. *GEM Inventory Data Capture Tool Risk Global Component*. Retrieved from http://www.nexus.globalquakemodel.org

HUYCK, C., GHOSH, S., & SELIGSON, H. (2006) Data Standardization Guidelines for Loss Estimation Population Inventory Databases for HAZUS MR-1. *California Office of Emergency Services*. Retrieved from https://www.usehazus.com/docs/loss_estimation_guide.pdf

HUYCK, C., HU, Z., EGUCHI, M., & BEVINGTON, J. (2014) User guide: Tool for spatial inventory data development, *GEM Technical Report* 2014-05 V1.0.0, 60 pp., GEM Foundation, Pavia, Italy, doi: 10.13117/GEM.DATA-CAPTURE.TR2014.05

HUYCK, C.K. & EGUCHI, M.T. (2017) GFDRR Africa Disaster Risk Financing-Result Area 5 Exposure Development. Replacement Cost Refinements to the Exposure Data. *Prepared for World Bank/GFDRR*.

HUYCK, C.K., ESQUIVIAS, G., GAMBA, P., HUSSAIN, M., ODHIAMBO, O., JAISWAL, K., CHEN, R. & YETMAN, G. (2011) D2.2 Survey of available input databases for GED, *Report produced in the context of the Global Exposure Databases for the Global Earthquake Model (IDCT)*.

ImageCat and ABS. (2006). Data Standardization Guidelines for Loss Estimation - Population Inventory Databases for HAZUS MR-1*. California Office of Emergency Services.* Retrieved from https://www.usehazus.com/docs/loss_estimation_guide.pdf

ImageCat. (February 2019[a]) Inhance ITV module. EO-Based building value estimation tool for the United States. *Available at: http://www.inhancedata.com*

ImageCat. (February 2019[b]) SeismiCat. Probabilistic Earthquake Modelling Platform for United States. *Available at: https://www.seismicat.com/*

JAISWAL, K., BAUSCH, D., ROZELLE, J., HOLUB, J. & MCGOWAN, S. (2017) Hazus® estimated annualized earthquake losses for the United States (No. FEMA P-366). *Federal Emergency Management Agency*.

JAISWAL, K., & WALD, D. J. (2008) Creating a Global Building Inventory for Earthquake Loss Assessment and Risk Management. *USGS Open File Report*, OF 2008-1160, 103 pp.

JOYNER, W.B. & BOORE, D.M. (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America*, *71*(6), 2011-2038.

LOS ANGELES COUNTY. (2018, October). GIS Maps*. Retrieved from Los Angeles County Office of the Assessor*: https://assessor.lacounty.gov/gis-maps/

LEE, Y. (2018). Catastrophe Loss Modeling Through Robust Simulation. *In*: Proceedings, Eleventh U.S. National Conference on Earthquake Engineering, June 25-29, 2018, Los Angeles, CA.

LEE, Y., W.P. GRAF & Z. HU (2018) Characterizing the Logic Tree Distribution in the USGS 2014 National Seismic Hazard Mapping Project (NSHMP). *Bulletin of Seismological Society of America*, 108 (3A), 1465–1480

MICROSOFT. (2018, October). US Building Footprints. United States. *Retrieved from https://github.com/Microsoft/USBuildingFootprints*

OPENSTREETMAP. (2018). OpenStreetMap Data. https://www.openstreetmap.org

PETERSEN, M.D., MOSCHETTI, M.P., POWERS, P.M., MUELLER, C.S., HALLER, K.M., Frankel, A.D., ZENG, YUEHUA, REZAEIAN, SANAZ, HARMSEN, S.C., BOYD, O.S., FIELD, NED, CHEN, Rui, RUKSTALES, K.S., LUCO, NICO, WHEELER, R.L., WILLIAMS, R.A., & OLSEN, A.H. (2014) Documentation for the 2014 update of the United States national seismic hazard maps: *U.S. Geological Survey Open-File Report,* 2014–1091, 243 p., *https://dx.doi.org/10.3133/ofr20141091*

POPGRID DATA COLLABORATIVE (2018) POPGRID Data Collaborative, Enhanced Population, Settlement and Infrastructure Data. https://sites.google.com/ciesin.columbia.edu/popgrid/

PORTER, K., HU, Z., HUYCK, C., & BEVINGTON, J. (2014) A Field Sampling Strategy for Estimating a Building Inventory, Inventory Data Capture Tools*. Retrieved from Global Earthquake Model*: https://www.globalquakemodel.org/single-post/2017/05/17/User-guide-Field-sampling-strategies-for-estimating-building-inventories

RS MEANS (2014) RSMeans Square Foot Costs*.* (35th Ed.) RS Mean Company. Retrieved from https://www.rsmeans.com/

RS MEANS (2018) RSMeans Square Foot Costs (39th Ed.). RS Means Company. Retrieved from https://www.rsmeans.com/

STEINBERG, S. (2014) Countywide Building Outlines – 2014 Update – Public Domain Release. *Retrieved from Los Angeles County GIS Data Portal:* https://egis3.lacounty.gov/dataportal/

TAYLOR, C. E. (2015) Robust Simulation for Mega-Risks: The Path from Single-Solution to Competitive, Multi-Solution Methods for Mega-Risk Management. *Springer International Publishing*.

UN-HABITAT (2007) Housing in the world- Demographic and Health Survey

VERRUCCI, E., BEVINGTON, J., & VICINI, A. (2014) Application of the GEM Inventory Data Capture Tools for Dynamic Vulnerability Assessment and Recovery Modelling. *In*: EGU General Assembly Conference Abstracts, Vol. 16.

VINAY, S., R. CHEN, M. BECKER, C. K. HUYCK, Z. HU, D. CAVALCA, E. GOLDONI, P. GAMBA, P., & JAISWAL, K. (2012) Description of the Global Exposure Database Schema. *Report produced in the context of the Global Exposure Database for the Global Earthquake Model (GED4GEM)*, GEM Foundation, Pavia, Italy.

WORLDPOP (2018) World population data products. http://www.worldpop.org.uk/data/get_data/