Technical Brief - 2022

Pilot study – autonomous identification of informal settlements in Pacific Islands using machine learning and satellite imagery

Impact and Innovation Grant

Port Vila and Suva – July 2022

International WaterCentre | Griffith University Queensland, Australia | UACS urban analytics & complex systems | USP The University of the South Pacific | Water For Women | Australian Aid

Full research title: Impact and Innovation Grant: Planning for climate-resilient urban water, sanitation and hygiene (WASH) in informal settlements in Melanesia

Research Partners:

With the support of:

Visit our website

Version history: FINAL V1, July 2022

Suggested citation: Patorniti, N., Vieira, A.S., Dashwant, A., Da Silvera, M. C., Sanderson, R., and Souter, R, 2022 *Pilot study – autonomous identification of informal settlements in Pacific Islands using machine learning and satellite imagery*. International WaterCentre at Griffith University. Brisbane, Australia

Cover images: Informal settlements in Suva, Fiji. Photograph captured by G. Ratidara

Introduction

Background

The International WaterCentre at Griffith University, UACS Consulting, and the University of the South Pacific are currently conducting research supported by the Water for Women Fund under the Australian Aid program, entitled *Planning for climate-resilient urban water, sanitation and hygiene (WASH) in informal settlements in Pacific Islands.* The research focuses on four cities in the Pacific, including: Port Moresby, Suva, Port Vila and Honiara. The key research question is:

> *How can urban planning processes in Melanesia be strengthened through participation and integration to improve the resilience of WASH service delivery in informal settlements and areas identified for housing growth within the urban footprint?*

The research aims to investigate how spatial and non-spatial datasets assist in the development of resilient WASH service models. Thus, it focusses on datasets that can be integrated to create city-wide WASH service delivery maps to assess the resilience of different WASH service delivery models in informal residential settlements.

Sliuzas and Mboup (2008) describes the creation and gradual spontaneous growth of informal settlements, including three phases of development from infancy to urban maturity (**Figure 1**). Informal settlements can present different socio-demographic and spatial characteristics at different developmental phases; although, population density, housing durability and security of tenure tend to be featured at all development phases and are broadly across informal settlements of different cultural backgrounds (Niva, Taka, & Varis, 2019; Owen & Wong, 2013).

FIGURE 1: ORGANIC DEVELOPMENT STAGES OF INFORMAL SETTLEMENTS (SLIUZAS & MBOUP, 2008).

A challenge faced by countries and cities when addressing the needs of informal settlement residents is a lack of accurate, relevant and timely data to describe demographics and living conditions within the settlements (Niva, Taka, & Varis, 2019; UN Habitat, 2015). Several global organisations are actively quantifying and expanding datasets to improve information and knowledge about informal settlements, including the World Bank's Urban Development projects, the UN's World Urbanization Prospects (UN Habitat, 2018) and UN-Habitat (2016). However, datasets often are not able to sufficiently describe the experiences of residents and their ability to access adequate WASH services in informal settlements. In Melanesia, a lack of timely data on urban populations and demographic changes, let alone incomplete knowledge of the location and extent of informal settlements, converge to create a challenging environment in which to find accurate and reliable data that can be used for planning purposes.

Pilot description

The research team undertook a pilot study based on machine learning techniques applied to remote sensing to autonomously detect urban landscape features from multi-spectral satellite imagery to assist in locating

and mapping informal settlements. This pilot was intentionally designed as an accessible and low-cost process. It is conducted remotely without the need for in loco investigation using open-source data and programs. Techniques are simple to use and deploy, and have resulted in a new approach to map dense urban settlements in Melanesia. The pilot aims, methodology, results and future directions are detailed in the pilot study section of this report.

Potential uses

Research efforts have been conducted towards fit-for-purposed solutions informed by tools tailored to inform end-users. In accordance with this theme, during the concurrent research activities under the project, stakeholder interviews (i.e., consultation process with urban planners and utility providers) in the region revelled the lack of comprehensive and up-to-date spatial information for informal settlement locations and extents. Government at different spheres face challenges to keep abreast of informal developments through up-to-date data and data collection systems.

To address such limitations, an autonomous process that can be conducted using open-source multi-spectral satellite imagery and low-resource settings can be implemented as an initial assessment to locate informal settlements within administrative areas across regions and countries in Melanesia. Notwithstanding such an autonomous process cannot reveal all important information about informal settlement areas (i.e., socioeconomics, populations and preferences), it provides means to identify the extent and density of informal settlements. Therefore, it may be a useful planning support tool for government authorities and other stakeholders involved in the provision of resilient WASH services.

The information provided by the tool developed to autonomously detection informal settlement across large areas can be then detailed by existing local socio-economic and planning datasets (e.g., population and dwelling growth, zoning, etc.), as well as new datasets from targeted surveys of areas identified by the tool as undergoing land use change related to informal settlements (e.g., new informal settlements, densification of existing informal settlements, etc.). In Suva and Fiji for example, growth pressure areas are identified through the number of applications for building and rezoning received for those areas, as well as the population in that area. Planning stakeholders have indicated that such information is used to indicate that changes in planning guidelines is needed. Suva City Council (SCC) attempt to collect this information and spatially record it. A process that helps to identify informal settlements, such as described here, would be a complementary tool to such analyses.

Limitations

It is recognised that informal settlements in the Pacific Islands are diverse in terms of both socioeconomic and geographical characteristics. For instance, some areas have high urban density and are significantly constrained for expansion (e.g., Nanuku in Suva), while other areas are more peri-urban and with lower residential densities (e.g., Blacksan in Port Vila). The process of settlement improvements in some cities has regularised infrastructure services (e.g., road improvements). As a result, some settlement areas that are more established, such as Freshwota in Port Vila, have more physical characteristics in common with formal settlements. However, such obvious formalisation attributes (e.g., road improvements) are not always reflected in the provision of adequate WASH services. The different of service standards in such areas may pose challenges to detect and classify informal settlement areas.

The pilot described herein used key physical characteristics to categorise different land covers to recognise settlements and then further separate formal from informal settlements. Building density was a key characteristic used in this regard and was applied using a threshold value of urban buildings density shown in the imagery. Low density informal settlements thus were not automatically identified using this process. Further work using other typical physical and demographic characteristics can be undertaken to create representative samples of low-density informal settlements, which, in turn, will improve the models to identify

low density informal settlements at a higher confidence level. , For example, disaggregated census data revealing average household size at an appropriate scale could be used to strengthen this identification. This trial was limited by the accuracy and representativeness of freely available data only.

Ethical considerations

Mapping informal settlements can present ethical questions of privacy, where persecution or eviction may occur, and more broadly how inclusive remotes sensing and remote mapping processes can be (Klemmer, Yeboah, de Albuquerque, & Jarvis, 2020; Oluoch, Kuffer, & Nagenborg, 2022). However, the use of GIS and remote sensing can serve to better identify and highlight informal settlement areas across an urban environment, thus increasing the potential for improved services, support and agency within those communities. Oluoch (2019) provided a code of ethics for GIS for informal settlements that includes individual or group rights to privacy, data uniformity and public data access, and this project seeks to abide by that code.

Case study - Identification of Melanesian urban and peri-urban informal settlements using machine learning and low resolution multi-spectral satellite imagery

Study aim

To develop a tool for the autonomous detection of informal settlements across large scales in Melanesia using accessible approaches. In the context of the study, 'accessible approaches' encompass freely available datasets with complete geographic coverage, low-cost methods requiring average technical skillsets, desktop survey techniques, and low resource intensive initiatives (i.e., low number of people required to perform the approach).

Method

The tool to detect informal settlements in the Pacific Island was developed using a pixel classification technique based on Random Forest algorithm and open-source multi-spectral satellite imagery from Copernicus. The pilot study area encompassed informal settlements in the Port Vila region, Vanuatu, and the west side of Fiji's largest island Viti Levu where the capital, Suva, is located. Random Forest has been also successfully used by Gram-Hansen et al. (2019) to classify informal settlements in other regions of the world. Figure 1 provides a summary of the pilot study method.

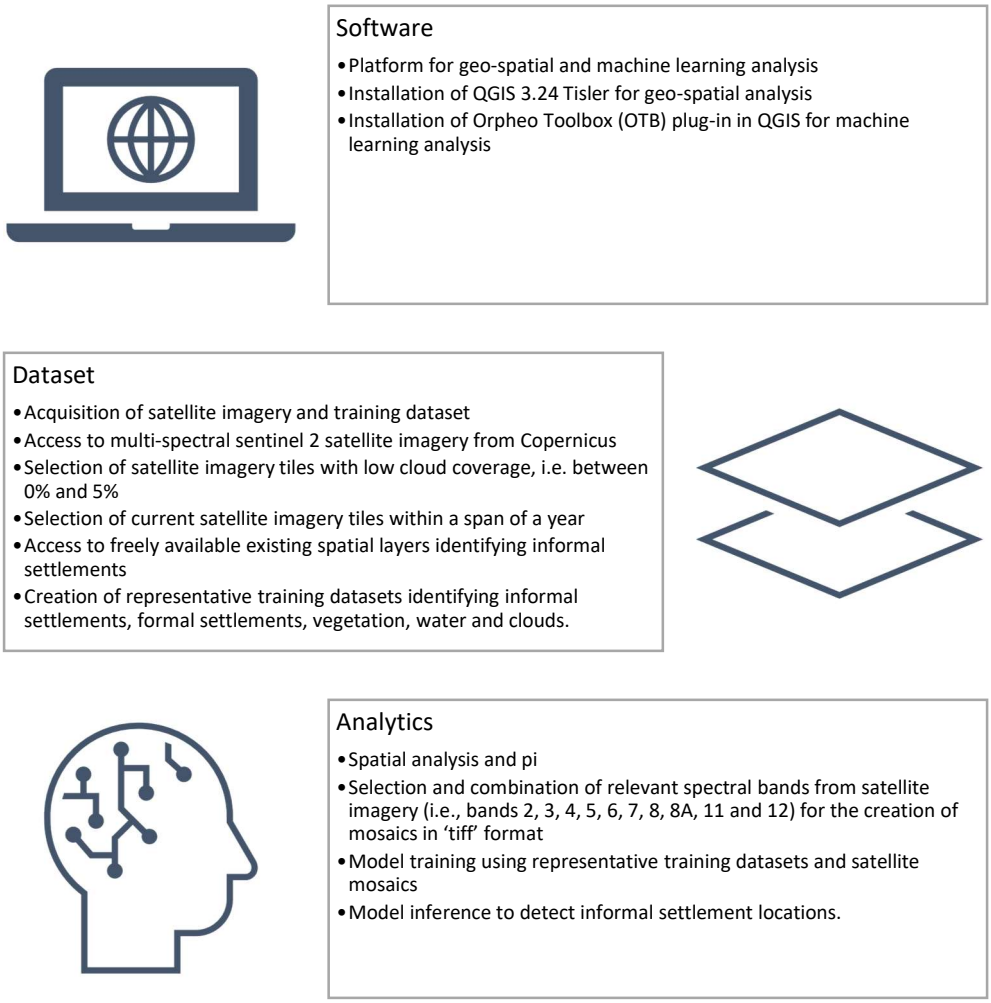

FIGURE 1 - FLOWCHART OF THE PIXEL CLASSIFICATION METHOD.

QGIS 3.24 Tisler, a free and open-source geographic information system from the Open-Source Geospatial Foundation Project, was selected to undertake this remote sensing exercise in accordance with the use of accessible approaches in this project. In addition, QGIS is a commonly used GIS platform across the Pacific and is promoted for use by the Pacific Community (SPC). The plugin Orfeo ToolBox (OTB) was installed in QGIS to enable pixel classification capabilities (Figure 1).

The pixel classification process to identify informal settlements was undertaken using imagery that is current, has good coverage and is freely available at no cost. For this purpose, the project used the Copernicus satellite imagery, which is an European constellation of satellites that provides free of charge images with pixel resolutions of 10 m, 20 m, and 30 m with worldwide coverage for imagery from the Sentinel-1 and Sentinel-2 satellites. In this project, low-resolution image from Sentinel-2 covering regions in the Port Vila, Vanuatu, and the west side of Viti Levu, Fiji's largest island.

The pixel classification model was trained to identify different land covers and elements using a collection of samples from the defined classes, including: informal settlement, formal settlement, vegetation, water, and clouds. The classes identifying formal and informal settlements, being the subject matter of this project, were sampled with higher precision, quantity and representativeness. To better identify where to sample for training informal settlement areas, tessellation layers were made which provided information of some areas known to have informal settlements. The boundaries of known informal settlements and their locations were provided as a vector layer (shapefile) from a study completed by the International WaterCentre in October 2020 (Sanderson & Souter, 2020). The water class did not need extensive sampling due to its low variability; most bodies of water are uniform as they have unique spectral signatures. Similarly, vegetation also required limited sampling as the spectral signature of vegetation is easily distinguishable from other classifications. Cloud samples were also needed as most images come with at least a small percentage of them included, so recognition of cloud by the model in imagery was required.

Once all data inputs were collected and the model was prepared, the model classification output was obtained using the Random Forest Classification using the Orfeo Toolbox plug-in in QGIS by following the following steps:

- To compute the 'Image Statistic', these steps are followed: 'Processing Tools' → 'OTB' → 'Learning' → 'ComputeImagesStatistics'. Then, select the image to be used to train the model and select a place to save the statistics folder. The '.xml' extension must be added to the name of the file;
- To train the pixel classifier, use the function 'TrainImagesClassifier' following the steps: 'Processing Tools' → 'OTB' → 'Learning' → 'TrainImagesClassifier'. For this process, it is necessary to add the image again, then in 'Input Vector Data List' add the shapefile containing all of the samples and add the image statistics from the previous step in the 'Input XML image Statistics file'. The other fields are left as default, with a 'training x validation scale' of 0.7. Select the field in the vector file that contains the class number identification. Select the 'RF' method and the number of decision trees used, in this case 100. Then save the output model file and add the the extension ". model" at the end of the file name and the extension ".csv" for the 'Confusion Matrix' file;
- The final step is to classify the image pixels using the 'ImageClassifier' function following the steps: 'Processing Tools' → 'OTB' → 'Learning' → 'ImageClassifier'. In this function, the target image is used to create inferences in the area of interest, then the model creates image statistics identifying land use classes for each pixel in the image. Finally, inferential results can be saved, including the predicted class, confidence and probability maps.

In tile 1 of Figure 2, the input low resolution multispectral imagery with pixel size of 10mby 10m from Sentinel 2 satellite is shown. In tile 2 of Figure 2, the Random Forest pixel classification model output is detailed. It shows each 10m by 10m pixel of the input imagery classified into the discrete classes set for the model.

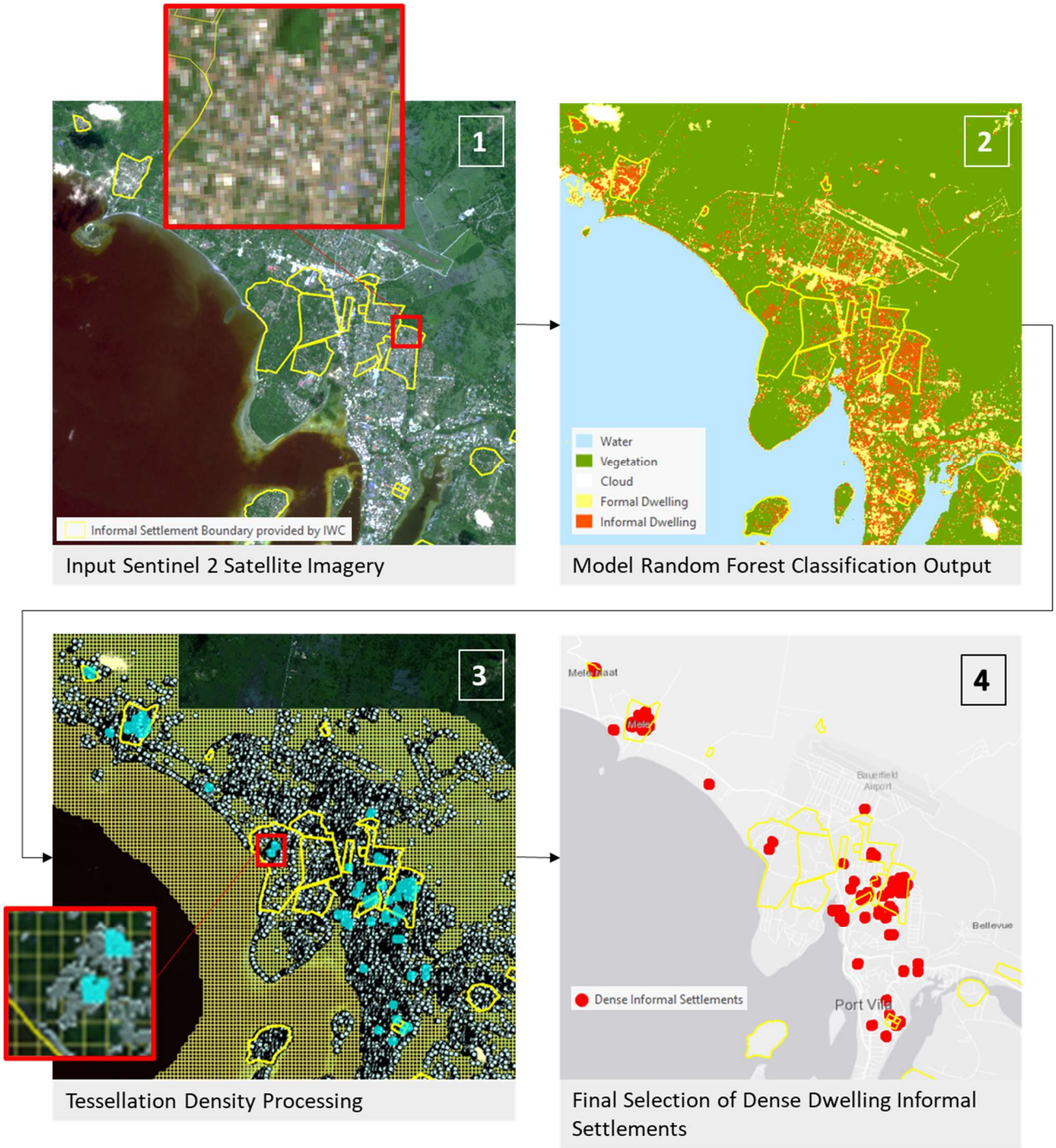

FIGURE 2 - RESULTS OF THE RANDOM FOREST CLASSIFICATION ON PORT VILA, VANUATU.

Following the process demonstrated in Figure 2, the model output was post-processed to identify the denser clusters of informal settlements. The post-processing results were then used to improve the classification of dense informal settlements, formal settlement areas and dispersed informal settlements. Tile 3 of Figure 2 illustrates the following post-processing step, summarised as follows:

- Create tessellation with grid of 75m by 75m across study area
- Convert raster pixels into centroids
- Summarise informal dwellings detected per tessellation cell
- Set threshold of informal dwellings per tessellation cell to select centroids within the dense cell. A threshold of 43 centroids per cell was used in the Vanuatu example.

Dispersed informal dwellings are identified by the model but are not captured in the post-processing step as they are inadvertently removed to identify the dense informal settlements. Detection of dispersed informal settlements are discussed briefly in the future research directions and application section below.

Results and discussion

Tile 4 of Figure 2 details the dense informal settlement areas selected from the tessellation density post-processing step. Annexure A provides extracts from the Vanuatu model. Results demonstrate the modelling approach provides a cost-effective method to detect dense informal settlements. For example, the modelling approach was able to autonomously detect the informal settlements using satellite imagery. The validation of model results showed the agreement of observed and modelled classes. Therefore, the developed method has advantages in relation to manual approaches, which are generally time-consuming and result in incomplete identification of targets. Results show the ability to detect settlement areas in general, including even dispersed formal and informal dwellings in the outer centre and peri-urban areas.

The Port Villa (Vanuatu) model was then upscaled to examine the model's ability to; 1) detect informal settlements in other pacific islands that the model had not been trained for; and 2) the model's ability to process large areas (i.e., country scale) on accessible resources. Annexure A provides extracts of the Fiji model. Results of the Fiji model detail the model's ability to detect dwellings and other built form structures using desktop resources. It does, however, provide evidence to support increased sampling in Vanuatu and then possibly sampling in Fiji to increase the number of training images. This will assist the model to more precisely differentiate between formal and informal dwellings. Further sampling will also help remove the incorrect classification (i.e., noise) surrounding the cloud classifications

Future research direction and application

A summary is provided of future research directions and applications of the approaches developed to be applied for the identification of informal settlement in Pacific Island countries. The summary is focused on the provision of water sanitation and hygiene services (WASH) regarding climate change and future population changes in Pacific Island cities.

- Refinement of detection models for informal settlements to enhance the identification of the location, size and density of informal settlements.
- Analytics can be applied to this new data to estimate the population and demand for WASH services. For example, each10m by 10m pixel classified as 'informal settlement' could represent 1.5 informal dwellings with an occupancy of5 people per dwelling for a total population per pixel of 7.5 people and a water demand of 100 litres per person per day for a total of 750 litres per pixel.
- Use for change detection to examine increases or decreases in informal dwellings. This would help identify informal settlement areas early by period detection model runs.
- Monitor informal settlement growth over time to scale up and match the provision of WASH services to local demand and changing needs.
- Use detection models with infrastructure network mapping to identify and examine areas outside service areas for water and sewerage
- Combine other data to help understand the characteristics of the informal settlement areas to better plan for their optimal servicing and resiliency.
- Investigation of ethical considerations with the application of this research.

References

Gram-Hansen, B. J., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V., & Bilinski, P. (2019). *Mapping informal settlements in developing countries using machine learning and low resolution multi-spectral data.* Paper presented at the Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.

Klemmer, K., Yeboah, G., de Albuquerque, J. P., & Jarvis, S. A. (2020). Population mapping in informal settlements with high-resolution satellite imagery and equitable ground-truth. *arXiv preprint arXiv:2009.08410*.

Niva, V., Taka, M., & Varis, O. (2019). Rural-urban migration and the growth of informal settlements: A socio-ecological system conceptualization with insights through a "water lens". *Sustainability, 11*(12), 3487.

Oluoch, I. (2019). *Mapping Out Responsibility: An Investigation of the Cartography of Slums.* University of Twente,

Oluoch, I., Kuffer, M., & Nagenborg, M. (2022). In-Between the Lines and Pixels: Cartography's Transition from Tool of the State to Humanitarian Mapping of Deprived Urban Areas. *Digital Society, 1*(1), 1-21.

Sanderson, R., & Souter, R. (2020). Investigating the transmission of faecal pathogens in urban informal settlements in and around Port Vila, Vanuatu, recognising biophysical and demographic diversity – Faecal pathogen transmission exposure typologies. In *(Research Brief)* (pp. 13). Brisbane, Australia: International WaterCentre, Griffith University.

Sliuzas, R., & Mboup, G. (2008). *Diversity of Global Slum Conditions—Is a Universal Spatial Definition of Slums Feasible.* Paper presented at the Report of Expert Group Meeting on Slum Identification and Mapping.

UN-Habitat. (2016). *Slum Almanac 2015-2016*. Retrieved from New York: https://unhabitat.org/sites/default/files/documents/2019-05/slum_almanac_2015-2016_psup.pdf

UN Habitat. (2015). *Habitat III issue paper 22—informal settlements*. Paper presented at the Paper presented at the Habitat III, United Nations Conference on Housing and Sustainable Urban Development, New York.

UN Habitat. (2018). *World Urbanization Prospects: The 2018 Revision*. Retrieved from https://population.un.org/wup/publications/Files/WUP2018-Report.pdf

Annexure A – Port Vila, Vanuatu Model Outputs

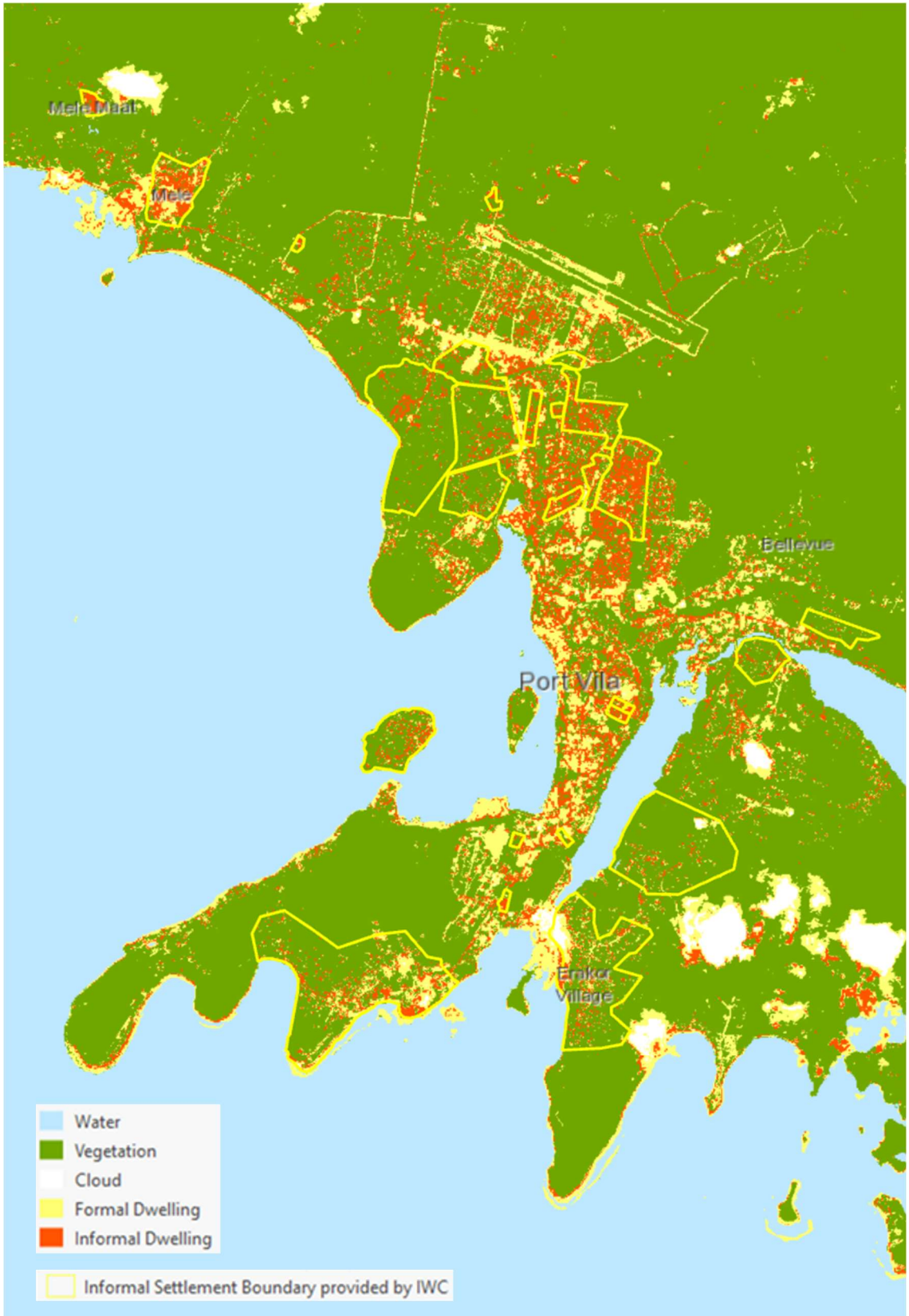

FIGURE A1 – PIXEL CLASSIFICATION MODEL OUTPUT IDENTIFYING INFORMAL DWELLING, FORMAL DWELLING, VEGETATION, WATER AND CLOUD CLASSES IN PORT VILA, VANUATU.

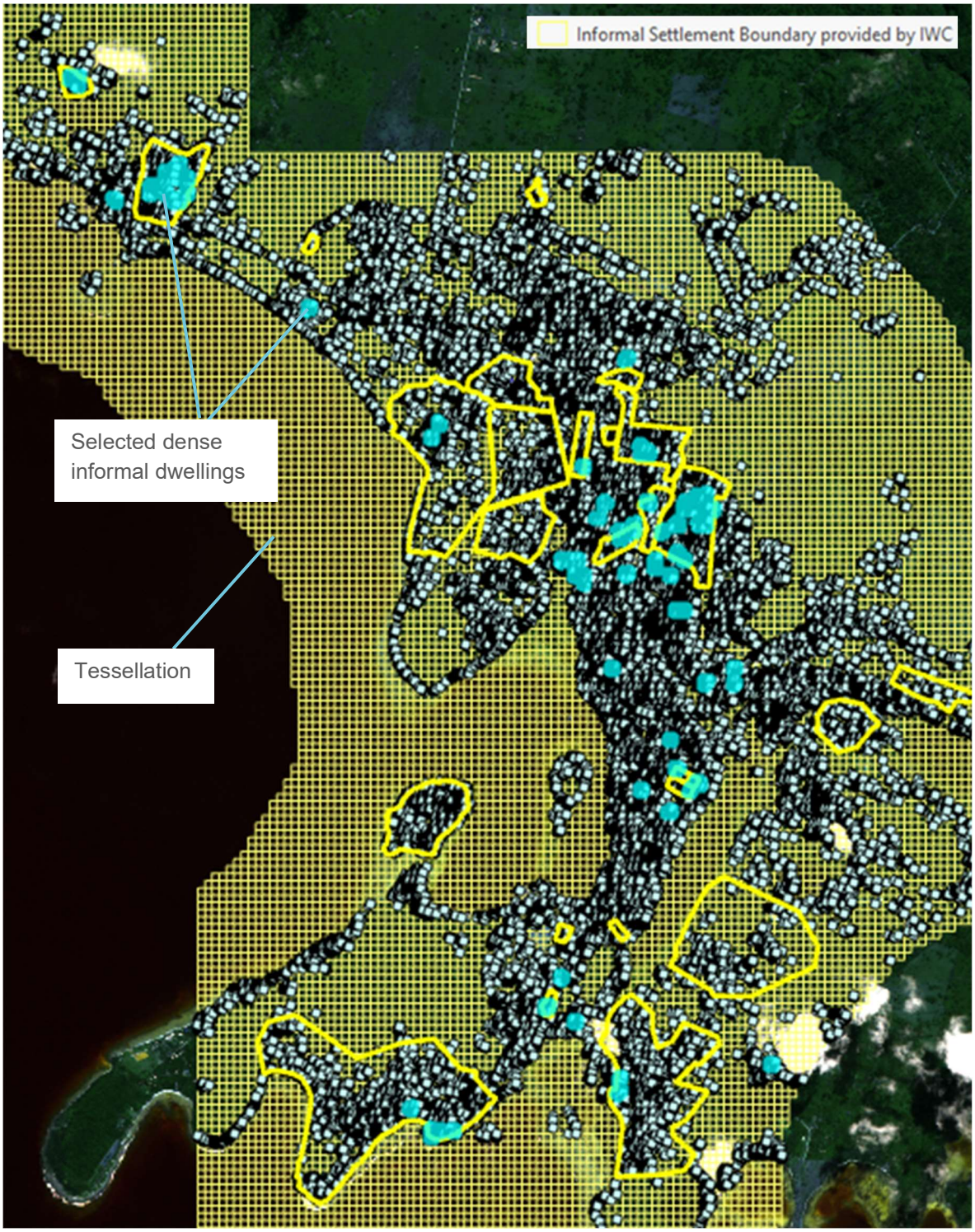

FIGURE A2 – PIXEL CLASSIFICATION TRAINING DATASET INCLUDING INFORMAL SETTLEMENT BOUNDARIES AND TESSELLATION GRID FOR MODELS OF PORT VILA, VANUATU.

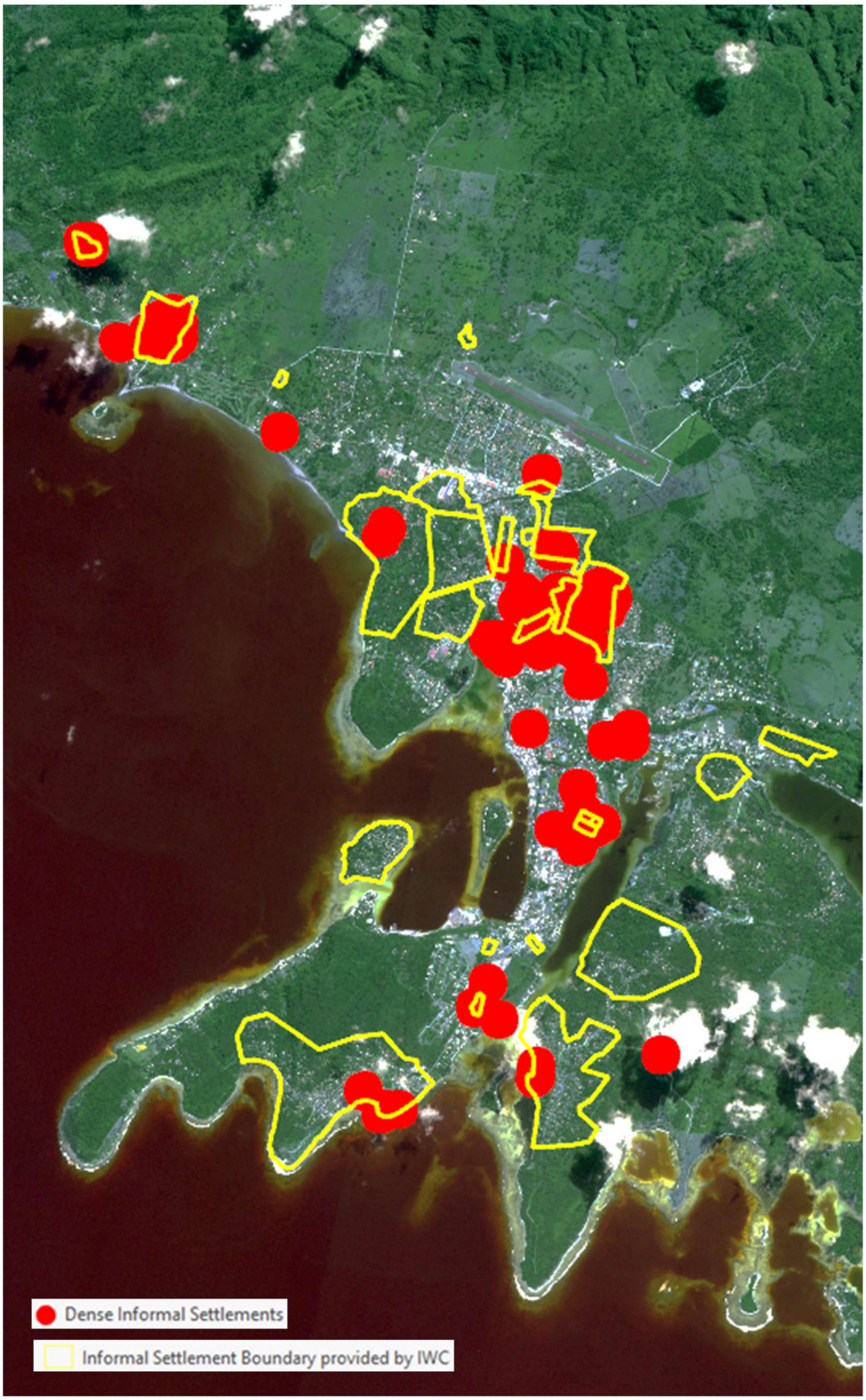

FIGURE A3 – PIXEL CLASSIFICATION MODEL TRAINING DATASET INCLUDING INFORMAL SETTLEMENT BOUNDARIES AND MODEL OUTPUT IDENTIFYING DENSE INFORMAL SETTLEMENTS IN PORT VILA, VANUATU (SATELLITE IMAGERY BACKGROUND).

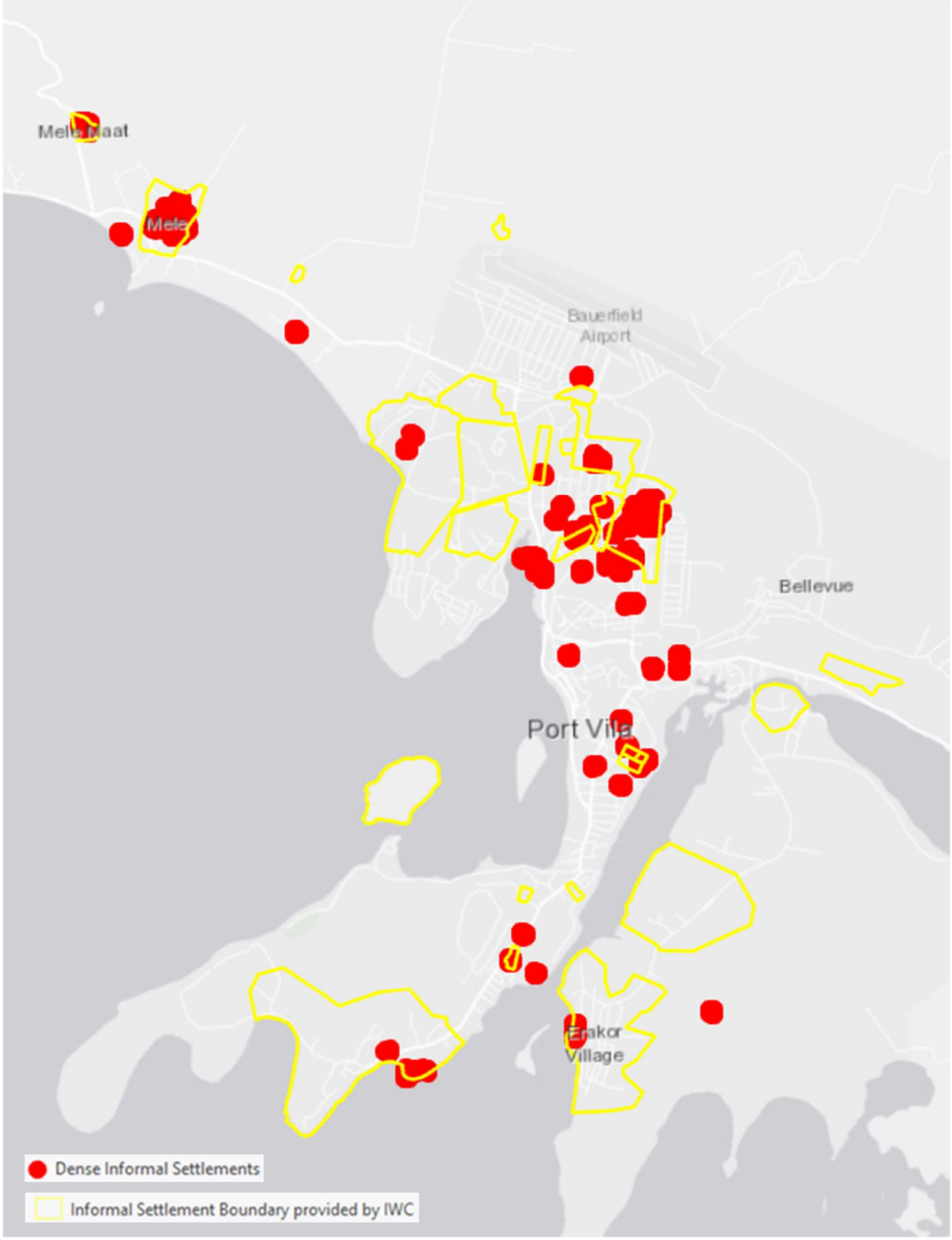

FIGURE A4 – PIXEL CLASSIFICATION MODEL TRAINING DATASET INCLUDING INFORMAL SETTLEMENT BOUNDARIES AND MODEL OUTPUT IDENTIFYING DENSE INFORMAL SETTLEMENTS IN PORT VILA, VANUATU (OPEN STREET MAP BACKGROUND).

Annexure B – Western Fiji Model Output

FIGURE B1 – PIXEL CLASSIFICATION MODEL OUTPUT IDENTIFYING INFORMAL DWELLING, FORMAL DWELLING, VEGETATION, WATER AND CLOUD CLASSES IN WESTERN FIJI.