

# Rethinking Predictive Analytics for Disaster Resource Allocation

Integrating Vulnerability and Sustained Impact into Risk Modelling

Hamish Patten, University of Oxford

#### ABSTRACT

Informed disaster management requires detailed knowledge of the affected environment. Predictive analytics can help to provide such insight. Both the Integrated Internal Displacement Population Sampler (IIDIPUS) statistical engine, and the Oxford Disaster Displacement Real-time Information Network (ODDRIN) interactive data visualisation software, have been developed through a collaboration between the Department of Statistics, University of Oxford, and the Internal Displacement Monitoring Centre (IDMC). This paper demonstrates the capacity of the IIDIPUS code to inform disaster resource allocation on the short to mid-term. The primary aim of IIDIPUS is to estimate human displacement, not damaged assets. This shift in focus increases the expected accuracy in the prediction of the spatial distribution of displacement, highlighting displacement hotspots. Sub-national and national vulnerability, sequential event modelling, and satellite image-based building damage assessment data are integrated into IIDIPUS. The IIDIPUS code predictions were more than 10 times more accurate in predicting the number of displaced people than two world renowned risk models: the Global Disaster Alert and Coordination System alert score and the United States Geological Survey PAGER risk score, over 101 historical earthquakes in 37 different countries. Mid to long term temporal displacement predictions utilise mobile phone data-based displacement information, with the potential for disaggregation by sex and age. Finally, coupling IIDIPUS with OpenRouteService facilitates the optimisation of the location and capacity of emergency shelters.

## **INTRODUCTION**

Predictive analytics are beginning to play a crucial role in disaster management. With economic budgets of limited flexibility, questions arise such as whether to allocate thousands or millions of dollars in humanitarian aid after rapid-onset natural hazards such as tropical cyclones or earthquakes. Spatial distribution of humanitarian aid is prioritised based on equity but requires intricate knowledge of worst-affected regions. Expert opinion must always guide management decisions and resource allocation. The knowledge and understanding that guides expert decisions come from a broad comprehension of historical events: to learn from the past. However, this seems counter-intuitive, as '*No two disasters are alike*'.<sup>1</sup>

An alternative way of stating this would be to say that natural hazards are inherently stochastic processes: a roll of the dice. Two similar cities with the same population can be struck by the exact same hazard intensity, yet the number of displaced persons can be more than ten times higher in one city than the other. The role of quantitative methods and risk modelling is to separate causal factors that result in disaster vulnerability (e.g. infrastructure quality or income disparity) from plain and simple bad luck. Exploring such questions allows for extrapolation in case of a future event, such as to assess the damage risk for higher hazard intensities or lower quality building infrastructure. This helps to guide disaster management, government policy and overall preparedness to learn from the mistakes/successes of the past. Informing such factors can also significantly reduce the total resources required for the next event occurrence. It is impossible to remove all uncertainty in resource allocation: the larger the population exposed to the hazard, the larger the expected error in the predictions. However, uncertainty can be significantly reduced through combining informative data (e.g. population maps) with well-suited risk models.

<sup>&</sup>lt;sup>1</sup> UNISDR (now UNDRR), "Global Assessment Report on Disaster Risk Reduction 2013". 2013. <u>https://www.undrr.org/publication/global-assessment-report-disaster-risk-reduction-2013</u>



## Open-Access Data

Recent years have witnessed a boom in the availability of open-access data. Creating accurate and effective risk models requires a minimum set of ingredients, all of which are now openly available in various forms and flavours. With relative ease, it is possible to access high-granularity gridded datasets that cover most countries in the world, providing estimates of population count and density, and Gross Domestic Product at Purchasing Power Parity (GDP-PPP) information disaggregated on a sub-national level. National-level indicators such as unemployment rates and income distributions can be retrieved through user-friendly access methods. Real-time, high quality hazard intensity mapping permits the visualisation of current and historical natural hazards in great detail. Furthermore, organisations such as the Internal Displacement Monitoring Centre (IDMC) hold decades-worth of records of the number of people displaced from the onset of tens of thousands of different events. The availability of such high quality and detailed databases greatly facilitates and even encourages organisations, academic institutions and the general public at large to contribute their own formulations of disaster risk.

#### **Research Contribution**

The research presented in this paper represents a collaboration between the University of Oxford and IDMC, made possible through the Engineering and Physical Sciences Research Council Impact Acceleration Account (EPSRC-IAA) grant. The aim is to question some traditional concepts of disaster risk modelling, address common issues and provide solutions. A principle focus is to provide an incentive to deviate away from calculating the total (financial or asset-based) cost of an event, towards first addressing the population demographic predicted to be displaced, then making the economic calculation. Financial and humanitarian resource allocation is often based on exposed financial assets or the number of exposed buildings, and can often fail to address the short and long term costs of displaced communities. Centering the question on estimating the displaced population instead of destroyed assets, this reformulation could decrease common long-term outcomes of disasters such as forced migration or even gentrification.

This article is structured in three main sections. The first section provides a definition of disaster risk modelling, highlighting some common issues and describing the solutions proposed in this research, then describes the IIDIPUS code model, data implementation and methodology. The second section presents the results, firstly through the use of global risk models for earthquakes, then through local-level disaster risk models, with Cyclone Harold as an example. Two main examples of the potential for future predictive analytics with IIDIPUS are also given in the second section, via the use of integrated information systems. This includes the use of mobile phone data and with open-source mapping software. The final section provides a conclusion and future outlook.

## DISASTER RISK MODELLING

Risk modelling for natural hazards tends to separate the required information for predictions into three categories: 'hazard', 'exposure' and 'vulnerability'. Note that sometimes a fourth category (here included in vulnerability) is 'coping capacity'. Hazard refers to the hazard intensity, such as the surface wind speed or flood level experienced at a specific location. Exposure refers to exposed elements, such as the number of people or the number of buildings or financial assets exposed to a non-negligible hazard intensity. Vulnerability is a very broad term, but assesses a causal susceptibility to disaster risk, and is usually based on both individual and collective effects. Examples of which are low individual income and poor government preparedness. Vulnerability is the most difficult of the three categories to quantify. The number of damaged elements (e.g. buildings or population) is predicted by combining these three categories via a risk model. Resource allocation is not directly calculated in this research, but can be estimated via the use of detailed micro and macroeconomic information, such as the cost of resources (shelter, food, medical supplies), transportation and resource distribution.



# Vulnerability

The more granular and high-quality the vulnerability information available, the more accurate the prediction. Local-level vulnerability can come in various forms. For example, in the Philippines, the Philippine Statistics Authority (PSA) regularly (every 2-5 years) publish detailed, census-based demographic information disaggregated by region<sup>2</sup>. Population and housing-related demographic information is disaggregated per city/municipality and building construction type, providing information such as the number of households and the household population size. Such high-quality vulnerability information can directly lead to improved disaster management and resource allocation. The difficulty with incorporating local vulnerability for globally predictive disaster risk models is that the information between countries must correspond directly with one another. If the risk model is built to require stratified housing quality information per region, every additional country or area calculated will also require this information. Furthermore, merging datasets - called 'data harmonisation' - is not an optimal solution, either: the definition of a 'strong' and 'weak' building varies significantly between countries. In this research, a focus is made on two local vulnerabilities: income and population density. These two variables were chosen due to the open-availability of globally harmonised datasets. Including local income and population density in the risk model provides insight into the etiology of socio-economic and rural-urban differences in disaster risk, respectively.

National vulnerability in the context of disaster risk is a vast and, so far, inconclusive topic. A meta-analysis study by Beccari, 2016, reviewed 106 methodologies that had the common aim of developing a disaster vulnerability index<sup>3</sup>. A disaster vulnerability index is a compound variable, which means that it is comprised of a specific combination of other variables, such as the unemployment rate or the percentage population aged over 65. In the 106 different methodologies aforementioned, 2298 different variables were used to form the vulnerability indices. This reflects that there is clearly no common consensus on which are the variables that correlate most with disaster vulnerability. The purpose of the research presented here is to demonstrate a proof-of-principle for integrated vulnerability. Therefore, as the combination of national vulnerability indicators is a poorly-resolved domain, a few common and intuitively-relevant variables were chosen for this study: physical infrastructure quality, disaster risk reduction index, government effectiveness, aid dependence, and poverty and development. Furthermore, as the global displacement model is based on past earthquake events, earthquake exposure was also included.

## Sustained Hazard Impact

Sustained impact from a hazard is something that is often neglected from risk modelling. Sustained impact is the accumulation of damage due to multiple impacts, not just considering the maximum hazard impact intensity over the entire event duration. For example, earthquake damage risk models tend to take the 'worst' earthquake, defined by differing criteria such as the largest maximum intensity, or the largest exposed population to a non-negligible intensity. The issue is to assume that pre/aftershocks are negligible, which is often not the case<sup>4</sup>. To provide another example, tropical cyclone damage risk models often take the global maximum wind speed of the cyclone path<sup>5</sup>. The accumulation of damage from experiencing hours or days of extreme wind speeds can be significantly more severe than being exposed to the maximum

https://doi.org/10.1371/currents.dis.453df025e34b682e9737f95070f9b970.

<sup>&</sup>lt;sup>2</sup> PSA, 'Housing Characteristics in the Philippines (Results of the 2015 Census of Population)', 2018. https://psa.gov.ph/population-and-housing

<sup>&</sup>lt;sup>3</sup> Beccari B. 'A Comparative Analysis of Disaster Risk, Vulnerability and Resilience Composite Indicators'. PLOS Currents Disasters. 9(1), 2016.

<sup>&</sup>lt;sup>4</sup> Hough, S. E., and Jones, L. M. '*Aftershocks: Are they earthquakes or afterthoughts?*' EOS, Transactions American Geophysical Union 78(45), 1997, p 505-508.

<sup>&</sup>lt;sup>5</sup> Aznar-Siguan, G., and Bresch, D.N. '*CLIMADA v1: a global weather and climate risk assessment platform.*' Geoscientific Model Development 12(7), 2019, p 3085-3097.



wind speed once. By including snapshots of the hourly maximum surface wind speed, over days or even weeks, this research aims to produce a more realistic and accurate risk model.

# Integrated Internal Displacement Population Sampler (IIDIPUS)

The main aim of the collaboration between the University of Oxford and IDMC was to build a software tool to predict disaster displacement risk, using state-of-the-art computational statistical algorithms and mathematical models. Many of the unique capabilities and reformulations of the IIDIPUS code is through the specific data used. The model is built by layering different datasets onto one another, providing insight into the hazard, exposure and vulnerability of the event. Figure 1 provides an illustration of the layering of the important datasets.



Figure 1: data layering in the IIDIPUS code. From the top layer downwards: satellite image-based building damage assessment data, hazard intensity mapping, census-based sub-national GDP-PPP per capita information, population count based on census-data and satellite-luminosity, and the location information such as region/country names and boundary shapes as well as elevation.

As mentioned previously, IIDIPUS focusses on the human aspect of disaster risk instead of the economic cost or damaged assets, by using population count data to predict the displaced population. The gridded population count data used is the UN WPP-Adjusted Population Count, taken from the SocioEconomic Data and Applications Center (SEDAC) Center for International Earth Science Information Network (CIESIN)<sup>6</sup>. To accurately model the spatial dynamics, as well as the variation in building damage (the stochastic component of the damage likelihood calculation), satellite image-based building data assessment data is implemented. The building damage data is extracted from both the United Nations Institute for Training and Development - Operational Satellite Applications Programme (UNITAR-UNOSAT)<sup>7</sup> and Copernicus<sup>8</sup>. The inclusion of satellite image-based building damage data is a novelty of the IIDIPUS code as compared to

<sup>&</sup>lt;sup>6</sup>Center for International Earth Science Information Network - CIESIN - Columbia University. 2018. '*Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11*'. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PN93PB.

<sup>&</sup>lt;sup>7</sup> Maps and Data – UNITAR. <u>http://www.unitar.org/unosat/maps</u>

<sup>&</sup>lt;sup>8</sup> Aifantopoulou, D. '*Copernicus emergency management Service*.' 10442/15383 (2016): 00-54. https://emergency.copernicus.eu/



alternative disaster risk models. Sequential event modelling requires high spatial resolution data that covers the entire period of the hazard impact. For earthquakes, the United States Geological Survey (USGS) high resolution shakemap intensity raster data<sup>9</sup> is accessed directly via the use of an API. For tropical cyclones, the hourly weather data uses the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) measurements, made and processed by the National Aeronautics Space Agency Goddard Earth Sciences Data and Information Services Center (NASA-GES-DISC)<sup>10</sup>. The national-level vulnerability parameters were extracted from the World Bank, the Joint Research Centre – European Commission (JRC-EC) and the World Inequality Database (WID). Sub-national vulnerability indicators were extracted from various sources. The GDP-PPP per capita gridded dataset was taken from Kummu, et al<sup>11</sup> to produce the sub-national socio-economic vulnerability. The socio-economic sub-national vulnerability indicator is calculated by combining the national income distribution (from the WID) with the local GDP value. If a country's population is ordered in terms of income, then the 50<sup>th</sup> percentile (the median income) of a population of 20 million would be the income of the 10 millionth person. The 10<sup>th</sup>, 20<sup>th</sup>, 30<sup>th</sup>, 40<sup>th</sup>, 50<sup>th</sup>, 60<sup>th</sup>, 70<sup>th</sup>, 80<sup>th</sup>, and 90<sup>th</sup> income percentiles were chosen to allow an accurate reflection of income distribution, without being strongly biased by the super-rich or super-poor. Before combining the income distribution with the GDP data, the income distribution is modified such that the different percentiles are divided by the median (50%) value. When combined with the local GDP-PPP per capita value, this can be interpreted as the local 'income', whereby the middle income is equal to the local GDP-PPP per capita, and differing percentiles become relative to the middle income. For example, at a location with a GDP-PPP per capita of \$10,000, the 50<sup>th</sup> percentile is assumed to earn \$10,000. If the 80<sup>th</sup> percentile earns two times more than the 50<sup>th</sup> percentile, then this population is assumed to earn \$20,000. In this way, it is possible to construct a sub-national vulnerability indicator that correlates with socio-economic factors, including income disparity. For the rural vs urban sub-national indicator, the aforementioned UN-adjusted population count was used. For short to mid-term spatially temporal displacement inference, mobile phone data can be used by the IIDIPUS code. Currently, no specific organisation/companies mobile phone-based displacement inference data has been directly integrated into IIDIPUS. However, the code has been adapted to easily implement data from organisations such as the Facebook Interactive Disaster Displacement Maps<sup>12</sup> and Flowminder<sup>13</sup>. Emergency shelter optimisation is calculated using OpenStreetMaps (OSM)<sup>14</sup> and the OpenRouteService (ORS)<sup>15</sup> open licensed products.

Mathematical models such as disaster risk models require very specific tuning of certain parameters in order to produce the most accurate predictions. Finding the model parameters that best correspond to the observed data is the role of computational statistical algorithms. The mathematical model presented in this paper uses a Bayesian framework, which is a specific formulation that allows expert knowledge to be included to guide the model tuning, and emphasises evaluating model uncertainty. Markov Chain Monte Carlo (MCMC) algorithms are implemented to find optimal solutions for the model parameterisation. The

<sup>&</sup>lt;sup>9</sup> Wald, D. J., *et al, 'ShakeMap manual: technical manual, user's guide, and software guide'*, USGS Publications Warehouse, 12-A1, (2005) <u>http://pubs.er.usgs.gov/publication/tm12A1</u>

<sup>&</sup>lt;sup>10</sup> Rienecker, M. M., *et al*, *'MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications', Journal of Climate, 24*(14), 3624-3648. (2011) https://journals.ametsoc.org/view/journals/clim/24/14/jcli-d-11-00015.1.xml

<sup>&</sup>lt;sup>11</sup> Kummu, M., et al. 'Gridded global datasets for gross domestic product and Human Development Index over 1990–2015.' Scientific data 5(1) (2018) <u>https://www.nature.com/articles/sdata20184</u>

<sup>&</sup>lt;sup>12</sup> Maas, Paige, *et al. 'Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery.' KDD*. Vol. 19. (2019)

<sup>&</sup>lt;sup>13</sup> Deville, Pierre, et al. 'Dynamic population mapping using mobile phone data.' Proceedings of the National Academy of Sciences 111.45 (2014)

<sup>&</sup>lt;sup>14</sup> OpenStreetMap contributors. <u>https://www.openstreetmap.org</u>

<sup>&</sup>lt;sup>15</sup> OpenRouteService contributors. <u>https://openrouteservice.org</u>



use of MCMC methods allows the user to access a detailed understanding of not just the most likely parameters, but the uncertainty in each, and even outputs the correlation between different parameters. For example, the strength of the influence of the physical infrastructure quality index on disaster risk is output as a plausible range of values, indicating the most likely value. Furthermore, the strength of the correlation between physical infrastructure quality and a socio-economic indicator can also be understood directly.

# RESULTS

In scientific research, the peer-review process is a crucial element to ensure high quality and accurate scientific publications. Before the peer-review process has been completed and a scientific article published, all results are preliminary. Therefore, the results published in this article are preliminary. The results are presented in four subsections. The first two subsections present the preliminary results from the IIDIPUS software, which illustrate the predictive performance on a global and local level, respectively. The third subsection describes the use of mobile phone data to infer spatially disaggregated temporal population displacement. Finally, the fourth subsection explores the capacity for emergency shelter optimisation.

## Global Displacement Model

A global risk model is defined as a model that is capable of making predictions for natural hazards at a global scale, and even compare events between countries through causal relations. When building such intricate global risk models from scratch, it is important to start off by addressing the most simple hazards, especially those where the required data is easily available. Storms and cyclones are inherently multi-hazard: rainfall, storm surge and surface wind speed all play a role. Additionally, floods can last for months, and such long-term effects are exceptionally difficult to accurately model. Producing a global displacement model for such intricate multi-hazard models requires a large amount of data and ensuring exceptionally high data quality. For simplicity, in the global displacement model presented here, a focus is made on earthquakes.

## Earthquakes

In the 101 earthquakes studied for this research, the total people displaced over all events was 9.8 million, ranging from 124 people in the Laos earthquake of November 2019 to 2.6 million people in the Nepal earthquake of April 2015. Earthquakes studied occurred in 37 different countries. The majority of the earthquakes occurred in Asia (69), with 17 in the Americas, 9 in Europe, 4 in Africa and 2 in Oceania. The earthquakes explored in this study dated between 2008 up to 2020.

The predictive power of the IIDIPUS code is calculated by comparing with two world leading hazard alert systems: the United States Geological Survey (USGS) PAGER score<sup>16</sup> and the Global Disaster Alert and Coordination System (GDACS) alert score<sup>17</sup>. The risk score predictions of USGS-PAGER are based on both the financial cost of the earthquake and the number of fatalities, and the GDACS alert score predictions are based on the number of fatalities. The USGS-PAGER alert is a four-category alert system: green, yellow, orange and red. The GDACS alert score is continuous, ranging from 0 for negligible risk to above 2 which is deemed a red alert. Comparison with IIDIPUS is done by converting the two alert scores into predictions for the number of people displaced, using Generalised Linear Models (GLM).

<sup>&</sup>lt;sup>16</sup> Jaiswal, K. S., et al. "*Earthquake casualty models within the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system*." Human casualties in earthquakes. Springer, Dordrecht, 2011. 83-94. <u>https://earthquake.usgs.gov/data/pager/background.php</u>

<sup>&</sup>lt;sup>17</sup> De Groeve, T., Vernaccini, L., and Annunziato, A. "*Modelling disaster impact for the global disaster alert and coordination system.*" Proceedings of the 3rd International ISCRAM Conference. Newark, NJ, USA, 2006. <u>https://www.gdacs.org/Knowledge/models\_EQ.aspx</u>



The definition of an accurate global model used for this research is one that reduces the absolute difference between the observed and predicted displaced population, relative to the observed estimate. Making the difference relative to the observed estimate acts to convert the error to a fraction or percentage of the total observed displacement. This is important when the displaced estimates range from the hundreds to the millions. A one percent error in the Nepal earthquake would lead to 26,000 people incorrectly predicted as displaced, but for Laos this would be only one person. If the relative value was not taken (diving by the observed value for each event), then small errors in the earthquakes which had a large displacement, such as for Nepal, would dominate the errors in the smaller-scale events. The average percentage error over all earthquakes was 426% for the IIDIPUS code, 4,894% for the GDACS alert score and 6,545% for USGS-PAGER. The median percentage error indicates that fifty percent of the events were within 95% of the observed value for IIDIPUS, 612% for the GDACS alert score and 749% for USGS-PAGER. These results clearly indicate that for the 101 earthquakes studied, the IIDIPUS code is significantly more accurate a predictor of the number of people displaced by an earthquake than both USGS-PAGER and the GDACS alert score. This improvement is expected to increase further with future modifications to the model, and by using machine learning algorithms to recognise correlation between a broad range of national vulnerability parameters and the displaced populations.

Out of the choice of the six national vulnerability parameters, the strongest correlated indicators with a damage/displacement vulnerability were the physical infrastructure quality index and the poverty and development index. This is an intuitive result: stronger house and road construction quality, low dependency on humanitarian aid, and high overall country development all lead to lower vulnerability. For the sub-national vulnerability parameters, the increase/decrease in the damage and disaster risk depends on the hazard intensity and local characteristics such as the local GDP-PPP per capita. Preliminary results show that for a person earning the median income in Nepal (8,620 USD-2021) as compared to in Italy (52,359 USD-2021), the risk of displacement is at least 1.6 times higher for an earthquake intensity of 5 MMI. If the earthquake intensity is increased to 7 MMI this factor drops to around 1.4, and the risk for both countries becomes equal at intensities above 9 MMI. A further benefit of using the income distribution instead of the GNI is that risk disparity can also be inferred. In Nepal, the poorest 10% are at least 2.7 times more at risk of being displaced than the richest 10% for a hazard intensity of 5 MMI. In Italy, the poorest 10% are at least 3.3 times more at risk of being displaced than the richest 10% for a hazard intensity of 5 MMI. As income inequality between the poor and the rich is larger in Italy than for Nepal, the relative displacement risk also increases.

## Local Displacement Model

The most reliable risk models are built at a local level. More descriptive and detailed population or income census data, or building quality information can significantly improve the predictions. However, this requires having ample amounts of accurate and detailed information, on at least one historical event. For IIDIPUS, generating local models requires, at the very least, building damage assessment data that covers a broad spatial extent, in order to accurately characterise the risk of damage and thus displacement. In this subsection, a focus is made on a climate-related hazard: Cyclone Harold, that passed over the Pacific islands in early April 2020. Cyclones, storms and floods are expected to increase in frequency and severity due to climate change. This means that developing tools to provide real-time, spatially distributed displacement estimates is expected to become increasingly important in the future.

April 2020 Cyclone Harold over Vanuatu





Figure 2: surface plot of the hourly maximum surface wind speed, in metres per second, around 7am on the 6th April, 2020 (EST). Data used for the plot was taken from the NASA - Goddard Earth Sciences Data and Information Services Center (GES DISC). Underneath the surface plot is a map of the Pacific islands, indicating that at this time, the cyclone passed directly over Vanuatu. Map plot taken from OpenStreetMaps.

Between the 3<sup>rd</sup> to 8<sup>th</sup> April, 2020, Cyclone Harold passed over Vanuatu. Figure 2 shows the maximum hourly wind speed around the time when the eye of the cyclone passed over the islands of Vanuatu, indicating surface wind speeds of over 40 metres per second (144kmph). ACAPS predicted that around 160,000, half the population of Vanuatu, were affected by the cyclone. The International Organization for Migration Displacement Tracking Matrix (IOM-DTM) estimated that the Internally Displaced Persons (IDP) stock on the 15<sup>th</sup> April 2020 was 18,538 on Vanuatu alone. This estimate was verified and input into the IDMC database. In the aftermath, UNOSAT assessed 12,494 buildings for damage using satellite imagery. 516 buildings were completely destroyed, 100 with moderate damage, 225 with possible damage and 11,653 described generally as having some form of damage. Figure 3 (right) shows a plot of the location of the damaged buildings.

To predict the displaced population, the IIDIPUS code used over 140-time stamps of the hourly maximum wind speed passing over Vanuatu, calculating the damage risk from the sustained impact, not the overall maximum intensity. For Vanuatu, the income disparity-related disaster vulnerability estimated that the poorest 10% were twice at risk of displacement than the richest 10% at 25m/s (90kmph) wind speeds and 1.5 times more at risk for 35m/s (126kmph) wind speeds. The spatial distribution of the displaced population predicted by IIDIPUS is shown in figure 3 (left). Note that the UNOSAT building damage and the IIDIPUS predictions are not meant to correlate directly with one another, as only specific regions tend to be assessed for damage by UNOSAT, not the entire area. Therefore, the IIDIPUS predictions can help to find hotspots where large displaced populations are foreseen, but no building damage assessment was made by UNOSAT, such as the island of Lambukuti. The largest displacements predicted were located in the region of Luganville, the second largest city in Vanuatu. Future improvements to the predictions made for Cyclone Harold include using the higher resolution population data provided by Facebook Data for Good<sup>18</sup>.

<sup>&</sup>lt;sup>18</sup> Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe.



Additionally, higher resolution information about the GDP-PPP information for each island, or more detailed information on the building quality per island would result in more precise spatial estimates.



Figure 3: (left) predictions of the displaced population made by IIDIPUS, noting that the colourmap is on a log-scale, (right) the UNOSAT building damage assessment data. Overlay onto the Vanuatu island map taken from OpenStreetMaps.

#### Mobile Phone Data Integration

Predictive analytics can be used to calculate the rate-of-return of displaced populations. However, producing accurate estimates is extremely difficult, due to non-linear dynamics such as government and humanitarian aid intervention, the government disaster policy and preparedness, and other effects such as social capital. Therefore, in this research, no attempt is made to predict the rate-of-return based on causal models, but to rely on available temporal data to infer the rate-of-return. One of the most accurate temporal displacement measurements is via inference from mobile phone data. It is even possible to disaggregate the population by sex or age. In order to provide anonymity to the general population, information on the displaced populations must be spatially aggregated by region or area. Figure 4 shows the combination of IIDIPUS displaced population estimates with an example of mobile phone data-based displacement information. The example mobile phone data was generated to realistically represent how such information would be output by organisations such as Facebook Data for Good or Flowminder. In this example, six of the islands of Vanuatu were chosen to aggregate the displaced population information. The significant importance of this capacity of the IIDIPUS code is clear: gender-differentiated humanitarian aid can be specifically targeted based on temporal estimates of the displaced populations. Furthermore, statistical methods can further be applied to mobile phone data to predict long-term trends of the rate-ofreturn, especially when combined with historical data.







Figure 4: an example of combining mobile phone data-based population displacement information with the IIDIPUS code. The upper two plots -(a) and (b) - are 40 days after the hazard occurrence, the lower two -(c) and (d) - are directly after. Sex disaggregation of the mobile phone data separates the female displaced population, shown in the left plots (a) and (c), from the male displaced population, shown in the right plots (b) and (d).

#### **Emergency Shelter Optimisation**

The time and distance it would take to travel between certain displaced population locations and a given shelter are calculated by use of the ORS. By combining IIDIPUS with the ORS and the locations and capacities of emergency shelters, it is possible to optimise emergency shelters for a specific event. The spatial distribution of the displaced population output by IIDIPUS can be used to understand whether emergency shelters local to specific displacement hotspots are sufficient in capacity. Such evaluations can highlight whether additional shelters should be installed in certain areas, and what shelter occupation might be expected from the locally displaced population.

#### CONCLUSION

This paper presents the recent successes of the collaboration between the University of Oxford and the Internal Displacement Monitoring Centre on disaster-induced displacement risk modelling. The software tool consists of two components: IIDIPUS, the statistical engine, and ODDRIN, for interactive data visualisation. The research presented in this article has been carried out with the primary purpose of informing disaster management and humanitarian aid allocation. The main novelties of the software tool are that it places emphasis on predicting displaced populations over damaged assets, includes local and national disaster vulnerability in the calculations, uses national income distributions and not the mean income, integrates satellite image-based building damage assessment data, and integrates sustained hazard impact instead of the maximum hazard impact. The IIDIPUS tool is therefore purpose built to try to further integrate the human aspect into disaster risk modelling. This can be used to provide insight into the influence of income disparity on risk, in addition to identifying the role of local and national-level vulnerability on the risk of displacement. Two main models are presented: global and local disaster risk models. Using the global risk model, predictions on the relative number of people displaced over 101 different earthquakes showed that IIDIPUS is, on average, more than 10 times more accurate than world leading organisations such as the USGS-PAGER and GDACS alert scores. By using specific information related to the country of Vanuatu, results for a local risk model were given for Cyclone Harold of April 2020. Developing local models can help provide more accurate and detailed predictions of displacement in the event of future events. By focussing on a climate-related event, it was illustrated that local models are a strong candidate for accurate disaster risk modelling in a future where climate change may induce more



frequent and severe weather-related events. A priority for future work should be to extend the model to include alternative natural hazards, such as storms and floods, and to calculate global disaster risk models for each. Furthermore, for tropical cyclones and storms, multi-hazard risk modelling should include all the necessary hazard components such as storm surges and flooding. Increasing the resolution in the population count and GDP-PPP per capita data should also be a high priority.