

Recurrent Flooding and Household Food Access in

Central Java, Indonesia

by Breanne Kristen Langlois, M.P.H., Ph.D. Candidate

A dissertation submitted to Tufts University

Friedman School of Nutrition Science & Policy

In conformity with the requirements for the degree of Doctor of Philosophy in the
Nutrition Epidemiology and Data Science Division

Committee:

Elena N. Naumova, Ph.D. (chair)

Timothy Griffin, Ph.D.

Erin Coughlan De Perez, Ph.D.

Magaly Koch, Ph.D.

i. Summary

Flooding affects food availability globally through direct losses to agricultural crops but the impact on diet-related aspects of food security at the household level are not well-understood. Central Java faces multiple types of flooding – coastal, flash, and river flooding – caused by heavy rainfall, tidal overflow, sea level rise, and land subsidence. Many households stay and adapt, but it is unclear how their adaptive capacity relates to diet and food access. Current evidence in this context is mostly focused on food production. Further, most research focuses on single flood events and the short-term impacts rather than flood recurrence and longer-term impacts over time. One reason is the lack of datasets designed for studying environmental impacts on household outcomes. Recent advances in flood detection using remote sensing have led to many data products, presenting opportunities for researchers to create multidisciplinary datasets. But merging geospatial and population data is challenging due to incompatible formats.

In this study, I compared the usability of 5 flood monitoring data portals to create a multidisciplinary dataset (Aim 1) and evaluated the relationship between recurring flooding and household diet quantity (Aim 2) and quality (Aim 3). I examined differential effects between recurring flooding and measures of adaptive capacity over 1984-2015. Usability aspects included data accessibility and availability, approaches to flood identification, data quality, and user feasibility. I identified an efficient method of merging raster data (geospatial data in gridded, imagery format) to a household survey using R programming and linked the Global Surface Water ‘recurrence’ layer (raster format) to the Indonesian Family Life Survey – a longitudinal health and economic survey from 1993 to 2015. I defined recurrent flooding as the fraction of district raster cells with episodic flooding over the period. I constructed household food expenditure share and dietary diversity score to measure diet quantity and quality based on the Data4Diets framework. I applied statistical and machine learning methods to model the

data – generalized linear mixed models and Random Forest regression models – and examined both linear and nonlinear relationships.

Statistical models detected joint effects with flooding and adaptive capacity. Wealth was associated with a decrease in food expenditure share, and this effect was strengthened in more flood affected districts. Access to credit was associated with higher odds of a more diverse diet, but this effect changed to reduced odds in more flood affected districts. Flooding had more predictive importance for food expenditure share than for diet diversity in Random Forest models. Other exploration showed that households in districts with the highest portion of flooding over the period had an increase in food expenditure share.

This study provides important methods for creating linked, multidisciplinary datasets that other researchers can replicate. I explored how recurrent flooding can be measured and found that access to credit, a factor that ordinarily improves food access, may not be effective in flood-prone areas for already vulnerable households. With the growing climate change concerns, wealthier households may be better able to adapt in terms of food access.

ii. Acknowledgements

This work could not have been completed without the support of my advisors, peers, friends, and family.

My advisor, thesis committee chair, mentor, life coach, and friend, Dr. Elena N. Naumova, has been my toughest critic and strongest advocate. Under her guidance, I have grown tremendously on both an academic and personal level. She takes her role as mentor seriously and teaches her students to believe in themselves while sharing her insight and wisdom. She is the greatest teacher I have had and one of the most thoughtful and compassionate people I have met. The originality and quality of this work is due to her mentorship and guidance.

The other members of my thesis committee – Dr. Timothy Griffin, Dr. Erin Coughlan de Perez, and Dr. Magaly Koch – have all shown incredible guidance throughout this process. They have all helped to shape this dissertation along the way, while sharing their different perspectives. I cannot think of a better group of individuals for this committee, and I am grateful to each of them.

This work would not have been possible without the collaboration from colleagues at Diponegoro University in Semarang, Indonesia. I am grateful to Dr. Aris Ismanto, Dr. Ambariyanto, Dr. Laila Kholid Alfirdaus, Dr. Muhammad Helmi, and Dr. Retno Murwani for their consistent feedback on the material in this dissertation.

Dr. Beatrice Rogers, Dr. Patrick Webb, Dr. Irwin Rosenberg, Dr. Kenneth Kwan Ho Chui and the entire USAID Food Aid Quality Review Team have provided me with invaluable mentorship during my time at Tufts. I have learned a lot from each of them. Thank you especially to Beatrice Rogers for her enthusiasm and mentorship. I thank Kenneth Kwan Ho Chui, whom I have learned a lot from during my time at Tufts, for his guidance and kindness.

The aims were inspired by the International Dietary Data Expansion project led by Dr. Jennifer Coates. I thank Jennifer Coates for her encouragement and support.

My peers, friends, and other students from Tufts have provided support throughout this process. Thank you to Ryan Simpson, Alex Carroll, Kyle Monahan, Elizabeth Marsh, Dolapo Anyanwu, Tyler Stotland, Nilakshi Chatterji, Tharina Messeroux, Aliya Magnuson, Leah Beaulac, and Katherine Berry. I am grateful for the entire NEDS seminar, who have listened, encouraged, and provided feedback on material. Thanks to Ellen-Marie Bransfield, Matt Hast and the Office of Student Affairs, and Michelle Frankfort for all of their support.

Renee Ciulla from the Stockbridge School of Agriculture at University of Massachusetts, Amherst has shown a great interest in this work. I thank her for sharing her commentary, insight, and ideas.

Thank you to my friends and family who have shown unconditional support, love, and understanding throughout this work. My partner, Bart McDonough, has been my biggest support. Thank you for sharing my enthusiasm and encouraging me to pursue this work, even when it seemed too daunting. Thank you for cooking dinner all those nights when I just could not. And thank you for being a wonderful, loving father to our 2-year-old son, Leon Xavier McDonough (Xavi), whom I could not imagine life without. Thank you, Xavi, for bringing such joy into our lives and for being my biggest inspiration.

My brothers – Matthew and Zachary Langlois – have been a constant source of encouragement and support throughout my life. I am deeply grateful to them. Thank you to my many aunts, uncles, and cousins for their love and support.

I am grateful for the generous financial support provided by USDA National Institute of Food and Agriculture and the National Science Foundation.

Lastly, thank you to participants of the Indonesian Family Life Survey for sharing such valuable information which we can learn from for generations.

iii. Financial Support and Declarations

I received generous financial support from the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Cooperative State Research, Education, and Extension Service Fellowship (grant number 2020-38420-30724) which provided a 3-year living stipend.

This research was partially supported by the NSF International Research Experiences for Students (IRES) grant: US-Indonesian Research Experience for Students on Sustainable Adaptation of Coastal Areas to Environmental Change (grant number 1826939—PI Naumova EN), the National Science Foundation's (NSF) Innovations in Graduate Education (IGE) Program's Solution-Oriented, Student-Initiated, Computationally-Enriched (SOLSTICE) training and learning approach (grant number 1855886), and the Tufts Institute of Environment (TIE) 2020 Environmental Research Fellowship.

I declare no financial conflicts of interest.

iv. Acronyms

DFO	Dartmouth Flood Observatory
FAO	Food and Agriculture Organization of the United Nations
GFMS	Global Flood Monitoring System
GloFAS	Global Flood Awareness System
GSW	Global Surface Water Dataset
GSWE	Global Surface Water Explorer
GSWIR	Green and Short-Wave Infra-Red
IFLS	Indonesian Family Life Survey
IPCC	Intergovernmental Panel on Climate Change
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NW	NASA Worldview
RSWIR	Red and Short-Wave Infra-Red
WI	Water Index

Table of Contents

i. Summary	ii
ii. Acknowledgements	iv
iii. Financial Support and Declarations	vi
iv. Acronyms	vii
v. List of Figures and Tables	xi
Chapter 1: Introduction	1
1.1 Background and Context	1
1.2 Flooding and Food Security – Existing Knowledge Gaps	3
1.3 Conceptual framework	5
1.4 Specific Aims	7
1.5 Visualizing Spatiotemporal Distribution of Floods Across Central Java through Dynamic Maps	9
1.6 References	14
Chapter 2: Vulnerability of households severely impacted by floods in Central Java	22
2.1 Summary	22
2.2 Household Flood Severity and Migration Extent in Central Java: Analysis of the Indonesian Family Life Survey (manuscript published in IJERPH)	23
2.2.1 Abstract	24
2.2.2 Introduction	24
2.2.3 Methods	26
2.2.4 Results	28
2.2.5 Discussion	33
2.2.6 Conclusions	35
2.3 References	36
Chapter 3: Usability of flood data portals for food security, nutrition, and health research (manuscript formatted for Nature Scientific Data)	38
3.1 Abstract	38
3.2 Introduction	39
3.2.1 Overview of remote sensing and flood mapping	40
3.2.2 Quantifying flood in studies examining the impact on nutrition and food security	41
3.2.3 Goals of this paper	42

3.2.4 Flood data portals included in this review	43
3.3 Results	45
3.3.1 Usability characteristics	45
3.3.2 External user assessment	55
3.4 Discussion	57
3.5 Methods	61
3.5.1 Usability characteristics	62
3.5.2 External user assessment	63
3.6 Code Availability	65
3.7 References	65
3.8 Acknowledgements	73
3.9 Author Contributions	74
3.10 Competing Interests	74
Chapter 4: Spatial and Survey Data Linkages	75
4.1 Linkage of Global Surface Water 1984-2015 raster dataset to a longitudinal household survey for integrative statistical analysis (TIES 2023 abstract)	75
4.2 Technical Workshop (presented at Tufts Nutrition Data Symposium 2023)	76
4.3 References	95
Chapter 5: Recurrent flooding and household food access in Central Java, Indonesia (manuscript formatted for Lancet Planetary Health)	96
5.1 Abstract	96
5.2 Introduction	97
5.3 Methods	99
5.3.1 Data Sources and Survey Design	99
5.3.2 Dataset Preparation	101
5.3.3 Definition of Flood and Outcome Variables	102
5.4 Statistical and Machine Learning Methods	104
5.5 Results	106
5.6 Discussion	115
5.7 References	118
5.8 Description of all other exploratory modelling	122
5.8.1 Linking of DFO archive data to IFLS	123
5.8.2 Exploration of variables to measure recurrent flooding	124
5.8.3 Statistical modelling and nonlinearity	124

5.8.4 Findings	125
Chapter 6: Conclusions.....	128
Appendices	130
1. Usability of existing global and national data for flood related vulnerability assessment in Indonesia (<i>manuscript published in Science of the Total Environment</i>).....	130
2. Usability of the Indonesian Family Life Survey (IFLS) To Examine Recurrent Flooding and Household Food Access in Central Java, Indonesia (<i>oral presentation at Nutrition 2022 conference</i>)	160
3. Template of usability characteristics for comparing flood monitoring portals	169
4. Webform Survey – User Feasibility Aspects of Flood Monitoring & Data Portals.....	172
5. Data and Code Sharing on Github	181
6. Comparison of IFLS modules.....	182
7. Tables of exploratory regression models	191

v. List of Figures and Tables

List of Figures (numbered in order of appearance)

1. Dimension and levels for food security measurement	3
2. Conceptual framework	7
3. Time series of flood events in Indonesia from Dartmouth Flood Observatory	12
4. Dynamic map showing progression of large floods across Indonesia since 1985	13
5. Dynamic map showing large floods across Indonesia by year since 1985	13
6. Percent agreement among raters of external user assessment of flood portals	56
7. Data linkages technical workshop title slide	77
8. Workshop objectives	78
9. Data gaps	79
10. Key differences in data formats	80
11. Key definitions	81
12. The R Studio interface	82
13. Description of raster data	83
14. Example of high-resolution raster data	84
15. Example #2 of raster data	85
16. Description of population survey data	86
17. Key challenge	87
18. Required steps to link raster and survey data	88
19. Data linking example	89
20. Step 1 of the data linking process: Download data	90
21. Step 2 of the data linking process: Crop and mask the raster	91
22. Step 3 of the data linking process: Extract and summarize raster values	92
23. Step 4 of the data linking process: Save raster as tabular dataset	93
24. Step 5 of the data linking process: Merge raster data to survey	94
25. Available data sources	95
26. Conceptual framework	106
27. Map of Central Java showing Global Surface Water Recurrence Values 1984-2015	107
28. Variable Importance Plots	109

List of Tables (numbered in order of appearance)

1. Description of Central Java Households	28
2. Generalized ordered logit/ partial proportional odds model assessing the extent moved among Central Java households	31
3. Summary of flood-related data products	46
4. Usability characteristics of flood-monitoring data portals	53
5. Usability characteristics included in review of flood data products	62
6. Evaluation metrics used in external user survey of flood data portals	63
7. Descriptive characteristics of Central Java households	110
8. Regression results	114

Chapter 1: Introduction

1.1 Background and Context

Climate changes represent a major challenge to global food security and nutrition through many possible pathways¹. Changing temperatures and rainfall patterns create more severe and variable weather events that lead to displacement of farm labor and productivity, disruption of markets, income losses, decreasing food storage safety, and increased food borne diseases². These changes could lead to a cycle of impacts having long lasting societal implications³. Furthermore, decreases in food access, diet diversity, and nutritional content of foods is expected to result from climate changes^{4,5}. Recently, NASA scientists have linked extreme wet and dry events to the warming planet and predict that climate changes will increase the frequency and intensity of floods and droughts⁶.

Flooding events are leading to significant agricultural losses^{2,7-13}. According to the Food and Agriculture Organization of the United Nations (FAO), flooding disasters accounted for nearly one fifth of total agricultural losses between 2008-2018^{13,14}. Floods are particularly damaging to agricultural crops compared to other types of disasters, such as droughts, for example, that are incremental. Floods have several attributes that determine how damaging they may be to agricultural crops including water depth, duration, velocity, sediment load, rate of water rise, and frequency of occurrence, with longer duration events being the most important characteristic¹⁵⁻¹⁷. FAO estimates that production losses also lead to nutrition losses – in terms of both energy and nutrients^{13,14}. This could have severe consequences for nutrition outcomes, especially in nations already undergoing a nutrition transition. However, to date, little is known about the impacts of floods on these outcomes.

As the world's largest island nation, Indonesia faces major threats due to flooding and inundation brought on by variable and more intense rainfall, combined with sea level rise, rising water temperatures, increasing tropical storms, and land subsidence^{18,19}. Due to this confluence of environmental and geological factors, Central Java experiences multiple types of floods – coastal flooding from high tides, flash flooding from heavy rainfall in low lying areas, and flooding from river overflow. Central Java is characterized by agricultural livelihoods and produces most of the rice in Indonesia²⁰. Currently, fresh-water and tidal flood events are increasing in northern Java and are significantly impacting crop production, income, and livelihoods of local families in the region. These hazards make the agricultural and fishery sectors highly vulnerable to losses and present a considerable challenge to food and nutrition security in addition to loss of biodiversity in fish populations^{10,20–25}. Under the current scenario, flooding could cause a severe rice deficit by 2050^{23,24}. Flooding could also impact food safety and utilization by increasing food and water borne diseases and increasing exposure to disease and infection³.

In addition, Indonesia is currently facing a double burden of malnutrition²⁶. Prevalence of obesity has steadily increased since 2000 along with the percentage of Indonesians living in urban areas, indicative of a nutrition transition²⁷. The prevalence of moderate or severe food insecurity, however, has declined in recent years. As of 2019, an estimated 6.2% were food insecure²⁷. This figure coincides with an estimated 6% of the population undernourished²⁷. Other food security indicators have also improved for Indonesia, such as utilization of basic drinking and sanitation services. These indicators, however, are single metrics and do not consider the full scope of food and nutrition security²⁸.

As an underlying driver of nutritional outcomes, understanding the role of climate changes and events in shaping food security is an important area of research. However, food security is a complex multidimensional construct consisting of three pillars: availability, access, and utilization. According to the most common definition adopted during the 1996 World Food Summit, it is a state in which “all people, at all

times, have physical and economic access to sufficient, safe and nutritious food”²⁹. This study conceptualizes food security using the Data4Diets framework and the six dimensions put forth by Jennifer Coates that better capture and understand diet-related experiences at the household level^{28,30}. Accordingly, these are nutritional adequacy, sufficient calories, preferability, safety, stability, and sustainability (**Figure 1**)³⁰.

	Quality	Quantity	Preferability	Safety	Stability	Sustainability
National, Market (Available)						
Market, Household, Individual (Accessible)						
Household, Individual (Utilizable*)						

Source: About Data4Diets | INDDEx Project [Internet]. [cited 2021 Jul 21]. Available from: <https://inddex.nutrition.ufls.edu/data4diets/about>

Figure 1. Dimensions and levels for food security measurement

1.2 Flooding and Food Security – Existing Knowledge Gaps

There is lack of evidence on how flooding is impacting diet and nutrition at the household level. The physical impacts of environmental changes on food production in the form of direct agricultural losses and decreased availability of food is evident around the globe and has been well-documented^{2,17,31–36}. Beyond the physical impacts on food production, however, it is not well-understood how climate factors, including floods, are impacting diet-related dimensions of food security and nutrition outcomes at the household level such as access to nutritious foods or other aspects of the food value chain^{5,35}. One recent study did find floods to have direct impacts on diet quality through impaired livelihoods and income³⁷. There is a wider body of literature on the impact of floods on livelihoods^{38–40}. The link to diet and nutritional outcomes, however, is unclear. A small but growing body of literature has focused on climate variability and nutritional

outcomes^{41–46}. Other related works have focused on impacts of rising temperatures, drought, or rainfall variability on food security in more arid climates^{3,47–51}. A few studies have found flooding to have effects on malnutrition^{52–54}. More evidence is needed, however⁵. Understanding what consequences flooding has on diet and nutrition may help mitigate long term negative impacts.

Most research focuses on single rather than *recurrent* flood events. A few studies have looked at diet and nutrition following disasters, but focused on the immediate recovery responses following a singular event or only on food production or availability^{55–64}. One study found long term health impacts associated with flood in Bangladesh, but it was focused on a single disaster⁶². Other studies have looked at migration as a coping response to food insecurity caused by variable rainfall or climate hazards^{60,65–71}. Some evidence from other agrarian contexts find vulnerability of agricultural livelihoods to variable rainfall patterns⁷². Overall, however, evidence on recurring flooding and household-level food security is poor. Recently, there has been more attention to this nexus in the literature. One recent study from Ghana found recurrent floods decreased food production of subsistent farmers³⁵. Another recent study from Nigeria, a context where flooding is recurrent, highlighted how flooding impacts all aspects of food security – availability, access, utilization, and stability – but has not received enough attention in the literature, despite the importance to the Sustainable Development Goals⁷³.

Current evidence in the Indonesian context is limited. Indonesia is one of the most vulnerable countries due to flood hazards, but current studies related to food security in Indonesia have mainly focused on crop production^{16,17,24,74}. It is unclear whether households are adapting in terms of food access outcomes like they are in other ways. In Indonesia, households living in areas with frequent flooding have developed strategies to adapt such as raising the floor levels (e.g. using stilts), using floating structures, and creating dams to block water from entering the house⁷⁵. But it is not understood whether these adaptation mechanisms are effective in terms of

household food access over time or whether there is a shift toward decreased food access despite being adapted to living with the flooding.

There is a lack of data designed specifically for the purpose of studying how environmental factors shape population outcomes such as food security, nutrition, and health. Remote sensing data products will play an important role in filling this gap. A few studies have linked spatial and survey data sources, including with the Indonesian Family Life Survey^{5,44,76–78}. But these cross-disciplinary data linkages are challenging due to differences in how data streams are captured, measured, and prepared and their different spatiotemporal attributes.

To address these gaps, I sought to link a geospatial flood data product to the longitudinal Indonesian Family Life Survey (IFLS) by comparing usability aspects of flood data portals. After reviewing several flood data products, I obtained exploratory measures of recurrent flooding which were merged to IFLS, creating a dataset for statistical modelling from different databases. The linked, longitudinal dataset contained information pertaining to flooding at the district level and food access at the household level. I sought to understand how flooding is shaping household access to foods that provide sufficient energy and are nutritionally adequate and whether there are differential effects based on socioeconomic status or access to credit.

1.3 Conceptual framework

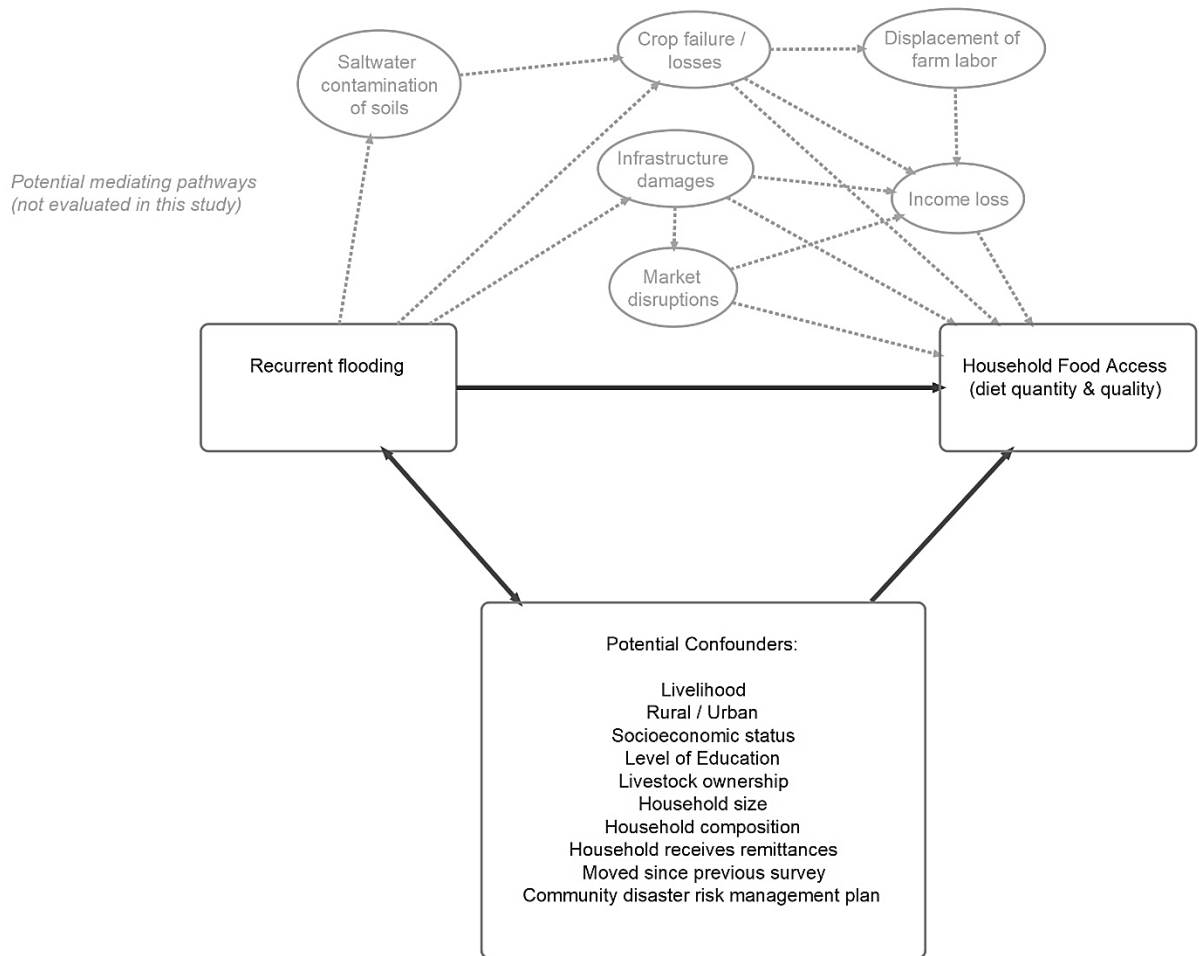
In the context of this study, I refer to *recurrent flooding* as repeated flooding over the 1984-2015 period as opposed to a single flood event. I explored recurrent flooding from different angles using the Global Surface Water dataset (GSW), which measures gradual changes in surface water at 30m resolution since 1984, and Dartmouth Flood Observatory (DFO), which records occurrences of large and extreme

floods since 1985. While DFO captures rare events, I explored the frequency of these events over time. The GSW variable served as the primary measure.

Figure 2 displays a conceptual framework developed for this research. The framework presents the suspected relationship between recurring flooding events and household food access. Household size, composition, and livestock ownership are potential confounders because these factors have been shown to positively impact food security in other studies^{51,55}. Rising household income has also been shown to improve food access³. Socioeconomic status and whether the household lives in a rural or urban area are potential confounders. Prior research has shown that the rural poor may be more vulnerable to negative impacts of flooding⁴. Each of these factors may also be associated with exposure to flooding, as they may be determinants of living in areas more prone to flooding.

There are several mediator variables. Saltwater intrusion from coastal floods leads to contamination of soils and consequently crop losses. It is known that flooding causes destruction of crops, which could affect household food access through decreased food production and availability or increased food prices. Flooding can also cause damage to infrastructure, which could impact food access in several ways. One study that assessed the impact of a flooding event on food security found household income losses and food insecurity in flood affected areas⁵⁹. These mediating pathways were not evaluated in this study, however.

Figure 2. Conceptual framework



1.4 Specific Aims

The long-term goal is to improve household access to nutritious foods in areas vulnerable to recurring flooding events. My research explored the relationship between flooding and food access outcomes over space and time, aiding our understanding of *how* floods are affecting household food access. My research adds to a growing body of literature on how climate events are impacting lives through diet and food security. Most importantly, my research addresses the data gap by demonstrating how environmental and survey data can be linked.

My specific aims were as follows:

1.4.1 Aim 1: Compare usability of different flood monitoring/data portals in terms of their applicability to the current study. Through this aim, I evaluated the following flood portals by reviewing and gathering information about their usability characteristics and conducting a feasibility survey: Global Flood Monitoring System (GFMS), Global Surface Water Explorer (GSWE), JRC's Global Flood Awareness System (GloFAS), NASA Worldview (NW), and Dartmouth Flood Observatory (DFO). I identified data products to merge with IFLS to obtain measures of flooding at the district-level.

1.4.2 Sub-aim 1.1: Determine and describe the spatiotemporal distribution of flood events across Central Java. I created a dynamic, timelapse map using flood event data from DFO to show the spatiotemporal distribution of floods across Central Java since 1985.

1.4.3 Aim 2: Evaluate how recurring flooding has impacted household access to sufficient quantities of food (diet quantity). Using household food purchase and consumption data from IFLS and the Data4Diets framework, I constructed the household food expenditure share (% of total spending) variable to measure household diet quantity³⁰. I assessed the linear relationship between flooding and food expenditure share using linear mixed models. I expected that more flooding would be associated with increased household food expenditure share (i.e. greater food insecurity) throughout the timespan (1984-2015).

1.4.4 Sub-aim 2.1: Assess the form and shape of the relationship between flooding and food quantity over time by exploring nonlinearity. Using nonlinear mixed models, I expected to see a threshold point in the relationship between flooding and food expenditure share where the relationship would reverse as households adapted.

1.4.5 Sub-aim 2.2: Assess whether differential effects exist based on indicators of adaptive capacity. I expected the effect of flooding on food expenditure share to be less severe for households with better adaptive capacity. I assessed the joint

effect between flooding and household socioeconomic status and access to credit.

1.4.6 Aim 3: Evaluate how recurring flooding has impacted household access to adequate nutrition (diet quality). Like Aim 2, I evaluated this aim using IFLS household food purchase and consumption data and the Data4Diets framework. I measured household diet quality by constructing the household dietary diversity score variable³⁰. I assessed the linear relationship between flooding and household dietary diversity score using mixed effects ordered logistic regression models. I expected that more flooding would be associated with decreased dietary diversity throughout the timespan (1984-2015).

1.4.7 Sub-aim 3.1: Assess the form and shape of the relationship between flooding and food quality over time by examining non-linearity. Using nonlinear mixed effect ordinal models, I expected to see a threshold point in the relationship between flooding and dietary diversity where the relationship would reverse as households adapted.

1.4.8 Sub-aim 3.2: Assess whether differential effects exist based on indicators of adaptive capacity. I expected the effect of flooding on dietary diversity score to be less severe for households with better adaptive capacity. I assessed the joint effect between flooding and household socioeconomic status and access to credit.

1.5 Visualizing Spatiotemporal Distribution of Floods Across Central Java through Dynamic Maps

To illustrate the spatiotemporal distribution of floods across Central Java, I created Dynamic Maps using R. Dynamic maps are helpful, exploratory visualization tools showing temporal and spatial information at the same time in the form of a movie,

like a google timelapse. This spatiotemporal distribution is represented by the model below:

$$E[X_{ts}] = \beta_0 + \beta_1 f(t, s) + e$$

where $E[X_{ts}]$ represents the number of floods at time t over area s . $\beta_1 f(t, s)$ represents a function of space-time. This method is described in detail in the 2009 article by Castronovo et al⁷⁹.

I used flood event data from the DFO's historical archive. DFO captures large events in coarsely drawn polygons. **Figure 3** shows time series of this data spanning 1985 to 2020. Over time, the number of floods oscillates, with an increase in the number of very large events after 2003 (**Figure 3**).

By exploring the spatiotemporal distribution, we can learn more about which areas are more affected and when through simultaneously visualizing distributions over both space and time in a moving timelapse map. To create the dynamic maps, I loaded shapefiles of Indonesia's administrative boundaries and DFO's archive of flood events into R using the 'sf' package. Next, I created month, year, and duration variables from the flood start and end dates. Then, I subset and simplified the map data using the 'rmapshaper' package to improve the speed of GIF rendering. I then created a base map of Indonesia from which to overlay the DFO data. DFO contains information about flood severity. To visualize this information, each event was highlighted in ascending severity (orange to red depending on severity level) in a map overlaying DFO data onto the base map. Finally, the dynamic maps were generated by animating these maps and saving the animation as a gif file. Two dynamic maps were created: one by flood event, and one by year. The code to create the dynamic maps was adapted from a previous example (<https://github.com/meghartwick>).

The dynamic maps are shown in **Figures 4 & 5** and can also be viewed on GitHub at <https://github.com/BreeLanglois>. Sumatra and Java are heavily hit by floods, with the most severe events occurring most frequently in Sumatra. In Java, the yearly intervals between flood occurrences shortens over time, especially after 2006, indicating

increasing frequency of these large events. It should be noted, however, that DFO is based mainly on media sources and may be subject to reporting bias.

Figure 3. Time series of flood events in Indonesia from Dartmouth Flood Observatory

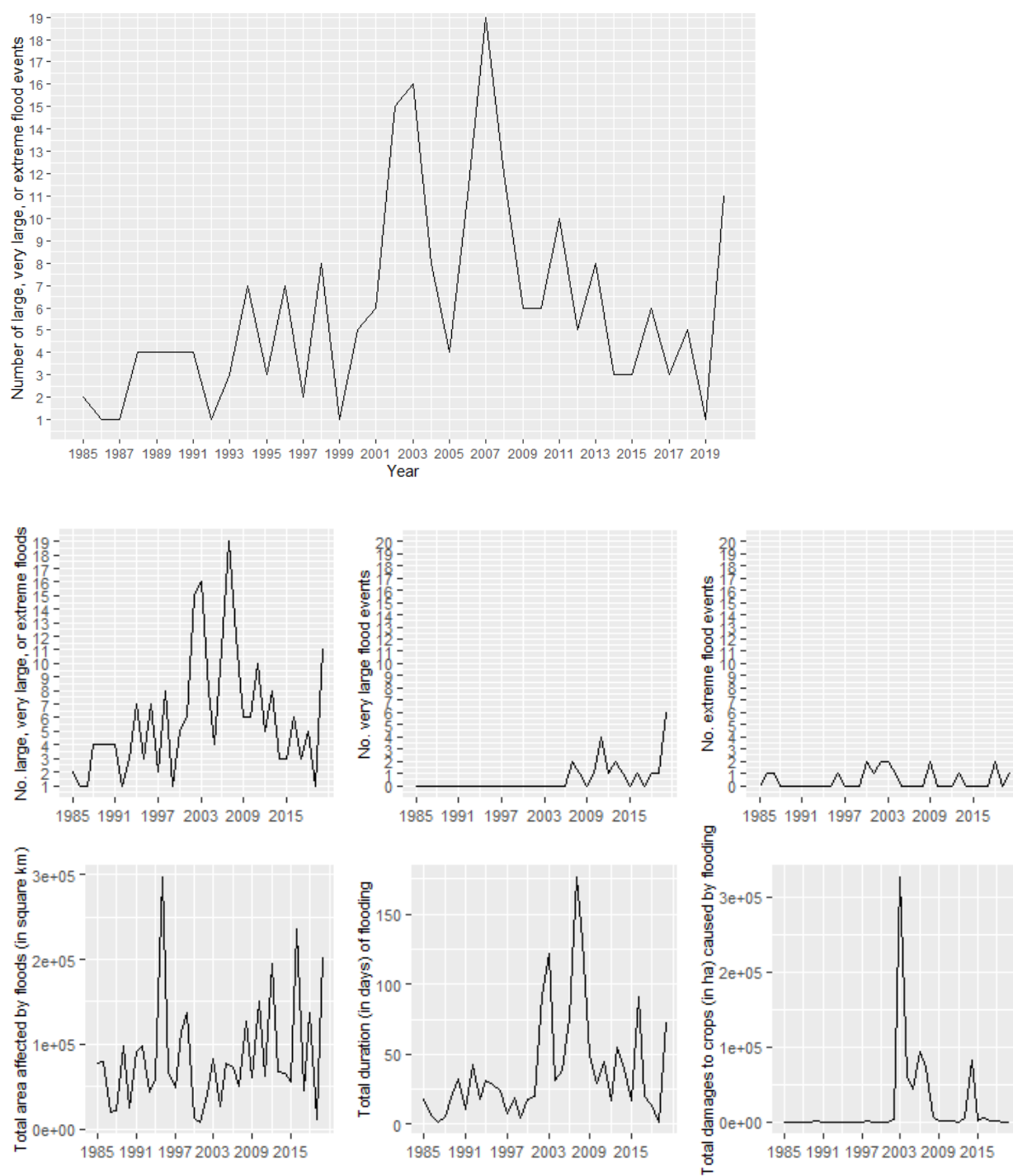


Figure 4. Dynamic map showing progression of large floods across Indonesia since 1985

1985-02-19

Frame 1 of 80



Figure 5. Dynamic map showing large floods across Indonesia by year since 1985

Year: 1985

Frame 1 of 36



1.6 References

1. Stringer, L. *et al.* Adaptation and development pathways for different types of farmers: key messages. (2019).
2. Salm, L., Nisbett, N., Cramer, L., Gillespie, S. & Thornton, P. How climate change interacts with inequity to affect nutrition. *Wiley Interdiscip. Rev. Clim. Chang.* **12**, e696 (2021).
3. Yadav, S. S., Redden, R. J., Hatfield, J. L., Ebert, A. W. & Hunter, D. Food Security and Climate Change. *Food Secur. Clim. Chang.* (2018)
doi:10.1002/9781119180661.
4. Fanzo, J., Davis, C., McLaren, R. & Choufani, J. The effect of climate change across food systems: Implications for nutrition outcomes. *Glob. Food Sec.* **18**, 12–19 (2018).
5. Niles, M. T. *et al.* Climate impacts associated with reduced diet diversity in children across nineteen countries. *Environ. Res. Lett.* **16**, 15010 (2021).
6. Rodell, M. & Li, B. Changing intensity of hydroclimatic extreme events revealed by GRACE and GRACE-FO. *Nat. Water* **2023 13 1**, 241–248 (2023).
7. Cobon, D. H., Ewai, M., Inape, K. & Bourke, R. M. Food shortages are associated with droughts, floods, frosts and ENSO in Papua New Guinea. *Agric. Syst.* **145**, 150–164 (2016).
8. Zhang, Q., Gu, X., Singh, V. P., Liu, L. & Kong, D. Flood-induced agricultural loss across China and impacts from climate indices. *Glob. Planet. Change* **139**, 31–43 (2016).
9. Adewuyi. Spatio-Temporal Analysis of Flood Incidence in Nigeria and Its Implication for Land Degradation and Food Security. *J. Agric. Sci.* **6**, 150 (2014).
10. *Floods ruin thousands of hectares of crops - Indonesia | ReliefWeb.*

11. Agricultural Loss Caused by 2007 Sidoharjo's Flood and its Household Impact - Agricultural & Environmental Science Collection - ProQuest.
12. Paul, B. K. & Rasid, H. Flood damage to rice crop in Bangladesh. *Geogr. Rev.* **83**, 150–159 (1993).
13. Food and Agriculture Organization of the United Nations. *The impact of disasters and crises on agriculture and food security: 2021*. (2021).
14. Food and Agriculture Organization of the United Nations. Agriculture on the Proving Grounds: Damage and loss. <https://www.fao.org/resources/digital-reports/disasters-in-agriculture/en/>.
15. Muh Aris Marfai, B. GIS Modelling of River and Tidal Flood Hazards in a Waterfront City. (International Institute for Geo-Information Science and Earth Observation, 2003).
16. Widiarto, L. A. & Kingma, N. Agricultural Loss Caused by 2007 Sidoharjo's Flood and its House-hold Impact. *Indones. J. Geogr.* **46**, 156 (2014).
17. Pacetti, T., Caporali, E. & Rulli, M. C. Floods and food security: A method to estimate the effect of inundation on crops availability. *Adv. Water Resour.* **110**, 494–504 (2017).
18. Marfai, M. A. *et al.* Natural hazards in Central Java Province, Indonesia: an overview. *Environ. Geol.* **56**, 335–351 (2008).
19. Marfai, M. A. & King, L. Tidal inundation mapping under enhanced land subsidence in Semarang, Central Java Indonesia. *Nat. Hazards* **44**, 93–109 (2008).
20. Lassa, J. A. *Emerging 'Agricultural Involution' in Indonesia: Impact of Natural Hazards and Climate Extremes on Agricultural Crops and Food System*. (2012).
21. Dewan Ketahanan Pangan, K. P. and W. F. P. (WFP). *Food Security and Vulnerability Atlas of Indonesia 2015*. (2015).

22. Marfai, M. A. & King, L. Potential vulnerability implications of coastal inundation due to sea level rise for the coastal zone of Semarang city, Indonesia. *Environ. Geol.* **54**, 1235–1245 (2007).
23. Marfai, M. Impact of coastal inundation on ecology and agricultural land use case study in central Java, Indonesia. *Quaest. Geogr.* **30**, 19–32 (2011).
24. Syaukat, Y. The Impact of Climate Change on Food Production and Security and Its Adaptation Programs In Indonesia. *J. ISSAAS* **17**, 40–51 (2011).
25. Rizal, A. & Anna, Z. Climate change and its possible food security implications toward Indonesian marine and fisheries. *World News Nat. Sci.* **22**, 119–128 (2019).
26. Lowe, C. *et al.* The double burden of malnutrition and dietary patterns in rural Central Java, Indonesia. *Lancet Reg. Heal. – West. Pacific* **14**, 100205 (2021).
27. FAOSTAT. <https://www.fao.org/faostat/en/#country/101>.
28. Coates, J. Build it back better: Deconstructing food security for improved measurement and action. *Glob. Food Sec.* **2**, 188–194 (2013).
29. Rome Declaration and Plan of Action.
<https://www.fao.org/3/w3613e/w3613e00.htm>.
30. Data4Diets: Food Security Indicators | INDDEx Project.
<https://inddex.nutrition.tufts.edu/data4diets/indicators>.
31. Aggarwal, P., Vyas, S., Thornton, P. & Campbell, B. M. How much does climate change add to the challenge of feeding the planet this century? *Environ. Res. Lett.* **14**, 043001 (2019).
32. Thornton, P., Dinesh, D., Cramer, L., Loboguerrero, A. M. & Campbell, B. Agriculture in a changing climate: Keeping our cool in the face of the hothouse: <https://doi.org/10.1177/0030727018815332> **47**, 283–290 (2018).

33. Thornton, P. K. *et al.* Is agricultural adaptation to global change in lower-income countries on track to meet the future food production challenge? *Glob. Environ. Chang.* **52**, 37–48 (2018).
34. Muhardiono, I. & Hamdani, A. Flood vulnerability impact for food estate potential in Central Kalimantan, Indonesia. *IOP Conf. Ser. Earth Environ. Sci.* **648**, (2021).
35. Porter L. Xie A.J. Challinor K. Cochrane S.M. Howden M.M. Iqbal D.B. Lobell and M.I. Travasso, J. R. *Food security and food production systems. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* (2014).
36. Baeribameng Yiran, G. A., Atubiga, J. A., Kusimi, J. M., Kwang, C. & Owusu, A. B. Adaptation to perennial flooding and food insecurity in the Sudan savannah agroecological zone of Ghana. *Environ. Res.* **214**, 114037 (2022).
37. Oskorouchi, H. R. & Sousa-Poza, A. Floods, food security, and coping strategies: Evidence from Afghanistan. *Agric. Econ. (United Kingdom)* **52**, 123–140 (2021).
38. Hoq, M. S., Raha, S. K. & Hossain, M. I. Livelihood Vulnerability to Flood Hazard: Understanding from the Flood-prone Haor Ecosystem of Bangladesh. *Environ. Manage.* **67**, 532–552 (2021).
39. Khayyam, U. & Noureen, S. Assessing the adverse effects of flooding for the livelihood of the poor and the level of external response: a case study of Hazara Division, Pakistan. *Environ. Sci. Pollut. Res. Int.* **27**, 19638–19649 (2020).
40. Ali, A. & Rahut, D. B. Localized Floods, Poverty and Food Security: Empirical Evidence from Rural Pakistan. *Hydrol.* **2020** **7**, 2 (2019).
41. Randell, H., Gray, C. & Shayo, E. H. Climatic conditions and household food security: Evidence from Tanzania. *Food Policy* **112**, 102362 (2022).
42. Grace, K., Davenport, F., Hanson, H., Funk, C. & Shukla, S. Linking climate change

- and health outcomes: Examining the relationship between temperature, precipitation and birth weight in Africa. *Glob. Environ. Chang.* **35**, 125–137 (2015).
43. Cooper, M. W. *et al.* Mapping the effects of drought on child stunting. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 17219–17224 (2019).
 44. Thiede, B. C. & Gray, C. Climate exposures and child undernutrition: Evidence from Indonesia. *Soc. Sci. Med.* **265**, 113298 (2020).
 45. Cornwell, K. & Inder, B. Child Health and Rainfall in Early Life. <http://dx.doi.org/10.1080/00220388.2014.976618> **51**, 865–880 (2015).
 46. Mank, I. *et al.* The Impact of Rainfall Variability on Diets and Undernutrition of Young Children in Rural Burkina Faso. *Front. public Heal.* **9**, 693281 (2021).
 47. Coughlan de Perez, E. *et al.* From rain to famine: assessing the utility of rainfall observations and seasonal forecasts to anticipate food insecurity in East Africa. *Food Secur.* **11**, 57–68 (2019).
 48. Tankari, M. R. Rainfall variability and farm households' food insecurity in Burkina Faso: nonfarm activities as a coping strategy. *Food Secur.* **12**, 567–578 (2020).
 49. Drysdale, R. E., Moshabela, M. & Bob, U. 'A creeping phenomenon': the association between rainfall and household food insecurity in the district of iLembe, KwaZulu-Natal. *Clim. Dev.* **13**, 128–138 (2021).
 50. Sassi, M. & Cardaci, A. Impact of rainfall pattern on cereal market and food security in Sudan: Stochastic approach and CGE model. *Food Policy* **43**, 321–331 (2013).
 51. Demeke, A. B., Keil, A. & Zeller, M. Using panel data to estimate the effect of rainfall shocks on smallholders food security and vulnerability in rural Ethiopia. *Clim. Change* **108**, 185–206 (2011).
 52. Gaire, S., Delbiso, T. D., Pandey, S. & Guha-Sapir, D. Impact of disasters on child stunting in Nepal. *Risk Manag. Healthc. Policy* **9**, 113 (2016).

53. Rodriguez-Llanes, J. M., Ranjan-Dash, S., Degomme, O., Mukhopadhyay, A. & Guha-Sapir, D. Child malnutrition and recurrent flooding in rural eastern India: a community-based survey. *BMJ Open* **1**, e000109 (2011).
54. Lieber, M., Chin-Hong, P., Kelly, K., Dandu, M. & Weiser, S. D. A systematic review and meta-analysis assessing the impact of droughts, flooding, and climate variability on malnutrition. <https://doi.org/10.1080/17441692.2020.1860247> (2020) doi:10.1080/17441692.2020.1860247.
55. Rukundo, P. M. *et al.* Household food insecurity and diet diversity after the major 2010 landslide disaster in Eastern Uganda: A cross-sectional survey. *Br. J. Nutr.* **115**, 718–729 (2016).
56. RISK FACTOR INFLUENCING FOOD SECURITY OF THE CHILDREN DURING FLOOD DISASTER - Agricultural & Environmental Science Collection - ProQuest.
<https://www.proquest.com/agricenvironm/docview/2061899548/2097818786F840DEPQ/21?accountid=14434>.
57. Doocy, S., Leidman, E., Aung, T. & Kirsch, T. Household economic and food security after the 2010 Pakistan floods. *Food Nutr. Bull.* **34**, 95–103 (2013).
58. Hidayat, A. R. T., Rachmawati, T. A. & Wahyuningtiyas, L. Food security in the disaster-prone area: an empirical study from the rural area of Indonesia.
doi:10.1088/1755-1315/653/1/012011.
59. Breisinger, C., Ecker, O., Thiele, R. & Wiebelt, M. Effects of the 2008 flood on economic performance and food security in Yemen: A simulation analysis. *Disasters* **40**, 304–326 (2016).
60. Sherman, M., Ford, J., Llanos-Cuentas, A. & Valdivia, M. J. Food system vulnerability amidst the extreme 2010–2011 flooding in the Peruvian Amazon: a case study from the Ucayali region. *Food Secur.* **8**, 551–570 (2016).
61. Salvucci, V. & Santos, R. Vulnerability to Natural Shocks: Assessing the Short-Term

- Impact on Consumption and Poverty of the 2015 Flood in Mozambique. *Ecol. Econ.* **176**, (2020).
62. del Ninno, C. & Lundberg, M. Treading water. The long-term impact of the 1998 flood on nutrition in Bangladesh. *Econ. Hum. Biol.* **3**, 67–96 (2005).
 63. Martalena Purb. Food and Nutrition Security During Natural Disaster Crisis: Indonesian Experience. in *Annals of Nutrition & Metabolism* (2019).
 64. Del Ninno, C. & Dorosh, P. A. Averting a food crisis: private imports and public targeted distribution in Bangladesh after the 1998 flood. *Agric. Econ.* **25**, 337–346 (2001).
 65. Warner, K. & Afifi, T. Where the rain falls: Evidence from 8 countries on how vulnerable households use migration to manage the risk of rainfall variability and food insecurity. *Clim. Dev.* **6**, 1–17 (2014).
 66. Etzold, B., Ahmed, A. U., Hassan, S. R. & Neelormi, S. Clouds gather in the sky, but no rain falls. Vulnerability to rainfall variability and food insecurity in Northern Bangladesh and its effects on migration. *Clim. Dev.* **6**, 18–27 (2014).
 67. Rainfall variability, food insecurity and migration in Cabricán, Guatemala - Agricultural & Environmental Science Collection - ProQuest.
 68. Rademacher-Schulz, C., Schraven, B. & Mahama, E. S. Time matters: Shifting seasonal migration in Northern Ghana in response to rainfall variability and food insecurity. *Clim. Dev.* **6**, 46–52 (2014).
 69. Afifi, T., Liwenga, E. & Kwezi, L. Rainfall-induced crop failure, food insecurity and out-migration in Same-Kilimanjaro, Tanzania. *Clim. Dev.* **6**, 53–60 (2014).
 70. Murali, J. & Afifi, T. Rainfall variability, food security and human mobility in the Janjgir-Champa district of Chhattisgarh state, India. *Clim. Dev.* **6**, 28–37 (2014).
 71. Ahmad, D. & Afzal, M. Flood hazards, human displacement and food insecurity in rural riverine areas of Punjab, Pakistan: policy implications. *Environ. Sci. Pollut.*

Res. **28**, 10125–10139 (2021).

72. Olayide, O. E. & Alabi, T. Between rainfall and food poverty: Assessing vulnerability to climate change in an agricultural economy. *J. Clean. Prod.* **198**, 1–10 (2018).
73. Echendu, A. J. Flooding, Food Security and the Sustainable Development Goals in Nigeria: An Assemblage and Systems Thinking Approach. *Soc. Sci.* **11**, 59 (2022).
74. Riadi, B. Strategy to Maintain Food Security in the Area of Flood Hazard in Karawang Regency. *IOP Conf. Ser. Earth Environ. Sci.* **165**, 12029 (2018).
75. Marfai, M. A. & Hizbaron, D. R. Community's adaptive capacity due to coastal flooding in Semarang coastal city, Indonesia. *Int. J. Ser. Geogr. Ann. Univ. Oradea* **21**, 2011 (2011).
76. Amare, M., Jensen, N. D., Shiferaw, B. & Cissé, J. D. Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agric. Syst.* **166**, 79–89 (2018).
77. Ji, M., An, R., Qiu, Y. & Guan, C. The Impact of Natural Disasters on Dietary Intake. *Am. J. Health Behav.* **44**, 26–39 (2020).
78. Bohra-Mishra, P., Oppenheimer, M. & Hsiang, S. M. Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9780–9785 (2014).
79. Castronovo, D. A., Chui, K. K. H. & Naumova, E. N. Dynamic maps: A visual-analytic methodology for exploring spatio-temporal disease patterns. *Environ. Heal. A Glob. Access Sci. Source* **8**, 1–9 (2009).

Chapter 2: Vulnerability of households severely impacted by floods in Central Java

2.1 Summary

This chapter examines the vulnerability of households suffering severe economic losses from floods through the lens of migration. It is commonly assumed that households will migrate from degraded areas and settle in new locations. However, emerging literature indicates that losses incurred from flood events may limit the ability of some households to migrate from heavily flood affected areas, creating “trapped” populations who may become increasingly vulnerable. Understanding migration of households most severely impacted by floods is important, as it may have implications for food insecurity and public health. A challenge is to identify flood affected households and distinguish the severity of impacts. This chapter defined a measure of flood severity based on the reported impacts to households using the IFLS. Findings revealed households that are most vulnerable to the impacts of floods tended to stay locally within affected areas, having implications for national disaster mitigation and response efforts. This highlights the importance of studying how flooding shapes food access in these areas in addition to the need for usable flood data products.



Article

2.2 Household Flood Severity and Migration Extent in Central Java: Analysis of the Indonesian Family Life Survey (*manuscript published in IJERPH*)

Breanne K. Langlois ^{1,*}, Leah Beaulac ¹, Katherine Berry ¹, Oyedolapo Anyanwu ¹, Ryan B. Simpson ¹, Aris Ismanto ², Magaly Koch ³, Erin Coughlan de Perez ^{1,4}, Timothy Griffin ¹ and Elena N. Naumova ¹

¹ Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA 02111, USA; leah.k.beaulac@gmail.com (L.B.); kgberry122@gmail.com (K.B.); oyedolapo.anyanwu@tufts.edu (O.A.); ryan.simpson@tufts.edu (R.S.); erin.coughlan@tufts.edu (E.C.d.P.); timothy.griffin@tufts.edu (T.G.); elena.naumova@tufts.edu (E.N.N.)

² Department of Oceanography, Faculty of Fisheries and Marine Science, Universitas Diponegoro, Semarang 50275, Indonesia; aris.ismanto@gmail.com

³ Center for Remote Sensing, Department of Earth & Environment, Boston University, Boston, MA 02215, USA; mkoch@bu.edu

⁴ Red Cross Red Crescent Climate Centre, 2502 KC The Hague, The Netherlands

* Correspondence: breanne.langlois@tufts.edu

2.2.1 Abstract

Central Java, Indonesia, is prone to river and coastal flooding due to climate changes and geological factors. Migration is one possible adaptation to flooding, but research is limited due to lack of longitudinal spatially granular datasets on migration and metrics to identify flood-affected households. The available literature indicates social and economic barriers may limit mobility from flood prone areas. The Indonesian Family Life Survey (IFLS) provides self-reported data on household experiences with natural disasters among 1501 Central Java households followed over two waves (2007 and 2014). We examined how the severity of flooding, defined by household-level impacts captured by the IFLS (death, injury, financial loss, or relocation of a household member), influenced the extent of household movement in Central Java using a generalized ordered logit/ partial proportional odds model. Households severely impacted by floods had 75% lower odds of moving farther away compared to those that did not experience floods. The most severely impacted households may be staying within flood-affected areas in Central Java. Public health, nutrition, and economic surveys should include modules focused on household experiences, impacts, and adaptations to facilitate the study of how climate changes are impacting these outcomes.

Keywords: flooding; disasters; climate change; migration; adaptation; vulnerability

Citation: Langlois, B.K.; Beaulac, L.; Berry, K.; Anyanwu, O.; Simpson, R.B.; Ismanto, A.; Koch, M.; Coughlan de Perez, E.; Griffin, T.; Naumova, E.N. Household Flood Severity and Migration Extent in Central Java: Analysis of the Indonesian Family Life Survey. *Int. J. Environ. Res. Public Health* **2023**, *20*, x. <https://doi.org/10.3390/xxxxx>

Academic Editors: Qingxiang Li, Steve Hung Lam Yim, Yonghong Li

Received: 31 December 2022

Revised: 12 April 2023

Accepted: 27 April 2023

Published: date



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2.2.2 Introduction

1. Introduction (*manuscript heading*)

Climate-related migration is directly linked to public health, as migrants and those left behind face food insecurity, malnutrition, and reduced access to services among other health threats [1–3]. While migration out of degraded areas is one example of adaption to climate change, the main push factors are often deteriorating livelihoods and economic situations [4]. There is global political interest in addressing these interconnected issues, with calls to prioritize migrant health interventions and include migration in national climate change plans during the 26th Conference of the Parties (COP26) for the United Nations Framework Convention on Climate Change [3].

While there is global attention to the issue of climate migration, especially in island nations experiencing frequent flooding from sea level rise and extreme weather, there are significant challenges to research in this area [5,6]. According to the most recent report from the International Organization for Migration (IOM), the various forms of migration across space and time (e.g., within a village or internationally, temporary or permanent), in addition to multiple push and pull factors, and the lack of suitable datasets make climate migration a difficult phenomenon to define [5]. While the “Migration, Climate Change, and the Environment” bibliographic database (CliMig) collates case studies related to climate migration, there is a lack of longitudinal and spatially granular data on migration to evaluate the multiple drivers and slow-onset environmental processes [7]. Some surveys, including the Indonesian Family Life Survey (IFLS), ask respondents directly about displacement due to disasters, but such

information only captures the immediate post-disaster response. Further, while flooding affects more people worldwide than any other climate-related disaster, drought and desertification remain the most studied hazards in the CliMig database [5].

In areas such as Central Java, flooding caused by both coastal inundation and river overflow from heavy precipitation is a regular occurrence [8–10]. Many studies on the impact of floods in the Indonesian context examine either coastal or river flooding or focus more broadly on sea level rise [11–16]. More research is needed on how all forms of flooding impact long term migration. The Internal Displacement Monitoring Centre (IDMC) estimated 155,000 internally displaced persons from disasters in Indonesia in 2021, but this figure does not distinguish temporary versus permanent movement or the main cause [17].

The literature indicates there may be substantial barriers to migration among poor, vulnerable households where climate hazards may limit capacity to move, challenging common assumptions about migration [18–24]. In one of the first robust analyses using longitudinal data to examine flood related migration in rural Bangladesh, researchers found flooding to have a minimal impact, while non-flood related crop failure had a large effect [19,20]. Studies from Indonesia show a preference to stay in affected areas and adapt [15]. Migration out of flood-affected areas is often a last resort due to social factors and the economic burden of migration, especially if the flooding is gradual or not severe [4,15]. Local adaptations to “the Rob” —a term used to describe the regular occurrence of seawater overflow during tides—include raising floor levels using concrete or stilts, using floating structures, or creating dams to block water from entering households [11,15]. This research points to the need to examine the household-level impacts and how vulnerability, economic barriers, and adaptive capacity may shape migration; however, identifying flood-affected households is challenging, and there is no metric of flood severity that incorporates the household-level impacts.

Self-reported data on experiences with natural disasters from the IFLS allows for the identification and study of flood-affected households over a 12-year span captured through two of the five cross-sectional survey waves. As a nationally representative, longitudinal economic and health survey that tracked Indonesian households from 1993 to 2014, the IFLS is robust in reflecting the overall population with its multi-stage sampling scheme, large sample size (15,921 interviewed households including 1949 in Central Java), and high tracking rate (94%) [25]. Other strengths of the IFLS include its multipurpose nature, extended time period (21 years), and availability of both current and retrospective data.

We examined how the severity of flooding, defined by household-level impacts of floods, influenced the extent of household movement in Central Java in the 1501 IFLS households that were tracked over two waves in 2007 and 2014. Specifically, we: (1) describe flooding and other type of disasters experienced by households and their reported direct impacts, and (2) evaluate how the severity of flooding, agricultural livelihood, and socio-economic characteristics influenced the extent moved. We discuss usability and data quality aspects of the IFLS for this purpose. This paper provides insight into how flooding disasters are impacting households in a largely agricultural area, Central Java, and may help with disaster planning and mitigation efforts.

2.2.3 Methods

2. Materials and Methods (*manuscript heading*)

2.1. Design of the Indonesian Family Life Survey

The RAND Corporation and Survey Meter conducted the IFLS. Ethical approval was obtained by the RAND Institutional Review Board in the United States and the University of Gadjah Mada Institutional Review Board in Indonesia. The IFLS is designed to study a wide range of behaviors and outcomes at individual and household levels including economic, well-being, consumption, income and wealth, education, and health indicators.

The IFLS is a longitudinal, population-based survey that was conducted over five cross-sectional waves fielded in 1993, 1997, 2000, 2007, and 2014. Due to the longitudinal design, sampling for the IFLS was based on wave one, with each subsequent wave attempting to contact and reinterview this sample. The original wave one survey sample from 1993 used a two-stage stratification process based on province and urban/rural location [26]. Households were then randomly selected from within each stratum. For cost reasons, the IFLS sampled from only half of the provinces. A total of 7224 households from 13 of 27 provinces participated in 1993, representing ~83% of the country's total population at the time. In addition to these "origin" households, the IFLS tracked newly formed households that contained a target member of an original household. The IFLS collected detailed demographic, economic, livelihood, and health information across all five waves from 6044 households (84% of the 1993 sample), including 828 from Central Java (94% of the 1993 Central Java sample). The survey incorporated a natural disasters module in waves four (2007) and five (2014). This module captured self-reported experiences with natural disasters in the prior five years, including the severity and impacts. More detailed information about the survey including data collection and verification process, instruments, and data cleaning can be found in the user guides [26].

A total of 1501 Central Java households were tracked across waves four and five (13% of all interviewed households followed over both waves, 93% tracking rate of wave four households in wave five). Of these, we analyzed 1472 (98%) that had complete data. We estimate our detectable effect size is 10 percentage points based on our sample size, an 8:1 group ratio, and 20% migrating among those not affected by flood.

2.2. Outcomes and Predictors

The main outcome was the extent the household had moved since the time of the last survey, between 2000 and 2007 for wave four and between 2007 and 2014 for wave five. This was a 6-level ordinal variable defined as: (0) the household did not move; or the household moved (1) within the same village, (2) within the same district, (3) within the same regency, (4) within the same province, or (5) to another province. We obtained this variable from the household tracking files provided with IFLS data. Field interviewers tracked households by returning to the address where the household was in the previous survey and, if the entire household was missing, locating all target members. While the IFLS included a module with direct questioning about the impacts of flooding and other disasters, including time spent without housing, it only captured the immediate response of some household members. Our interest was in the longer-term impact. Therefore, we used whole-household moves as our measure of migration.

All independent variables were self-reported from the household survey. Severity of flooding experienced by households was the main independent variable, defined as: (0) the household did not experience a flood disaster; (1) experienced a flood disaster, but the impacts were not severe; or (2) experienced a flood disaster that had severe impacts. We identified households that experienced floods by the survey questions of whether there were any natural disasters in the area where the households resided in the prior 5 years and by the reported types of disasters. The IFLS Natural Disaster module captured both natural disasters and events that cause economic disruption—such as floods, earthquakes, tsunamis, mudslides, civil strife, death of household head or member, serious illness, job loss, etc.—collected from a check-all-that-apply list with 18 possible response options (See Table S1 for a complete listing). “Severity” was predefined by a question within the IFLS Natural Disaster module of whether any disasters were “severe enough to cause death or major injuries of a household member, cause direct financial loss to the household, or cause household members to relocate”. The second independent variable of interest was whether the household was agricultural. To be as broad as possible in our definition of an agricultural household, we constructed an indicator of whether the household had any members engaged in a farm business or activity. See Table S1 for the metrics and corresponding survey questions.

The rural poor may be more vulnerable to negative impacts of flooding according to prior research [27]. We evaluated the relationship to migration by including cofactors collected in the Household Characteristics module in the model. These were urban or rural areas, household size, main water source, type of toilet, sewage, and garbage disposal systems as measures of socio-economic status. Each of these factors may also be associated with a household’s experience with flooding, as they may be determinants of living in areas more prone to flooding.

Cross-tabulations of each categorical independent variable stratified by the ordinal outcome variable showed exceedingly low cell counts (<15), and the model with the six-level ordinal outcome did not converge; therefore, we collapsed it into a five-level category for modelling defined as: (0) the household did not move, or the household moved (1) within the same village, (2) within the same district, (3) within the same regency, (4) outside of the regency or to another province. Similarly, due to low frequencies in the response categories, we collapsed other socio-economic characteristics into fewer categories or into binary variables for modelling as follows: type of toilet to a binary variable indicating owning a septic system, type of sewage draining system to a binary variable indicating flowing drainage ditch, type of garbage disposal system to a three-level categorical variable (collection by sanitation service/burned/or other), and type of drinking water to a four-level categorical variable (pipe/well with pump/well water/or other).

2.3. Data Analysis

All data management and analysis were done using Stata 17 (Stata Statistical Software: College Station, TX, USA: Stata Corp LP). We calculated percentages and 95% confidence intervals of the types of disasters, including floods and related impacts, socio-economic and demographic characteristics for the wave four and wave five cross-sections. To evaluate whether occurrence and severity of flooding influenced migration, we fit a generalized ordered logit and partial proportional odds model as follows [28]:

$$P(Y_i > j) = \frac{\exp(\beta_{0j} + \beta_1 X_{i,u} + \beta_2 Z_{i,v})}{1 + [\exp(\beta_{0j} + \beta_1 X_{i,u} + \beta_2 Z_{i,v})]} \quad (1)$$

where i represents household, and j represents the cumulative logit models for the $M - 1$ levels of the ordinal outcome Y (with M representing the number of levels).

The model produced four sets of coefficients for the four cumulative logit models based on the five-level ordinal outcome, specifically for comparing the odds of (1) not moved versus moved; (2) not moved or moved within the same village versus moved out of the village; (3) not moved or moved within the same district versus moved out of the district; and (4) not moved or moved within the same regency versus moved out of the regency. This is a more parsimonious alternative to the ordered logit model, which relies on the assumption of proportional odds (i.e., the relationship between the independent and dependent variables are the same for each of the cumulative logistic regressions). The partial proportional odds model allows for beta coefficients that are both constrained and unconstrained to the proportional odds assumption. For independent variables that meet the assumption, one coefficient is given. For variables that do not meet the assumption, four sets of coefficients are given. The independent variables, such as household experience with flood, indicator of agricultural livelihood, socio-economic characteristics, indicator of whether the household was in an urban or rural area, household size, and survey wave are denoted by u and v . In the above equation, $\beta_1 X_{i,u}$ represents the coefficients for the independent variables, u , that are the same across the values of j , and $\beta_{2j} Z_{i,v}$ represents the coefficients for the independent variables, v , that are not the same across the values of j . We present odds ratios and interpret them based on direction, with higher odds indicating leaving the locality and lower odds indicating staying within the locality. We determine statistical significance at the $\alpha = 0.05$ level.

We fit univariate models for each independent variable before fitting the full, adjusted model. We used complete case analysis (less than 1% missing for each variable included in the model). We examined the collinearity of independent variables using variance inflation factor (with values of 5 or higher considered problematic), as well as assessing changes in model output between univariate and adjusted models. There was no evidence of multicollinearity, with all VIF < 2. Other disaster variables, including household experience with multiple disasters, were not included in the models due to very little variation (<3%).

2.2.4 Results

3. Results (*manuscript heading*)

Descriptive statistics are presented in Table 1. The population can be characterized as predominantly rural (>55%), with most households having running water and a septic system. About 40% were agricultural households, with at least one household member working as a farmer.

Table 1. Description of Central Java Households.

Household Characteristics	Wave 4, 2007 (n = 1458)		Wave 5, 2014 (n = 1472)	
	%	(n)	%	(n)
Rural	58	(846)	55.2	(813)
Agricultural	39.9	(582)	37	(544)
Main source of drinking/cooking water				
Pipe water	22.4	(327)	27.9	(411)
Well/pump (electric, hand)	27.4	(400)	26.6	(392)
Well water	22.9	(334)	9.9	(145)
Spring water	15.1	(220)	12.2	(179)

Rainwater	0.6	(9)	0.1	(2)
River/Creek water	0.3	(5)	0.5	(8)
Water collection basin	2.2	(32)	0.4	(6)
Aqua/Mineral water	6.9	(101)	15.9	(234)
Other	0.2	(3)	0.9	(13)
Type of toilet/Where household members defecate				
Own toilet with septic	61.4	(895)	69.4	(1021)
Own toilet without septic	8.8	(129)	9	(133)
Shared toilet	6.7	(98)	4.2	(62)
Public toilet	3.2	(47)	3.1	(46)
Creek/river/ditch (without toilet)	11.5	(168)	4.3	(64)
Yard/field (without toilet)	0.6	(9)	0.5	(8)
Pond/fishpond	4.9	(71)	2.6	(39)
Sea/lake	0.8	(11)	0.3	(5)
Other	0.1	(2)	0.9	(13)
Sewage disposal system				
Drainage ditch (flowing)	61.2	(892)	54.4	(801)
Drainage ditch (stagnant)	2.1	(30)	2.6	(39)
Permanent pit	10.2	(149)	12.8	(189)
Disposed into river	7.1	(103)	7.5	(110)
Disposed in yard/garden	7.3	(106)	6.6	(97)
Pond/Fishpond/Lake/Pool	6.8	(99)	5.6	(83)
Hole (without permanent lining)	3.2	(46)	3.8	(56)
Paddy or other field	0.3	(5)	0.5	(8)
Other	0.1	(1)	0.5	(8)
Type of garbage disposal system				
Trash can, collected by sanitation service	23	(335)	24.8	(365)
Burned	37	(540)	44.5	(655)
River/Creek/Sewer	7.7	(112)	8.1	(119)
Disposed in yard and let decompose	14.9	(217)	11.5	(169)
Pit	10.8	(157)	2.8	(41)
Forest, mountain	0.5	(7)	0.1	(1)
Sea, lake, beach	1.2	(18)	1.2	(18)
Paddy or other field	2.4	(35)	0.5	(8)
Other	0.7	(10)	1	(15)
Experience with disasters				
Experienced any disaster in area where household lives in past 5 years	14.4	(210)	19.5	(287)
Flood	5.2	(76)	8.4	(123)
Landslide	2.1	(31)	2.9	(43)
Volcanic	0	(0)	3.8	(56)
Earthquake	4.3	(63)	1.6	(24)
Windstorm	0.8	(11)	2.7	(40)
Fire	0.1	(2)	0.8	(12)
Drought	0	(0)	2.2	(33)

Experienced multiple disasters	0.9	(13)	2.6	(38)
Flood severity				
Did not experience a flood disaster	92.9	(1355)	86.1	(1267)
Experienced a flood disaster with no severe impacts	2.9	(43)	6.9	(101)
Experienced a flood disaster with severe impacts	2.3	(33)	1.5	(22)
Migration outcome				
Extent moved since previous survey				
Household did not move	79.7	(1162)	88.5	(1302)
Moved within same village	1.6	(23)	2.2	(33)
Moved within same district	3.4	(49)	2.1	(31)
Moved within same regency	5.8	(84)	1.8	(26)
Moved within same province	5.1	(75)	2.3	(34)
Moved to another province	4.3	(62)	3	(44)

3.1. Reported Types of Disasters and Their Impacts

Over the time span, 31% (466) of Central Java households reported being affected by a select disaster (Table S1) in the area where they live, with 3% (52) experiencing multiple disasters. Floods were the most common disaster, involving 13% (190) of Central Java households over the period, followed by earthquakes at 6% (92), volcanoes at 4% (58), windstorms at 3% (49), drought at 2% (33), and fire at 1% (17) (Table 1). A total of 7% (104) of Central Java households (22% of disaster-affected households) reported being affected by a disaster severe enough to result in death, significant injuries, financial loss, or relocation. These severely impacted households were mainly affected by flood at 55% (58), followed by an earthquake at 21% (22), windstorms at 13% (13), drought at 10% (10), and volcanoes at 9% (9).

In addition to the types of disasters above, respondents reported the direct impacts of those disasters from a set of questions in the Natural Disaster module (Table S1). Of the 58 households that were severely impacted by floods, 17% (10) had multiple disasters at the same time, mostly failed harvests (5). This subpopulation did not indicate damage or destruction of homes due to flooding, but temporary displacement was standard with 60% (35) reporting some time without housing or in temporary housing. All who spent time without housing indicated they did return or planned to return to their home. Most households (55%) received some government or NGO assistance. Most assistance was from the regional government (56% of households that received assistance). No deaths, injuries, or illnesses due to the flood were reported by these households.

3.2. Household Migration Response

Most households, 72% (1081), did not move during the time span, while 28% (420) moved at least once; 21% (309) moved once and 7% (111) moved in both waves.

Table 2 displays results of the generalized ordered logit/ partial proportional odds model. The model included 1472 households with complete data. The proportional odds assumption held for flood severity, flowing drainage system, and owning a septic system, but was not imposed for agricultural livelihood, type of garbage disposal system, water source, household size, rural area, and survey wave. Compared to households not exposed to floods, those that were exposed to floods with severe impacts had 75% lower odds of moving farther away (OR 0.25, 95% CI 0.09,

0.70) and agricultural households had 42% lower odds (OR 0.58, 95% CI 0.44, 0.77). Drinking water source, household size, and rural area were associated with the extent moved. Compared to households with piped water, there were higher odds of moving farther away among those with pumped well water (OR 1.36, 95% CI 1.01, 1.83) and other water sources including spring, rain, or river (OR 1.93, 95% CI 1.42, 2.61). Rural households had 62% lower odds of moving farther away (OR 0.38, 95% CI 0.29, 0.51). An increase in household size had 7% lower odds of moving farther away (OR 0.93, 95% CI 0.87, 0.99). Compared to the period from 2002 to 2007, the odds of moving farther away from 2009 to 2015 decreased by 56% (OR 0.44, 95% CI 0.35, 0.55).

Table 2. Generalized ordered logit/ partial proportional odds model assessing the extent moved among Central Java households ^{1,2}.

	Odds Ratio	SE	p-Value	Lower Bound	Upper Bound
<i>Main variables of interest</i>					
Flood severity (no experience with flood is ref.)					
Experienced a flood with no severe impacts	0.93	0.21	0.745	0.60	1.44
Experienced a flood with severe impacts	0.25	0.13	0.008	0.09	0.70
Agricultural household	0.58	0.08	0.000	0.44	0.77
<i>Other household socioeconomic factors</i>					
Owns toilet with septic	1.23	0.16	0.096	0.96	1.58
Sewage drains from flowing drainage ditch	0.88	0.10	0.260	0.70	1.10
Garbage disposal (collection by sanitation service is ref.)					
Burned	1.09	0.17	0.581	0.81	1.47
Other (river, yard, pit, etc.)	0.91	0.15	0.558	0.66	1.25
Main source of drinking/cooking water (pipe is ref.)					
Well, pump (electric, hand)	1.36	0.21	0.043	1.01	1.83
Well water	1.32	0.22	0.104	0.95	1.84
Other (spring, rain, river, etc.)	1.93	0.30	0.000	1.42	2.61
Household size	0.94	0.03	0.077	0.89	1.01
Rural	0.38	0.06	0.000	0.29	0.51
Survey wave (wave 5 vs. 4)	0.49	0.05	0.000	0.40	0.60
Constant	0.42	0.10	0.000	0.27	0.66
Variables not constrained to the proportional odds assumption					
Model 2 ²					
Agricultural household	0.62	0.10	0.002	0.45	0.84
Garbage disposal (collection by sanitation service is ref.)					
Burned	0.92	0.16	0.640	0.65	1.30
Main source of drinking/cooking water (pipe is ref.)					

Well, pump (electric, hand)	1.35	0.21	0.055	0.99	1.84
Well water	1.69	0.33	0.008	1.15	2.49
Household size	0.93	0.03	0.028	0.87	0.99
Rural	0.32	0.06	0.000	0.23	0.45
Survey wave (wave 5 vs. 4)	0.44	0.05	0.000	0.35	0.55
Constant	0.43	0.10	0.000	0.27	0.68
Model 3 ²					
Agricultural household	0.50	0.09	0.000	0.36	0.71
Garbage disposal (collection by sanitation service is ref.)					
Burned	0.72	0.13	0.070	0.50	1.03
Main source of drinking/cooking water (pipe is ref.)					
Well, pump (electric, hand)	0.94	0.17	0.735	0.66	1.34
Well water	1.05	0.24	0.822	0.68	1.64
Household size	0.83	0.03	0.000	0.77	0.90
Rural	0.45	0.09	0.000	0.31	0.66
Survey wave (wave 5 vs. 4)	0.25	0.04	0.000	0.19	0.35
Constant	0.70	0.19	0.195	0.41	1.20
Model 4 ²					
Agricultural household	0.62	0.13	0.025	0.41	0.94
Garbage disposal (collection by sanitation service is ref.)					
Burned	0.77	0.16	0.203	0.52	1.15
Main source of drinking/cooking water (pipe is ref.)					
Well, pump (electric, hand)	1.07	0.22	0.727	0.72	1.59
Well water	1.20	0.32	0.506	0.71	2.02
Household size	0.79	0.04	0.000	0.71	0.87
Rural	0.42	0.09	0.000	0.27	0.64
Survey wave (wave 5 vs. 4)	0.35	0.06	0.000	0.25	0.49
Constant	0.41	0.12	0.003	0.23	0.75

¹ The model includes a total of 1472 unique households (2816 observations over two cross-sectional survey waves in 2007 and 2014). ² For variables not constrained to the proportional odds assumption, the model produces four sets of coefficients for the four cumulative logit models for comparing the odds of (1) not moved versus moved; (2) not moved or moved within the same village versus moved out of the village; (3) not moved or moved within the same district versus moved out of the district; and (4) not moved or moved within the same regency versus moved out of the regency.

2.2.5 Discussion

4. Discussion

4.1. Migration Extent

This study assessed how flooding and the severity of impacts effected the extent of household movement in Central Java, Indonesia using data from a household survey. We distinguished between local and long-distance moves by using the household migration extent as our outcome. We quantified flood severity using information about the reported impacts to households. We found households that are agricultural, rural, and those most severely affected by floods had lower odds of moving farther away. In other words, they tended to stay locally.

In a similar study in 2014 by Bohra-Mishra et al., which analyzed farmers using four waves of IFLS data (1993–2007), the findings indicated that precipitation, which can be a precursor to flooding events, had a smaller impact on migration than temperature change [29]. Results from the IFLS are consistent with other Indonesia studies indicating a tendency to prioritize adaptation strategies other than migration when faced with flooding disasters [15,29,30]. In the IFLS, almost all who reported spending some time without housing due to flood indicated they returned. In a 2011 study in Semarang, Indonesia, by Harwitasari et al., the findings indicated a deep connection to dwellings and noted that abandoning their homes was not realistic [30]. The study found that community members, especially homeowners, were willing to pay and coordinate together on adaptation practices to remain on their land. In a 2018 study by Buchori et al. in Semarang, Indonesia, 81% of respondents indicated a preference to stay rather than migrate [15]. The most reported reason for remaining was an emotional attachment to their homes (41%), followed by proximity to work (34%) [15]. For households that did decide to migrate, a significant proportion (44% of the 19% that reported migrating) chose local hillier areas within the Semarang area [15]. Knowledge and policy gaps related to migration in the context of slow-onset impacts of climate change remain [4,5].

Our finding that severely affected households tended not to move as far away may signify a turning point where adaptation and adaptive capacity—including a household’s ability to relocate out of degraded areas—may be diminished. These findings are consistent with other studies and with migration theory that the losses from flooding disasters may make longer distance moves more difficult [15,18,19]. This hypothesis is reinforced by the most recent IPCC report which detailed projected limits to adaptation in coastal and agricultural communities as global warming continues [4]. The losses caused by floods may create “trapped” populations, as households that are most affected become the ones who lack the capital needed to start a new life somewhere else. This phenomenon has been documented in other flood-affected settings [6]. In our study, findings indicated that the most vulnerable were larger, rural, agricultural households.

Floods were the most reported disaster to be severe enough to result in death, major injuries, financial loss, or relocation. Flooding in the Indonesia context can carry a multitude of definitions, all resulting in potential implications for agricultural households. For example, “the Rob”, which refers to the combination of temporary flooding and mostly permanent inundation, is common in the northern coastal region of Java [15]. Additionally, Indonesia is vulnerable to various compounding climate change factors such as sea level rise and variability in intensity and duration of monsoon rains [31]. In a 2011 study by Syaukat analyzing the impact of climate change on food production in Indonesia, the analysis demonstrated significant changes in wet and dry seasons when comparing periods between 1961 and 1990, and 1991 and 2003, 10–20 days later and 10–60

days earlier, respectively [32]. This study found climate change related factors could lead to an estimated deficit of 90 million tons of husked rice by 2050. Considering the variety of climatic factors related to flooding that may impact agriculture in Indonesia, future studies should seek to better understand and differentiate between type and severity of floods. While most studies are specific to either riverine or coastal flooding, the measure used in this study encompassed all types of floods and parsed out severely affected households.

4.2. Usability Aspects of the IFLS

The IFLS was designed for studying various aspects of social and economic well-being of the Indonesian population and is a feasible data source for studying how environmental factors impact households. The self-reported information on natural disasters over two survey waves in 2007 and 2014 with respondent recall spanning 10 years and with direct questioning on the impacts allows users to understand the immediate damages. Beyond the immediate damages, users can use the IFLS to explore how disasters affected other aspects of life through the breadth of information related to livelihoods, expenditures, migration, and health at individual, household, and community levels. The IFLS data are provided in multiple formats in several modules, requiring user processing to reshape and merge selected files. The IFLS maintains consistency of variables throughout survey waves, facilitating cross-sectional or longitudinal data analyses. The IFLS has spatial aggregation at the district-level.

The IFLS allows for identification of flood-affected households directly through self-report. The IFLS most likely captures sudden-onset events because it measures a recall of disasters during the previous five years and does not provide a good measure of recurrent, or regular, flooding and the slow-onset changes. Another similar study also used direct survey responses to measure floods, however, as discussed in Oskorouchi et al., this measure may be biased, as those who were more affected may have been more likely to report a flood [33]. The self-reporting of floods also has limitations in terms of precision. Other studies have used a range of approaches to classify flooding experienced by households, including linking rainfall, disaster, and other spatial data, or using satellite precipitation-driven modeling [15,33–36]. In a previous study of the IFLS, the authors linked IFLS data with the DesInventar database, which provides information on disasters from official sources, academic records, newspapers and institutions [29]. While these data linkages are possible with the IFLS, there is risk of misclassification of flood-affected households, as the smallest administrative boundary is at the district level.

While the IFLS captures all types of flooding, including river and coastal flooding, it did not delineate between types of floods or main cause. The survey tool asks whether households experienced a disaster (flood or other type of disaster), how many times, and when was the most severe event (see Table S1). The information about the frequency of events could have provided another measure of flood but with very low frequencies when subset to Central Java due to a logical skip pattern in the survey. Findings from prior studies indicate that flood severity does impact the types of adaptation strategies chosen, such as migration or household modifications [15]. We classified flood severity based on a survey question of whether the disaster was severe enough to result in death, major injuries, financial loss, or relocation, but these impacts could not be disaggregated; however, this information did allow for identification of the most severely impacted households from which to derive a flood severity metric.

The IFLS had an excellent tracking record of households across survey waves, allowing for an accurate measure of whole household movement. Findings from our study differ from the Internal Displacement Monitoring Centre (IDMC) database, which reports consistently higher displacement figures as a result of sudden-onset disasters, primarily flooding, throughout the IFLS

survey period [17,37]. This is likely because we examined whole-household migration rather than temporary displacement. While the IFLS asked whether any household members spent time without housing due to a disaster, it did not ask directly if households relocated as a result. Further, the direct questioning captures the immediate rather than long-term impacts. Reasons why households stayed or moved were not captured by the IFLS but would enhance understanding of vulnerability and immobility in the context of flooding disasters.

Due to the cross-sectional nature of the data, it could not be determined whether floods preceded household move, leading to potential antecedent-consequent bias. This may have the effect of biasing toward the null, meaning there may be a larger effect than what was detected in this study. The IFLS contained an individual-level module on migration, which was not examined in our study, that is suited towards studying the social and economic drivers. This data could be explored more to understand the various drivers, including slow-onset environmental changes, which a previous paper has examined through data linkages [29].

2.2.6 Conclusions

5. Conclusions (*manuscript heading*)

Flood-affected households in Central Java that suffered losses tended to stay locally compared to those that were unaffected by floods, while flood-affected households that did not suffer losses showed no difference. We found differences in the extent moved in households that were agricultural and rural, which also tended to stay locally. These findings provide evidence that more vulnerable households may be staying within flood-affected areas in Central Java. Immobility in the context of climate hazards should be further researched and better understood. Future surveys should incorporate experiences with small, gradual shifts in the environment and compounding effects of multiple disasters. Disaggregation in data, and evaluating non-linearity in relationships, can allow for deeper understanding of how environmental disasters are impacting lives.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: Survey questions and derived variables. ¹ Not part of the household survey, information was gathered through the survey's tracking of panel households. ² This variable was generated by the IFLS from the household tracking forms and provided to users in the 'htrack' data file.

Author Contributions: Conceptualization, B.K.L.; methodology, B.K.L., R.S., and E.N.N.; software, B.K.L.; formal analysis, B.K.L.; investigation, B.K.L.; data curation, B.K.L. and R.S.; writing—original draft preparation, B.K.L. and O.A.; writing—review and editing, B.K.L., L.B., R.S., O.A., K.B., E.N.N., E.C.d.P., T.G., M.K., and A.I.; supervision, E.N.N., E.C.d.P., T.G., M.K., and A.I.; funding acquisition, E.N.N., T.G., and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Cooperative State Research, Education, and Extension Service Fellowship (grant number 2020-38420-30724, ENN and TG); the National Science Foundation's (NSF) Innovations in Graduate Education (IGE) Program's Solution-Oriented, Student-Initiated, Computationally-Enriched (SOLSTICE) training and learning approach (grant number 1855886, ENN); and the NSF International Research Experiences for Students (IRES) grant: US-Indonesian Research Experience for Students on Sustainable Adaptation

of Coastal Areas to Environmental Change (grant number 1826939, ENN and MK). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of RAND in the United States and the University of Gadjah Mada in Indonesia (s0064-06-01-CR01.).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data reported in this manuscript is publicly available from the RAND corporation website: <https://www.rand.org/well-being/social-and-behavioralpolicy/data/FLS/IFLS.html> (accessed on 27 August 2021). Additionally, the data can be made available upon request by contacting the corresponding author.

Acknowledgments: The authors would like to thank colleagues at the center for Coastal Disaster Mitigation and Rehabilitation Studies, Universitas Diponegoro in Semarang City, Central Java, Indonesia, for their continued guidance and critical feedback on presentations and manuscript drafts. We would like to acknowledge Alex Carroll, Tyler Stotland, and Nilakshi Chatterji for their prior work with Tufts Institute of the Environment which inspired this analysis. Finally, we would like to thank participants of the IFLS for sharing their data which has been an invaluable resource.

Conflicts of Interest: The authors declare no conflicts of interest.

2.3 References

- Schwerdtle, P.; Bowen, K.; McMichael, C. The health impacts of climate-related migration. *BMC Med.* **2017**, *16*, 1.
- McMichael, C.; Barnett, J.; McMichael, A.J. An ill wind? Climate change, migration and health. *Env. Health Perspect* **2012**, *120*, 646–654. <https://doi.org/10.1289/ehp.1104375>.
- COP26—Direct Linkages between Climate Change, Health and Migration must be Tackled Urgently—IOM, WHO, Lancet Migration. Available online: <https://www.who.int/news/item/09-11-2021-cop26---direct-linkages-between-climate-change-health-and-migration-must-be-tackled-urgently-iom-who-lancet-migration> (accessed on 8 August 2022).
- Pörtner, H.-O.; Roberts, D.C.; Adam, H.; Adekan, I.; Adler, C.; Adrian, R.; Aldunce, P.; Ali, E.; Begum, R.A.; Bednar-Friedl, B.; et al. *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Pörtner, H.-O., Roberts, D.C., Poloczanska, E.S., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Eds.; Cambridge University Press: Cambridge, UK, 2022.
- MCauliffe, M.; Traindafyllidou, A. *World Migration Report 2022*; International Organization for Migration (IOM): Geneva, Switzerland, 2021.
- Williams, L.; Arguillas, M.J.B.; Arguillas, F. Major storms, rising tides, and wet feet: Adapting to flood risk in the Philippines. *Int. J. Disaster Risk Reduct.* **2020**, *50*, 101810. <https://doi.org/10.1016/j.ijdr.2020.101810>.
- Climig/The Migration, Climate Change and the Environment Database. Available online: <https://climig.com/> (accessed on 8 August 2022).
- Najibi, N.; Devineni, N. Recent trends in the frequency and duration of global floods. *Earth Syst. Dyn.* **2018**, *9*, 757–783. <https://doi.org/10.5194/ESD-9-757-2018>.
- EM-DAT EM-DAT/The International Disasters Database. Available online: <https://www.emdat.be/> (accessed on 3 November 2019).
- The Flood Observatory. Available online: <https://floodobservatory.colorado.edu/index.html> (accessed on 28 January 2021).
- Marfai, M.A.; King, L.; Sartohadi, J.; Sudrajat, S.; Budiani, S.R.; Yulianto, F. The impact of tidal flooding on a coastal community in Semarang, Indonesia. *Environmentalist* **2008**, *28*, 237–248.
- Marfai, M. Impact of coastal inundation on ecology and agricultural land use case study in central Java, Indonesia. *Quaest. Geogr.* **2011**, *30*, 19–32.

13. Nurjaya, W.; Lumban-Gaol, J.; Tambunan, E.; Osawa, T.; Pasaribu, B. Sea Level Rise Impact on Eastern Coast of North Sumatra, Indonesia. In Proceedings of the 2nd International Forum on Sustainable Future in Asia, Bali, Indonesia, 26–28 January 2017.
14. Demirkesen, A.C.; Evrendilek, F.; Berberoglu, S.; Kilic, S. Coastal Flood Risk Analysis Using Landsat-7 ETM+ Imagery and SRTM DEM: A Case Study of Izmir, Turkey. *Environ. Monit. Assess.* **2007**, *131*, 293–300. <https://doi.org/10.1007/s10661-006-9476-2>.
15. Buchori, I.; Pramitasari, A.; Sugiri, A.; Maryono, M.; Basuki, Y.; Sejati, A.W. Adaptation to coastal flooding and inundation: Mitigations and migration pattern in Semarang City, Indonesia. *Ocean Coast. Manag.* **2018**, *163*, 445–455. <https://doi.org/10.1016/j.ocecoaman.2018.07.017>.
16. Abdullah, F.; Marzuki, M.; Yanis, M.; Ismail, N. Food security strategies toward flood hazards along the Cinendang river, Aceh Province. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *667*, 012040. <https://doi.org/10.1088/1755-1315/667/1/012040>.
17. Internal Displacement Monitoring Centre. *Global Report on Internal Displacement*; Internal Displacement Monitoring Centre: Geneva, Switzerland, 2022;.
18. Black, R.; Bennett, S.R.G.; Thomas, S.M.; Beddington, J.R. Migration as adaptation. *Nature* **2011**, *478*, 447–449. <https://doi.org/10.1038/478477a>.
19. Gray, C.L.; Mueller, V. Natural disasters and population mobility in Bangladesh. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6000–6005. https://doi.org/10.1073/PNAS.1115944109/SUPPL_FILE/SAPP.PDF.
20. Chen, J.J.; Mueller, V.; Jia, Y.; Tseng, S.K.-H. Validating Migration Responses to Flooding Using Satellite and Vital Registration Data. *Am. Econ. Rev.* **2017**, *107*, 441–445. <https://doi.org/10.1257/aer.p20171052>.
21. Benveniste, H.; Oppenheimer, M.; Fleurbaey, M. Climate change increases resource-constrained international immobility. *Nat. Clim. Chang.* **2022**, *12*, 634–641. <https://doi.org/10.1038/s41558-022-01401-w>.
22. Schewel, K. Understanding immobility: Moving beyond the mobility bias in migration studies. *Int. Migr. Rev.* **2020**, *54*, 328–355. <https://doi.org/10.1177/0197918319831952>.
23. Findlay, A.M. Flooding and the scale of migration. *Nat. Clim. Chang.* **2012**, *2*, 401–402. <https://doi.org/10.1038/nclimate1554>.
24. Adams, H. Why populations persist: Mobility, place attachment and climate change. *Popul. Environ.* **2016**, *37*, 429–448. <https://doi.org/10.1007/s11111-015-0246-3>.
25. RAND Indonesian Family Life Survey (IFLS)/RAND. Available online: <https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html> (accessed on 27 August 2021).
26. Strauss, J.; Sikoki, B. *Working Paper User's Guide for the Indonesia Family Life Survey, Wave 5 Volume 2*; 2016. RAND Corporation, Santa Monica, CA, USA.
27. Fanzo, J.; Davis, C.; McLaren, R.; Choufani, J. The effect of climate change across food systems: Implications for nutrition outcomes. *Glob. Food Sec.* **2018**, *18*, 12–19. <https://doi.org/10.1016/J.GFS.2018.06.001>.
28. Williams, R. Understanding and interpreting generalized ordered logit models. *J. Math. Sociol.* **2016**, *40*, 7–20. <https://doi.org/10.1080/0022250X.2015.1112384>.
29. Bohra-Mishra, P.; Oppenheimer, M.; Hsiang, S.M. Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9780–9785. https://doi.org/10.1073/PNAS.1317166111/SUPPL_FILE/PNAS.201317166SI.PDF.
30. Harwitasari, D.; van Ast, J.A.A. Climate change adaptation in practice: people's responses to tidal flooding in Semarang, Indonesia. *J. Flood Risk Manag.* **2011**, *4*, 216–233. <https://doi.org/10.1111/j.1753-318X.2011.01104.x>.
31. Douglas, I. Climate change, flooding and food security in south Asia. *Food Secur.* **2009**, *1*, 127–136.
32. Syaukat, Y. The Impact of Climate Change on Food Production and Security and Its Adaptation Programs In Indonesia. *J. ISSAAS* **2011**, *17*, 40–51.
33. Oskorouchi, H.R.; Sousa-Poza, A. Floods, food security, and coping strategies: Evidence from Afghanistan. *Agric. Econ.* **2021**, *52*, 123–140. <https://doi.org/10.1111/AGEC.12610>.
34. Amare, M.; Jensen, N.D.; Shiferaw, B.; Cissé, J.D. Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agric. Syst.* **2018**, *166*, 79–89. <https://doi.org/10.1016/J.AGSY.2018.07.014>.
35. Oskorouchi, H.R.; Nie, P.; Sousa-Poza, A. The effect of floods on anemia among reproductive age women in Afghanistan. *PLoS ONE* **2018**, *13*, e0191726. <https://doi.org/10.1371/journal.pone.0191726>.
36. Gaire, S.; Delbiso, T.D.; Pandey, S.; Guha-Sapir, D. Impact of disasters on child stunting in Nepal. *Risk Manag. Healthc. Policy* **2016**, *9*, 113. <https://doi.org/10.2147/RMHP.S101124>.
37. Cazabat, C.; O'Connor, A. *Internal Displacement Index Report*; 2021. Internal Displacement Monitoring Centre: Geneva, Switzerland, 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Chapter 3: Usability of flood data portals for food security, nutrition, and health research (*manuscript formatted for Nature Scientific Data*)

Breanne K Langlois^{1*}; Aliya Magnuson¹; Magaly Koch²; Timothy Griffin¹; Erin Coughlan de Perez^{1,4}; Elena Naumova¹

1 Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

2 Center for Remote Sensing, Boston University, Boston, MA, USA

3 Departemen Oseanografi, Universitas Diponegoro, Semarang, IDN

4 Red Cross Red Crescent Climate Centre, 2502 KC The Hague, The Netherlands

*Corresponding author: breanne.langlois@tufts.edu

3.1 Abstract

There are many flood data portals due to recent improvements in flood identification using satellites, providing opportunities to study the human impacts. We compared usability of NASA Worldview (NW), Global Flood Awareness System (GloFAS), Global Flood Monitoring System (GFMS), Global Surface Water Explorer (GSWE), and Dartmouth Flood Observatory (DFO) by reviewing the portals' data accessibility and availability, approaches to flood identification, and data quality. We evaluated their user feasibility through a survey of graduate students. GSWE and GloFAS were rated most favorably overall. Respondents had discrepancies in their opinions on the clarity of the goals and scope and platform accessibility for GFMS, DFO, and NW, and in data and visualization quality for all portals. Historical data and measures of flood recurrence and other characteristics are needed. Portals should incorporate the user-perspective by providing multiple data formats, aggregated by sub-national boundaries, and with mechanisms that delineate incomplete or unreliable data. Flood data portals should include interdisciplinary research as part of their purpose and mission. Their longevity and maintenance should be secured to preserve these important data sources for future research.

Keywords: flood, natural disasters, extreme events, data usability, food security

3.2 Introduction

The frequency and duration of flood events is increasing globally, presenting threats to public health and nutrition through vector borne diseases, water contamination, and agricultural losses^{1–10}. Flooding due to river overflow, coastal seawater intrusion, or flash floods are all impacted by climate factors such as extreme precipitation and sea level rise^{11–13}. Other factors such as urbanization, land use, mitigation infrastructures, and socioeconomic conditions also impact flood risk^{14,15}. With a growing global population exposed, floods are among the most impactful environmental hazards^{6,16,17}. The destructive nature of floods warrants data for situational awareness and disaster planning efforts, and research of the human impacts.

Currently, datasets to study the impacts of environmental changes on food security, nutrition, and health outcomes are lacking. Spatial and survey data linkages – the process of assembling information from these different data sources into one dataset – are needed. Remotely sensed imagery from optical and synthetic-aperture radar sensors are instrumental for this purpose^{13,18,19}. After decades of Earth observing satellite missions, abundant data are now available to the public for flood identification, monitoring, forecasting, and assessment of long-term environmental changes. These data present an opportunity for researchers to study the human impacts of disasters and environmental changes. The challenge, however, is integrating geospatial and population data that traditionally do not speak to each other. There is a heavy user burden for researchers outside the earth sciences to process geospatial data products into formats that can be linked to population surveys (e.g., extracting raster values into polygon attributes and aggregating data by administrative boundaries).

In this paper, we address this data gap by evaluating data quality and usability aspects to enable data linkages for this critical inter-disciplinary research. Other papers have evaluated data usability aspects within public health and disaster relief, but inter-disciplinary perspectives are lacking^{20–22}.

3.2.1 Overview of remote sensing and flood mapping

Thirty meter resolution imaging (once considered high resolution, but currently considered moderate resolution) provided by optical sensors aboard Landsat satellites has been valuable to monitor seasonality and long-term changes in surface water over four decades²³. With the launch of NASA's Terra (1999) and Aqua (2002) satellites, daily tracking of wide areas (250 m resolution) of the Earth's surface water has been made possible by the optical sensor MODIS (moderate-resolution imaging spectroradiometer)^{6,24–26}. Synthetic aperture radar sensors (SAR) such as those aboard Sentinel-1 satellites, operated by the Copernicus program, have the advantage of penetrating cloud cover and provide information on river discharge^{19,27}. Multi-satellite precipitation data products from NASA's Global Precipitation Measurement Mission are used in flood detection models²⁸.

Identifications and monitoring of floods and long-term changes remains challenging, despite availability of data provided by satellites²³. Landsat imagery is limited by a wide temporal coverage (8 days), however, this has been recently resolved by a new product that combines Landsat and Sentinel-2 images to derive daily surface water extent maps with global coverage^{23,29}. While MODIS sensors can monitor large floods daily, it cannot detect floods where there is persistent cloud cover, complex terrain such as urban environments, or smaller flash floods⁶. Sentinel-1's SAR sensors are limited in flood detection capability in the presence of wind or vegetation as the contrast between flooded and non-flooded areas is reduced^{19,27,30}. Due to the long temporal cycle of Sentinel-1 satellites (12 days), SAR flood detection is more accurate for longer duration events rather than flash floods^{19,27,30}. Several indices derived from satellite imagery, such

as NDWI (Normalized Difference Water Index), WI (Water Index), RSWIR (Red and Short Wave Infra-Red), or GSWIR (Green and Short Wave Infra-Red), have been used for flood mapping, but each have limitations (e.g. tendency to classify cloud shadow as water)^{31,32}. NDWI, for example, tends to underestimate inundated areas while GSWIR, in contrast, tends to overestimate³¹.

In recent years many websites and data archives containing geospatial flood data products (i.e. flood data portals) have come into existence, each with their own mission and purpose^{6,23,28,33–37}. These portals use various sources of information from satellite imagery, government, or media sources to provide information or generate usable products about floods. Some portals such as USDA's Global Agricultural & Disaster Assessment System are interactive web map viewers that allow users to visualize floods without any special software^{23,37–39}. Some offer data products in different formats^{24,28,36,37,40,41}. The Dartmouth Flood Observatory, for example, provides a downloadable database of flood events^{6,24,39}. Others, such as the Global Surface Water Explorer, provide high resolution map products of surface water changes²³.

3.2.2 Quantifying flood in studies examining the impact on nutrition and food security

Measurement of flood varies widely within the literature on nutrition and food security. One challenge is measuring flood at the community or household level. Another measurement challenge is distinguishing the type of flood – e.g. river, pluvial, coastal, or multiform floods – which can have different impacts^{42,43}. Many studies that have examined nutrition and food security impacts of floods have focused on single flood disasters, comparing pre/post or flooded and non-flood areas^{44–55}. Some studies have measured flood through self-report using surveys, having the unique advantage of directly identifying flood-affected households but raising the risk of bias due to potential overreporting of flood exposure^{45,56,57}. Others have used direct measures from satellite or other spatial data, aggregated at various levels^{10,58–60}. These studies typically do not differentiate the type of flood or main cause. Few studies in this body of literature have

linked spatial data to household surveys. One recent study examined the effect of flooding on food security across broad regional, country, and all-continent scales in Africa and found both positive and negative impacts of floods and their related meteorological conditions. This study highlighted the important data gaps for assessment of flooding and food security outcomes across different spatial and temporal aggregations⁴³.

Linking spatial and survey data is challenging due to usability aspects. While flood data portals classify whether an area is flooded, with varying approaches, the differing spatial and temporal attributes make it difficult to link this information to surveys which include community, sub-district, or district level boundaries. In contexts where flooding is recurrent and commonplace, such a measure could deepen understanding of the impacts and how to respond. The heterogeneity in the literature makes it difficult to compare studies across contexts to determine, broadly, the human impacts of floods. But also, this heterogeneity makes it difficult to compare across studies within the same context. Currently there is no universal definition or classification system of flood for use in disaster relief efforts or research.

3.2.3 Goals of this paper

The goal of this paper is to compare the usability of NASA Worldview (NW), Global Surface Water Explorer (GSWE), Global Flood Awareness System (GloFAS), Global Flood Monitoring System (GFMS), and Dartmouth Flood Observatory (DFO) for nutrition and public health research. This paper aids researchers wanting to link spatial and survey data to study how natural disasters and environmental changes are impacting human outcomes, e.g., health, economics, migration, food security, or nutrition. Specifically, we review the portals for data accessibility and availability, approaches to flood identification, and data quality. In addition, we evaluated user feasibility aspects of the portals through a survey of graduate students reviewing and rating each of the portals for several criteria.

3.2.4 Flood data portals included in this review

All flood data portals included in this review are publicly accessible. This is not an exhaustive list or systematic review of all flood monitoring or flood data portals in existence, of which there are many. We selected five commonly used flood data portals that had global coverage based on a literature review and exploratory search for data in consultation with an expert in the application of Remote Sensing and Geographic Information Systems. Our criteria were that the data were freely and publicly available. We included:

- **NASA Worldview (NW):** NW is a large satellite data archive with both an interactive, dynamic (spatiotemporal) map viewer and data repository offering download of multiple spatial data layers. A few of the many data products it offers are related to floods or surface water extent³⁶. Hosted by the United States government's National Aeronautics and Space Administration (NASA) and built by the Earth Science Data and Information System Project, NW includes satellite derived data from all aspects of the earth system and aims to support monitoring of natural hazards by providing access to current, daily satellite imagery data³⁶.
- **Global Surface Water Explorer (GSWE):** GSWE is an interactive, dynamic map viewer allowing visualization of long-term changes in the Earth's surface water over 37 years^{23,41}. GSWE displays datasets comprising the Global Surface Water collection (since 1984) produced using Landsat satellite imagery. While these data are continuously being updated with no end date, the data products are made available periodically. Currently, the 1984-2021 dataset is available. Data are accessible via several download options provided through a separate data access page. The European Commission's Joint Research Centre developed this portal and corresponding datasets under the Copernicus Program to support a range of applications such as water resource management, conservation, and food security⁴¹. While not specific to flood

events, this portal provides information on the extent and change of surface water including development of new permanent water bodies, e.g. from seasonal to permanent inundated areas.

- **Global Flood Awareness System (GloFAS):** GloFAS provides an interactive map viewer to visualize current flood forecasts and detailed information about how to access the underlying data^{34,40}. Data are derived from satellite imagery and are accessible through the Copernicus Climate Data Store and can also be downloaded through the map viewer by creating an account. Implemented by the European Commission and Copernicus Program, the Global Flood Awareness System aims to support flood disaster preparedness efforts through global flood monitoring and forecasting of large river basins⁴⁰. This is a component of the Copernicus Emergency Management Service.
- **Global Flood Monitoring System (GFMS):** GFMS is an experimental, quasi-global flood monitoring system using real-time, multi-satellite precipitation data to create a gridded flood model that users can view on 1/8th degree resolution maps^{28,61}. GFMS uses TRMM multi-satellite precipitation analysis and global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG) to create the flood model which is based on land surface and river tracing routing models²⁸. The maps allow users to change the different parameters and view time series of regional areas. The portal provides several variables that can be viewed, but the only data available for download is flood intensity (in depth) above threshold. This portal is developed and managed by the University of Maryland and funded by NASA.
- **Dartmouth Flood Observatory (DFO):** DFO is a resource of information about large flood events worldwide, provided through a range of dynamic maps and visualizations and a downloadable historical archive developed from media reports and other sources²⁴. DFO originated at Dartmouth College in 1993 and then transferred to the University of Colorado in 2010 where it is hosted currently. DFO is often the first to map major flood events. Its mission is to

support research and other humanitarian and water management applications.

3.3 Results

3.3.1 Usability characteristics

Table 1 shows the suite of flood-related data products offered by the flood data portals. A comparative summary of the usability characteristics across the flood data portals is presented in **Table 2**.

3.3.1a Data accessibility and availability

All five portals in this review are in the public domain and free to users. All are currently maintained but it was unclear whether they receive long-term funding. The aims and missions varied and included flood monitoring (NW, GloFAS), flood forecasting (GloFAS), disaster preparedness (GloFAS), humanitarian (DFO), water-management decision-making and applications (DFO, GSWE). Only DFO specified a research purpose. The GFMS is an experimental system and did not specify an aim or mission.

All portals make their data products accessible directly from the site but NW and GloFAS require login. The portals offer 18 different flood-related data products (**Table 1**). Most provide information about flood occurrences from satellite imagery. For example, the MODIS NRT Global Flood Product from NW is a daily global map of flooding derived from Terra and Aqua satellite imagery. The collection of Global Surface Water datasets from GSWE capture changes in surface water occurrence at 30m resolution (previously considered high, but currently considered moderate resolution) from Landsat imagery since 1984. While not specifically measuring flood events, GSWE datasets can be used to identify flooded areas. GloFAS data products measure ongoing

and upcoming river floods from Sentinel-1 imagery. While the other portals offer multiple data products, GFMS offers only 1, flood intensity (in depth), derived from multi-satellite precipitation data. DFO data provide information on occurrence of large flood events, but the database is compiled mostly from media sources and not from satellite imagery.

In addition to data on flood occurrence, NW and GSWE capture recurrence and NW, GloFAS, GFMS, and DFO provide data on other characteristics of floods (**Table 1**). The Flood Hazard: Frequency and Distribution gridded data from NW measures frequency of floods between 1985-2003 using DFO data. The Recurrence dataset from GSWE measures inter-annual frequency of surface water between 1984 to present time, capturing areas that were episodically flooded. NW offers Mortality Risk and Economic Risk Flood Hazard layers, providing information on mortality and economic losses. Both are derived from the International Emergency Events Database (EM-DAT). GloFAS flood monitoring has data layers of the affected population and affected land use. GFMS provides data on flood depth. DFO provides the most information about floods in its Global Archive of flood events including main cause, area affected, detailed locations, names of rivers, duration, known number of deaths and displaced, and damages in USD. None of the portals' data products distinguish or classify the type of flood, but DFO does provide the main cause and a measure of flood severity. GloFAS data only includes river floods.

Most of the data products are geospatial, provided in raster or vector data formats. Only DFO provides a non-geospatial format – an excel sheet of flood events.

Table 1. Summary of flood-related data products

Portal	Data Product	Timespan	Temporal aggregation	Spatial aggregation	Format	Data Source	Description of variables
NW	MODIS NRT Global Flood Product	2022-present	Daily ¹	Global, ~232-m pixel res.	HDF-EOS or GeoTIF F ²	Satellite (Terra & Aqua)	Flood occurrence

GSWE	Flood Hazard: Frequency and Distribution	1985-2003	Total timespan	Global 2.5-minute grid	ASCII	DFO archive	Frequency of floods
	Flood Hazard: Mortality Risk	1985-2003	Total timespan	Global 2.5-minute grid	ASCII	Emergency Events Database (EM-DAT)	Mortality loss
	Flood Hazard: Economic Risk	1985-2003	Total timespan	Global 2.5-minute grid	ASCII	EM-DAT	Economic loss as proportion of Gross Domestic Product
	Water Occurrence	1984-2021	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Frequency with which surface water was present
	Change Intensity	1984-2021	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Change in surface water occurrence as increase, decrease or no change
	Water Seasonality	2020	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Intra-annual behavior of surface water
	Water Recurrence	1984-2021	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Frequency with which water returns from year to year
	Water Transitions	1984-2021	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Change in seasonality between the first and last years
	Maximum Water Extent	1984-2021	Total timespan	Global, 30-m pixel res.	GeoTIF F ²	Satellite (Landsat)	Flag indicating if water was ever

							detected or not
GloFA S ³	Forecasts	2011-present	Daily	Global 20-m pixel res.	Grib or NetCDF	Satellite (Sentinel-1)	River discharge
	Seasonal forecasts	2017-present	Monthly	Global 20-m pixel res.	Grib or NetCDF	Satellite (Sentinel-1)	River discharge
	Global Flood Monitoring ⁴	2021-present	Near Real Time	Global 20-m pixel res.	GeoTIF or shapefile	Satellite (Sentinel-1)	Flood extent, number of people affected, affected land use
GFMS	Flood Detection / Intensity (in depth)	2001-2022	Daily	Quasi-global (50°S to 50°N) 1/8th degree res.	Binary files	Multi-satellite precipitation data from NASA's Global Precipitation Measurement Mission News, government, & remote sensing sources from NASA, the Japanese Space Agency, & the European Space Agency	Depth above threshold
	Global Active Archive of large flood events	1985-present	Date of flood	Global, flood extent	Excel or shapefile		Flood location, river names, duration, #dead, #displaced, damages in USD, main cause, severity
DFO	River and Reservoir Watch Version 4.5	1998 - present	Daily	Global	Tabular	Satellite microwave radiometry	Present status and rate of change using River Discharge & Reservoir Area measurement

Master Index of Rapid Response Inundatio n Maps ⁵	2000 - 2008	Date of flood	Global, flood extent	Tabular	ts, flood magnitude Glide number, flood locations, river names, duration
¹ The Global Flood Product is based on Near Real Time satellite imagery and is provided daily for 2 compositing periods (2-day and 3-day). ² Downloaded as 10x10 degree tiles. ³ GloFAS does not provide information on flash flood risk, coastal flooding, or inundated areas. Only focuses on rivers. ⁴ The Global Flood Monitoring product provides 11 different output layers accessible via the GloFAS map viewer. ⁵ Not currently being updated (last update was 2008).					

3.3.1b Approaches to flood identification and their potential errors

The portals had different approaches to flood identification and minimizing type 1 (false positive) or type 2 (false negative) errors by: compositing (NW), ensemble of different flood detection algorithms (GloFAS), classification over an extended time (GSWE), or combining land surface and river runoff models (GFMS).

NW classifies floods via the MODIS Near Real Time Global Flood Product and the Flood Hazard: Frequency and Distribution layer. The Global Flood Product is derived from the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor datasets from both Terra and Aqua satellites. Flood identification is based on time compositing water detections over 3 periods (1, 2, and 3-day) and differentiating them from surface water using a reference mask⁶². If a water detection does not match the reference layer, it is classified as flood. The compositing approach minimizes false flood detections by combining multiple observations, thus reducing error from clouds. For each period, all observed water detections are combined, and pixels are classified as water if they exceed thresholds of 1, 2, or 3 observations based on the periods, respectively. While this approach minimizes error, the Global Flood Product still has limitations and false detections cannot be entirely removed, especially if cloud cover is persistent over multiple observations⁶². False-negative detections may result from obscuration from buildings, canopy, or clouds, while false-positives may still result from cloud or terrain

shadow or dark rock or soil. Users are encouraged to compare flood maps to reflectance imagery to determine potential sources of false-positives or false-negatives⁶². Another potential source of error is the reference mask used to differentiate surface water from floods, which is outdated and planned for improvement⁶².

The Flood Hazard: Frequency and Distribution layer from NW is derived from DFO data (discussed further below).

GSWE uses high resolution Landsat satellite imagery to classify changes in surface water from 1984-2021²³. This portal differs from the others in its goal to detect and classify surface water, not flood events, over multiple decades. Challenges include variations in measurement across the Landsat sensors (TM, ETM+, OLI) and in the conditions in which they're observed²³. False water detection may be caused by shadows (e.g. from buildings, terrain, or cloud), or lava flow. GSWE applies a combination of expert systems, visual analytics, and evidential reasoning techniques to classify pixels and prevent water classification errors²³. This is followed by manual visual inspection and cleaning of the maps to remove false detections, and further validation to handle remaining errors²³. An important feature of GSWE data is its classification over an extended period. Because classifications are made over time, the temporal behavior of the pixels is also used in determining their likelihood of being water, thus strengthening the accuracy of the classification.

GloFAS captures river flooding using Sentinel-1 SAR satellite data and a combination of 3 separate flood detection algorithms⁶³. A final ensemble algorithm generates the final flood product based on a consensus of the 3 flood detection algorithms⁶³. Like NW, a reference water mask is used to differentiate flooding from permanent surface water⁶³. The ensemble approach helps prevent flood classification errors attributed to any one of the flood detection algorithms. However, classification errors are possible due to limitations inherent with SAR data. False positives can occur from land characteristics such as very dry soil, frozen ground, wet snow, or impervious surfaces or from radar shadowing⁶³. False negatives can occur from urbanization or dense

vegetation or strong, heavy winds and rainfall⁶³. To address these classification challenges, GloFAS produces exclusion mask (for effects due to ground surface characteristics) and advisory flag (for effects due to weather conditions) layers along with their flood product layers⁶³. Users can apply these layers to determine which pixels to disregard.

GFMS uses a real-time multi-satellite precipitation data product from NASA's Global Precipitation Measurement Mission to identify floods using a dual land surface with river runoff model – the Dominant River Tracing-Routing Integrated with Variable Infiltration Capacity Environment (DRIVE) model²⁸. Flood is defined when the routed runoff is greater than the flood threshold of that grid cell. Precipitation biases and the size and duration of flood events can impact GFMS flood detection capability²⁸. False positives can occur in areas with dams or related to winter surface features in higher latitudes. False negatives can occur due to underestimation of orographic rainfall (due to hills or mountains). Generally, GFMS has improved flood detection performance with longer, larger floods.

Unlike the other portals, the DFO historical archive is based mainly on media sources (news reports, governmental and international relief agency websites). DFO captures large flood events spanning large geographic extents which are mostly hand drawn based on where the media reported flooding. In some instances, the archive might also have satellite observations to determine the flood extents. While it is a good source for identifying where large flood events occur globally, it is not useful for identifying which specific locations or communities are affected, making it prone to type 1 error. The differences in news coverage by country and region may impact the accuracy of information about a particular flood event.

3.3.1c Data Quality

We assessed data quality through *alignment*, *velocity*, and *variety* (**Table 3**). All portals provide data products structured by spatiotemporal attributes, but only DFO provides a dataset of flood events. All provide daily or Near Real Time data except for GSWE which classifies the whole timespan. DFO is the most important source of historical data about flood events from the 1980s. NW and GSWE also provide historical data products from the 1980s, but the products available through NW are derived from DFO data and EM-DAT. While GSWE has historical data, it captures changes over time and is not specific to flood events. GSWE and GloFAS provide the highest resolution data products, but only DFO provides detailed locations including names of rivers and towns.

All portals update their data products regularly. GFMS, NW, GloFAS, GFMS, and DFO (River and Reservoir Watch), provide daily updates. DFO may have some poor or missing data in the 1990s. The flood data products provided by NW cover the 1985-2003 and January 2022-present timespans, but there is a gap between 2003 to January 2022.

Table 2. Usability characteristics of flood-monitoring data portals

		DFO	GSWE	GloFAS	NW	GFMS
Accessibility	Host organization	University of Colorado ¹	European Commission's Joint Research Centre, Copernicus Programme	European Commission's Joint Research Centre, Copernicus Emergency Management Service	NASA	University of Maryland
	Intended for research	✓				
	Intended for flood monitoring, forecasting, or humanitarian efforts	✓		✓		✓
	Any other purpose		✓ ²		✓	
	Data can be downloaded directly from portal	✓	✓	✓	✓	✓
	Data access requires user login			✓	✓	
	Has a user guide or tutorial		✓	✓	✓	✓
	Portal is currently being maintained	✓	✓	✓	✓	✓
Availability	Number of flood related data products	3	6	3	4	1
	Spatial data product in vector format	✓		✓		
	Spatial data product in raster format		✓	✓	✓	
	Spatial data in other format (ASCII, HDF-EOS, binary)				✓	✓
	Commonly used dataset format (e.g. excel, Stata, SAS, R)	✓				
	Requires further processing for use with statistical software		✓	✓	✓	✓
Data sources	Satellite imagery	✓	✓	✓	✓	✓ ³
	Non-satellite sources (media, government, other)	✓			✓ ⁵	
Alignment	Provides a dataset structured by flood event	✓				
	Provides a spatiotemporal dataset		✓	✓	✓	✓
	Provides historical data (1980s to current)	✓	✓		✓	
	Provides detailed locations (towns, rivers, etc.)	✓				
Variety	Classifies type of flood		N/A			

	Includes river floods	✓	N/A	✓		✓
	Includes coastal/tidal floods	✓	N/A			
	Classifies main cause of flood	✓	N/A			
	Classifies flood severity	✓	N/A			
	Classifies or measures flood frequency/recurrence		✓		✓	
	Other flood information (e.g. duration, depth, damages, etc.)	✓	N/A	✓	✓	✓
Volume	Number of variables	15	6	3	2	1
	Size/extent	Global	Global	Global	Global	Quasi-global
Velocity	Data products are updated regularly	✓	✓	✓	✓	✓
Completeness	Documents incomplete or missing data			✓	✓	
	Gaps between time points of collected data	✓ ⁶			✓ ⁷	
Visualizations	Has dynamic / timelapse maps	✓	✓		✓	✓
	Has interactive web-map viewer/tool		✓	✓	✓	✓

¹ Dartmouth Flood Observatory originated at Dartmouth College in New Hampshire, USA, in 1993 and moved to University of Colorado in 2010. ² Purpose is to quantify changes in global surface water over time to support water management decision making. ³ Uses multi-satellite precipitation data. ⁴ The Flood Hazard: Frequency and Distribution data product is derived from Dartmouth Flood Observatory data. The Flood Hazard: Mortality Risk and Flood Hazard: Economic Risk are derived from the Emergency Events Database (EM-DAT). ⁵ Derived from media reports and include hand drawn areas of where the media reported flooding. ⁶ Poor or missing data in the early/mid 1990s. ⁷ Does not provide data between 2003-2022.

3.3.2 External user assessment

We present an exploratory descriptive and qualitative assessment of graduate student raters from an external user assessment survey. Eight raters participated in the survey, with 6 completing ratings for all 5 portals. **Figure 1** shows the percent agreement. According to raters' opinions, GSWE and GloFAS were ranked most favorably for their usability on average (73% average agreement), while DFO and NW were ranked least favorably (39% and 37%, respectively) and GFMS was in the middle (58%). **Figure 1** shows the breakdown by the evaluation metrics and areas where the portals can improve their usability. For goals and scope, raters agreed for GSWE, GloFAS, and GFMS, but did not seem to agree for DFO and NW. Ratings for data and visualization quality were mixed across the portals, with raters agreeing on different aspects for each portal. For platform accessibility, raters agreed for GSWE, but less so for the other portals. For contacts, communication, and conflicts of interest, raters agreed for GFMS, GloFAS, and GSWE and did not seem to agree for DFO and NW. Inter-rater reliability for each of the portals ranged from slight to fair. When examined across all 5 rating categories and across 2 rating categories (either agree or disagree), respectively, Kappa coefficients were 0.06 and 0.40 for GSWE, 0.03 and 0.16 for GloFAS, 0.25 and 0.40 for GFMS, 0.05 and 0.08 for DFO, and 0.08 and 0.11 for NW.

Figure 1. Percent agreement among raters of external user assessment of flood portals



Notes provided useful feedback about the portals. For DFO, comments indicated the difficulty of navigating the website. E.g.,

“The website is not intuitive and difficult to navigate.”

“There is a wealth of useful data on this portal, however, the site is not user-friendly.”

For GFMS, the ease of navigating the portal was commonly described. E.g.,

“This website was easy to navigate.”

“This portal is well organized and provides easy to-use instructions.”

But improvements could be made to the visualization tools, e.g.:

“The visual presentation and design leaves somewhat to be desired - there are missing spaces and grammatical errors, and the formatting changes throughout the webpage.”

“Some data links were dead, and some visualization options resulted in blank white areas without any stated reason as to why data could not be produced.”

Comments about GloFAS were positive and indicated that it was clear and user-friendly, e.g.:

“This website is the most clearly laid out and the outputs are all described well.”

Comments about GSWE were mixed. Comments indicated the site was well-designed. One rater felt the data download process was difficult, while another commented on the difficulty of the graphs:

“You are able to explore the data but downloading it is a bit more complicated. The website is well designed though and easy to use!”

“The tool to explore global surface water was impressive... however, either I was not able to properly understand the graphs, or the graphs weren't working properly.”

Finally, comments about NW were focused on difficulty of accessing and downloading flood-specific data, e.g.:

“While this tool is visually impressive, I couldn't even find a way to view flooding or surface water after several minutes of exploration.”

3.4 Discussion

We reviewed 5 flood monitoring portals for their usability for public health and nutrition research applications, addressing an important data gap in interdisciplinary research of how climate events are impacting human outcomes. We provide a detailed comparison of different flood data portals which is useful to researchers and host organizations seeking to improve usability of these important and valuable data products. We did not evaluate the correctness of the flood detection algorithms in our assessment, as it depends largely on the intended use. We found considerable heterogeneity in the data products and information provided, and identified areas where these portals could improve. Overall, these portals should include research within their scope and provide products that facilitate interdisciplinary research.

There are several case examples of how data from these portals have been used in prior research. DFO and GloFAS have been used more frequently than the other portals. DFO has had a broader range of applications in the literature including spatiotemporal variability of floods, impacts on mortality, displacement, and economy, while GloFAS has been used mainly for early warning systems^{64–72}. DFO has also been used in validation of an early flood warning system⁷³. A study from 2012 used DFO to examine human determinants of floods⁷⁴. Several studies have linked DFO to other datasets^{69,71,74}. It appears GFMS, NW, and GSWE have been used less frequently for research. In one study, GFMS was used to examine the accuracy of flood monitoring systems⁷⁵. Another study used GFMS to model surface water elevation to predict floods and identify risk zones in Yamuna river, India⁷⁶. NW has been used for studying the impact of palm oil expansion and wildfires on land use, but less so for floods^{77,78}. In one study, NW was used to analyze forest fire risk and flooding as an after-effect, demonstrating the use of this data in GIS (Geographic Information System)⁷⁹. Few studies seem to use GSWE, and these have focused on changes in delta morphology and coastal morphology over time due to human activity^{80,81}. One study examining opinions about the state of rivers in India used GSWE data as a backdrop to the information revealed by the survey⁸². However, we are not aware of any studies using GSWE variables to measure flooding. These case examples point to the need for improvements to usability so these

important data products can be used more efficiently in longitudinal studies to examine human impacts of flooding.

Historical data products that span several decades are needed for longitudinal studies examining impacts of flooding on health, nutrition, or other population outcomes. A few of the portals provided historical data products, with DFO providing an important historical archive of floods since the 1980s. Historical data from NW was derived from DFO data and from EM-DAT but is not ongoing. GSWE also provided important historical data taken from satellite imagery, but it captured the overall change over time. Data products that capture flood events over a long period of time are valuable for population studies, though deriving such products retrospectively is difficult due to quality of data from earlier decades.

Most data captures flood occurrences, but measures of flood recurrence are needed. Only GSWE and NW provided this type of information. GSWE provides a 'recurrence' layer, which captures information about the inter-annual behavior showing areas that are inundated regularly versus those that are flooded episodically⁴¹. NW captures flood recurrence through its 'Flood Hazard: Frequency and Distribution' layer, providing information about flood frequency using data from the DFO historical archive³⁶. Additionally, a planned future improvement to the NW Global Flood Product is to include a 'recurring flood' layer to identify regularly flooded areas⁶². Incorporating other information about floods is also important, such as size, duration, depth, or damages. Only DFO and NW provided other information about floods, including the human and economic impacts. Other classifications of flood exist that incorporate potential damages^{13,24,83}. For example, the US National Weather Service uses a 3-level classification system based on the level of damage caused¹³. If flood data portals can extend beyond flood monitoring, to incorporating these other elements, the usability of their products will improve greatly.

Spatial and survey data linkages are challenging and not intuitive for researchers in public health and nutrition related fields. All the data products in this review require

extra steps to be linked to surveys. None provide aggregations by administrative boundaries, which could facilitate this process. Having multiple data formats available that are accessible to a broad range of researchers could also facilitate this process. High resolution data products are important for data linkages, as aggregating up to community, sub-district, or district levels is more advantageous than downscaling coarse data.

Survey respondents highlighted areas where the portals could improve or provide better clarification. DFO and NW could improve the clarity of their goals and scope. All portals could improve data quality aspects, especially classifying flood types, providing more information about floods, and measuring recurrence. Sub-national and sub-yearly classifications are also needed to facilitate data linkages. GSWE was the most accessible platform, with raters agreeing on some aspects for GloFAS and improvements to user-friendliness, outdated links, tutorials, and data accessibility and formats could be made to the other portals. Taken together, this external user review can help these portals determine where they could improve either in terms of their products or the clarity of their websites.

The low inter-rater agreement could reflect either the variability of the raters or of the portals themselves. The raters in this review were graduate students who were familiar with spatial and survey data, although some had more expertise than others. The low agreement was more likely due to the variability of the portals, indicating a need for a common system of evaluation to improve their usability. Some evaluation criteria may have been easier to judge than others. Regardless, the percent agreement still provides valid and important information.

We did not conduct an extensive review of all flood or disaster data portals in existence, of which there are many. We highlighted major aspects of 5 common flood data portals that we identified. However, it is possible that other flood data portals not included in this review have different usability aspects. While not specific to other flood data portals, our findings may provide general guidance on how to improve these

important data sources. The portals included in this review – NW, GloFAS, GFMS, GSWE, DFO – may take specific recommendations from these findings.

Flood data portals should include research as part of their purpose and mission and should incorporate the interdisciplinary user perspective. We have shown how an evaluation system that incorporates the user perspective could be created. Usability of the data products could be improved by providing sub-national aggregations, different data formats such as statistical datasets in Stata or R format, and mechanisms that clearly indicate where data may be incomplete or unreliable. Historical data products that span decades can provide an opportunity for longitudinal research of the human impacts of climate changes and extreme events. Measures of flood recurrence, not just occurrence, are needed. Data on other characteristics of flooding such as type of flood, depth, duration, and damages are also needed. Finally, the longevity and maintenance of these data portals should be secured through long-term funding for future generations and research.

3.5 Methods

This review was conducted in two parts. First, we gathered detailed information about the portals' data usability characteristics (section 3.5.1 below). Secondly, we conducted a survey evaluating feasibility aspects of the portals from an external user perspective (section 3.5.2 below) using metrics adapted from a previous study⁸⁴. This framework used principles of evidence, efficiency, emphasis, and ethics (**Table 4**). Under the evidence principle are metrics assessing the quality of a portal's data products and whether the data is presented in line with the goals and objectives. Under the efficiency principle are metrics assessing the content organization and visual properties of a portal. Under the emphasis principle are metrics assessing a portal's "user-friendliness" and ability to recognize the needs of a target audience and tailor the content to their perceptions, abilities, and interests. Finally, the ethics principal seeks to evaluate whether best practices are applied.

3.5.1 Usability characteristics

From February to March 2022, we explored the flood data portals in depth and completed a structured form containing the usability characteristics defined in **Table 3** (see Appendix for data gathering form). The forms were completed with as much detail as possible that could be obtained from the websites and associated articles. We assessed data availability, approaches to flood identification through the data sources and flood detection methods, and data quality through alignment, velocity, and variety (**Table 3**). We summarized the information in a table.

Table 3. Usability characteristics included in review of flood data portals.

Aim	Characteristic	Description
Data availability	Accessibility	Refers to the aim and mission of the portal and the ease of access of their data products
	Availability	Refers to availability of data products and type of information about floods
Approach to flood identification	Data sources	The sources of data used to create products
	Algorithms/models	Refers to how flood is identified and classified from the raw data
Data quality	Alignment	Refers to how the data products are structured with respect to time and space, the time span, levels of spatial and temporal aggregation, and how well the data allow linkages to other datasets
	Velocity	Refers to how the data are collected and processed, and the frequency at which data products are updated, collected, reported, and made available
	Variety	Refers to the types of data products, formats available for download, and variable types

3.5.2 External user assessment

From May to September 2022, we recruited graduate students from Tufts University and Boston University to review each portal and complete a webform survey (see Appendix). The survey captured participants' opinions about the portals' goals and scope, data and visualization quality, and platform accessibility. For each portal, participants were asked to spend no more than 20 minutes exploring the portal and rating the evaluation metrics using 5-point Likert responses. The survey was created and conducted using the Kobo Toolbox platform. Ethical approval for this portion of the study was obtained from the Tufts Social, Behavioral & Educational Research Institutional Review Board.

We used Stata 17 (StataCorp LLC, College Station, TX) for analysis of the external user assessment survey. Survey responses were recorded using a 5-point Likert scale defined as: Strongly Disagree, Disagree, Agree, Strongly Agree, Not clear/ Difficult to assess. For each of the 25 evaluation metrics in the survey, we calculated percent agreement (i.e. rated either 'Strongly Agree' or 'Agree') for each portal. Then, we assessed interrater agreement using a generalization of weighted Kappa allowing for the situation of more than 2 raters, more than 2 ratings, and an incomplete design⁸¹. We calculated the Kappa statistic for each flood data portal and interpreted the level of agreement as 0 to .20 = slight, .21 to .40 = fair, .41 to .60 = moderate, .61 to .80 = substantial, and .81 to 1 = perfect. Survey respondents were also asked to leave notes about each of the portals they reviewed. We summarized common themes from these notes inductively.

Table 4. Evaluation metrics used in external user survey of flood data portals.¹

Principle	Indicator	Sub-indicator	Metric
Evidence	Goals & Scope		Overall goals are clearly stated.
			The available data products are clearly listed with a sufficient description.
			Overall, it was quick and easy to understand what data products are offered.
	Data Quality	Integrity	The data sources and metadata are clearly described.

		Standardization	The level of available geospatial aggregations is clear: e.g. flood extent, postal code, village/city/town, state/province/region, and country.
			The level of available temporal aggregations is clear: e.g. day, week, month, quarter, year, etc.
			Historical data are provided (1980s to present).
		Granularity	The data can be classified into sub-national administrative boundaries (e.g. province, district, or village).
			The data can be classified into sub-yearly temporal units (e.g. month, week, or day).
		Definition of flood	A definition of flood or classification of how a flooded area was identified is clearly listed.
			A definition or classification of flood exposure is provided (e.g. how a person, household, or community could be classified as being exposed to flood across space and time).
			The available data classifies type of flood (e.g. river, tidal, or flash flood).
			The available data classifies or provides a measure of flood recurrence for a given spatio-temporal aggregation.
			The available data includes other information about flood (such as duration, depth, damages)
		Completeness	If data is not available for a given variable, reasons are clearly stated.
Efficiency	Visualization Quality	Readability	The portal provides adequate and effective visualizations to demonstrate the data distribution, time trend, quality, quantity and relationship of the data.
			The visualizations clearly communicate the data.
Emphasis	Platform Accessibility		The portal is user-friendly and allows users to easily navigate and explore data and visualizations.
			The portal includes a tutorial or user guide (e.g., informative pop-ups, a tutorial video).
			The data are easily accessible.
			Multiple data formats are available to facilitate a range of uses in different programs (statistical packages and spatial).

		Links are updated and provide updated information.
	Contacts & Communication	The portal includes contact information for questions, suggestions and feedback.
Ethics	Conflicts of Interest	The host organization and purpose of providing the data is clearly stated.
		Funding sources and roles of funders are clearly stated.

¹ Adapted from B. Zhou et al., "Food and Nutrition Systems Dashboards: A Systematic Review," Adv. Nutr., vol. 10, no. Supplement_4, pp. S308–S319, 2019, doi: 10.1093/ADVANCES/NMAC022.

3.6 Code Availability

No code was used in the preparation of this manuscript.

3.7 References

1. *The Human Cost of Weather-Related Disasters 1995-2015*. (2015).
2. Food and Agriculture Organization of the United Nations. Agriculture on the Proving Grounds: Damage and loss. <https://www.fao.org/resources/digital-reports/disasters-in-agriculture/en/>.
3. Salm, L., Nisbett, N., Cramer, L., Gillespie, S. & Thornton, P. How climate change interacts with inequity to affect nutrition. *Wiley Interdiscip. Rev. Clim. Chang.* **12**, e696 (2021).
4. Food and Agriculture Organization of the United Nations. *The impact of disasters and crises on agriculture and food security: 2021*. (2021).

5. Najibi, N. & Devineni, N. Recent trends in the frequency and duration of global floods. *Earth Syst. Dyn.* **9**, 757–783 (2018).
6. Tellman, B. *et al.* Satellite imaging reveals increased proportion of population exposed to floods. *Nat.* **2021 5967870 596**, 80–86 (2021).
7. Alfieri, L. *et al.* Global projections of river flood risk in a warmer world. *Earth's Futur.* **5**, 171–182 (2017).
8. Muh Aris Marfai, B. GIS Modelling of River and Tidal Flood Hazards in a Waterfront City. (International Institute for Geo-Information Science and Earth Observation, 2003).
9. Widiarto, L. A. & Kingma, N. Agricultural Loss Caused by 2007 Sidoharjo's Flood and its House-hold Impact. *Indones. J. Geogr.* **46**, 156 (2014).
10. Pacetti, T., Caporali, E. & Rulli, M. C. Floods and food security: A method to estimate the effect of inundation on crops availability. *Adv. Water Resour.* **110**, 494–504 (2017).
11. Brunner, M. I. *et al.* An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Commun. Earth Environ.* **2021 21 2**, 1–11 (2021).
12. Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A. & Maher, N. More extreme precipitation in the world's dry and wet regions. *Nat. Clim. Chang.* **2016 6**, 508–513 (2016).
13. Klemas, V. Remote Sensing of Floods and Flood-Prone Areas: An Overview. *J. Coast. Res.* **314**, 1005–1013 (2015).
14. Liu, X. *et al.* High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nat. Sustain.* **2020 37 3**, 564–570 (2020).

15. Jongman, B. *et al.* Declining vulnerability to river floods and the global benefits of adaptation. *Proc. Natl Acad. Sci. USA* **112**, E2271–E2280 (2015).
16. Ceola, S., Laio, F. & Montanari, A. Satellite nighttime lights reveal increasing human exposure to floods worldwide. *Geophys. Res. Lett.* **41**, 7184–7190 (2014).
17. Jongman, B., Ward, P. J. & Aerts, J. C. J. H. Global exposure to river and coastal flooding: long term trends and changes. *Glob. Environ. Chang.* **22**, 823–835 (2012).
18. Serpico, S. B. *et al.* Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proc. IEEE* **100**, 2946–2970 (2012).
19. Mateo-Garcia, G. *et al.* Towards global flood mapping onboard low cost satellites with machine learning. *Sci. Reports 2021 111* **11**, 1–12 (2021).
20. Sweta, L. O. & Bijker, W. Methodology for assessing the usability of earth observation-based data for disaster management. *Nat. Hazards* **65**, 167–199 (2013).
21. Simpson, R. B., Gottlieb, J., Zhou, B., Hartwick, M. A. & Naumova, E. N. Completeness of open access FluNet influenza surveillance data for Pan-America in 2005–2019. *Sci. Reports 2021 111* **11**, 1–17 (2021).
22. Zhang, Y. *et al.* Evaluating Completeness of Foodborne Outbreak Reporting in the United States, 1998–2019. *Int. J. Environ. Res. Public Heal.* **2022**, Vol. 19, Page 2898 **19**, 2898 (2022).
23. Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
24. The Flood Observatory. <https://floodobservatory.colorado.edu/index.html>.
25. Brakenridge, G. R. For Research, Humanitarian, and Water Management

Applications.

26. MODIS Web. <https://modis.gsfc.nasa.gov/about/>.
27. Twele, A., Cao, W., Plank, S. & Martinis, S. Sentinel-1-based flood mapping: a fully automated processing chain. <http://dx.doi.org/10.1080/01431161.2016.1192304> **37**, 2990–3004 (2016).
28. Wu, H. *et al.* Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* **50**, 2693–2717 (2014).
29. OPERA. OPERA Dynamic Surface Water Extent from Harmonized Landsat Sentinel-2 Provisional Product. (2023) doi:10.5067/OPDSW-PL3V1.
30. Tarpanelli, A., Mondini, A. C. & Camici, S. Effectiveness of Sentinel-1 and Sentinel-2 for flood detection assessment in Europe. *Nat. Hazards Earth Syst. Sci.* **22**, 2473–2489 (2022).
31. Memon, A. A., Muhammad, S., Rahman, S. & Haq, M. Flood monitoring and damage assessment using water indices: A case study of Pakistan flood-2012. *Egypt. J. Remote Sens. Sp. Sci.* **18**, 99–106 (2015).
32. Sanyal, J. & Lu, X. X. Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review. *Nat. Hazards 2004 332* **33**, 283–301 (2004).
33. The Flood Observatory. <https://floodobservatory.colorado.edu/index.html>.
34. Hirpa, F. A. *et al.* Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.* **566**, 595–606 (2018).
35. EOSDIS Worldview. <https://worldview.earthdata.nasa.gov/?v=-206.31832957697642,-46.546875,5.787079576976424,53.015625&t=2022-04->

26-T23%3A12%3A34Z.

36. Worldview | Earthdata. <https://earthdata.nasa.gov/worldview>.
37. UNU-INWEH Flood Mapping Tool. <https://floodmapping.inweh.unu.edu/>.
38. GADAS. <https://geo.fas.usda.gov/GADAS/index.html>.
39. Global Flood Database. <https://global-flood-database.cloudtostreet.ai/>.
40. Global Flood Awareness System – global ensemble streamflow forecasting and flood forecasting. <https://www.globalfloods.eu/>.
41. Global Surface Water Explorer. <https://global-surface-water.appspot.com/#features>.
42. Kruczkiewicz, A. *et al.* Multiform flood risk in a rapidly changing world: what we do not do, what we should and why it matters. *Environ. Res. Lett.* **17**, 081001 (2022).
43. Reed, C. *et al.* The impact of flooding on food security across Africa. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2119399119 (2022).
44. Health Effects of Flooding in Rural Bangladesh on JSTOR. <https://www.jstor.org/stable/23214183>.
45. Oskorouchi, H. R. & Sousa-Poza, A. Floods, food security, and coping strategies: Evidence from Afghanistan. *Agric. Econ. (United Kingdom)* **52**, 123–140 (2021).
46. del Ninno, C. & Lundberg, M. Treading water. The long-term impact of the 1998 flood on nutrition in Bangladesh. *Econ. Hum. Biol.* **3**, 67–96 (2005).
47. Rodriguez-Llanes, J. M., Ranjan-Dash, S., Degomme, O., Mukhopadhyay, A. & Guha-Sapir, D. Child malnutrition and recurrent flooding in rural eastern India: a community-based survey. *BMJ Open* **1**, e000109 (2011).

48. Sherman, M., Ford, J., Llanos-Cuentas, A. & Valdivia, M. J. Food system vulnerability amidst the extreme 2010–2011 flooding in the Peruvian Amazon: a case study from the Ucayali region. *Food Secur.* **8**, 551–570 (2016).
49. Salvucci, V. & Santos, R. Vulnerability to Natural Shocks: Assessing the Short-Term Impact on Consumption and Poverty of the 2015 Flood in Mozambique. *Ecol. Econ.* **176**, (2020).
50. Cutter, S. L. The Perilous Nature of Food Supplies: Natural Hazards, Social Vulnerability, and Disaster Resilience. <http://dx.doi.org/10.1080/00139157.2017.1252603> **59**, 5–15 (2016).
51. Alias, N. A. & Nor, N. M. Measurement needs of food security during flood disaster in Bera, Pahang. *Curr. Res. Nutr. Food Sci.* **6**, 83–88 (2018).
52. Akukwe, T. I., Oluoko-Odingo, A. A. & Krhoda, G. O. Do floods affect food security? A before-and-after comparative study of flood-affected households' food security status in South-Eastern Nigeria. *Bull. Geogr. Socio-economic Ser.* **47**, 115–131 (2020).
53. Parvez, M., Islam, M. R. & Dey, N. C. Household food insecurity after the early monsoon flash flood of 2017 among wetland (Haor) communities of northeastern Bangladesh: a cross-sectional study. *Food Energy Secur.* **11**, e326 (2022).
54. Doocy, S., Leidman, E., Aung, T. & Kirsch, T. Household economic and food security after the 2010 Pakistan floods. *Food Nutr. Bull.* **34**, 95–103 (2013).
55. Rodriguez-Llanes, J. M., Ranjan-Dash, S., Mukhopadhyay, A. & Guha-Sapir, D. Flood-Exposure is Associated with Higher Prevalence of Child Undernutrition in Rural Eastern India. *Int. J. Environ. Res. Public Health* **13**, 210 (2016).
56. Sam, A. S. *et al.* Flood vulnerability and food security in eastern India: A threat to the achievement of the Sustainable Development Goals. *Int. J. Disaster Risk*

Reduct. **66**, 102589 (2021).

57. Ahmad, D. & Afzal, M. Flood hazards, human displacement and food insecurity in rural riverine areas of Punjab, Pakistan: policy implications. *Environ. Sci. Pollut. Res.* **28**, 10125–10139 (2021).
58. Riadi, B. Strategy to Maintain Food Security in the Area of Flood Hazard in Karawang Regency. *IOP Conf. Ser. Earth Environ. Sci.* **165**, 12029 (2018).
59. Rosalia, A. C. T., Suryanto & Hakim, L. Spatial analysis of the impact of flood and drought on food security index. *Nat. Environ. Pollut. Technol.* **20**, 721–727 (2021).
60. Rahman, M. S. & Di, L. A systematic review on case studies of remote-sensing-based flood crop loss assessment. *Agric.* **10**, (2020).
61. Global Flood Monitoring. <http://flood.umd.edu/>.
62. Slayback, D. (NASA). MODIS NRT Global Flood Product. <https://www.earthdata.nasa.gov/global-flood-product>.
63. GFM Product User Manual. <https://extwiki.eodc.eu/GFM/PUM>.
64. Passerotti, G. *et al.* Hydrological Model Application in the Sirba River: Early Warning System and GloFAS Improvements. *Water* **2020** **12**, 620 (2020).
65. Alfieri, L. *et al.* GloFAS – global ensemble streamflow forecasting and flood early warning. **17**, 1161–1175.
66. Moraes, M. A. E., Neto, L. A. dos S., Pimentel, A. & Brown, V. L. R. GloFAS as a Flood Alert System in Acre Civil Defense.
67. Chen, A., Giese, M. & Chen, D. Flood impact on Mainland Southeast Asia between 1985 and 2018—The role of tropical cyclones. *J. Flood Risk Manag.* **13**, e12598 (2020).

68. Kundzewicz, Z. W., Pińskwar, I. & Brakenridge, G. R. Large floods in Europe, 1985–2009. *Hydrol. Sci. J.* **58**, 1–7 (2013).
69. Cunado, J. & Ferreira, S. The Macroeconomic Impacts of Natural Disasters: The Case of Floods. *Land Econ.* **90**, 149–168 (2014).
70. Kundzewicz, Z. W., Pińskwar, I. & Brakenridge, G. R. Changes in river flood hazard in Europe: a review. *Hydrol. Res.* **49**, 294–302 (2018).
71. Hu, P., Zhang, Q., Shi, P., Chen, B. & Fang, J. Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. *Sci. Total Environ.* **643**, 171–182 (2018).
72. Halgamuge, M. N. & Nirmalathas, A. Analysis of large flood events: Based on flood data during 1985–2016 in Australia and India. *Int. J. Disaster Risk Reduct.* **24**, 1–11 (2017).
73. Idowu, D. & Zhou, W. Performance evaluation of a potential component of an early flood warning system-a case study of the 2012 flood, lower Niger River Basin, Nigeria. *Remote Sens.* **11**, (2019).
74. Ferreira, S. & Ghimire, R. Forest cover, socioeconomics, and reported flood frequency in developing countries. *Water Resour. Res.* **48**, 8529 (2012).
75. Kumar, N. *et al.* Applicability of HEC-RAS 2D and GFMS for flood extent mapping: a case study of Sangam area, Prayagraj, India. *Model. Earth Syst. Environ.* **6**, 397–405 (2020).
76. Kumar, N., Lal, D., Sherring, A. & Issac, R. K. Applicability of HEC-RAS & GFMS tool for 1D water surface elevation/flood modeling of the river: a Case Study of River Yamuna at Allahabad (Sangam), India. *Model. Earth Syst. Environ.* **3**, 1463–1475 (2017).

77. Pittman, A., Carlson, K., Curran, L. & Ponette-González, A. NASA Satellite Data Used to Study the Impact of Oil Palm Expansion Across Indonesian Borneo. *Earth Obs.* **25**, 12–16 (2013).
78. Sherbatov, A., Hsiang, E., Kilburn, C., Ortiz, J. & Koppe, B. Case Study of Camp Fire Employing Novel Metric for Time Series Analysis of Vegetation Recovery. *ESS Open Arch.* (2021) doi:10.1002/essoar.10508247.1.
79. Peters, B. V., Bagwell, R. E. & Wong, M. M. Flood Analysis of the Thomas Flood in California Using NASA Data in a GIS. (2018).
80. Mentaschi, L., Voudoukas, M. I., Pekel, J. F., Voukouvalas, E. & Feyen, L. Global long-term observations of coastal erosion and accretion. *Sci. Rep.* **8**, 1–11 (2018).
81. Nienhuis, J. H. *et al.* Global-scale human impact on delta morphology has led to net land area gain. *Nature* **577**, 514–518 (2020).
82. Gupta, N. & Kochhar, I. Examining stakeholders' views and opinions for river conservation in India: a case study from Pune, Maharashtra. *NeBIO An Int. J. Environ. Biodivers.* **9**, 7–12 (2018).
83. Lassa, J. A. *Emerging 'Agricultural Involution' in Indonesia: Impact of Natural Hazards and Climate Extremes on Agricultural Crops and Food System.* (2012).
84. Zhou, B. *et al.* Food and Nutrition Systems Dashboards: A Systematic Review. *Adv. Nutr.* **10**, S308–S319 (2019).

3.8 Acknowledgements

Thank you to Katherine Berry and Leah Beaulac for their guidance on this project during their time as masters' students in the Tufts Sustainable Water Management program. We greatly acknowledge support from the Nutrition Epidemiology and Data Science Doctoral Seminar at the Tufts University Friedman School of Nutrition Science

and Policy for providing critical feedback of the tools and information presented in this paper. Thank you to the survey participants from Tufts University and Boston University who took the time to evaluate the flood data portals and provide their valuable opinions. We would like to acknowledge Ryan Simpson, Alex Carroll, Tyler Stotland, and Nilakshi Chatterji for their prior work with Tufts Institute of the Environment which inspired this analysis. Finally, thank you to colleagues at the center for Coastal Disaster Mitigation and Rehabilitation Studies, Universitas Diponegoro in Semarang City, Central Java, Indonesia, for their continued guidance and critical feedback on all aspects of this project.

3.9 Author Contributions

Conceptualization, B.K.L.; methodology, B.K.L., M.K., and E.N.N.; formal analysis, B.K.L.; investigation, B.K.L. and A.M.; data curation, B.K.L. and A.M.; writing—original draft preparation, B.K.L. and A.M.; writing—review and editing, B.K.L., M.K., A.M., T.G., E.C.d.P., E.N.N.; supervision, E.N.N., E.C.d.P., T.G., M.K.; funding acquisition, E.N.N., T.G. and M.K. All authors have read and agreed to the published version of the manuscript.

3.10 Competing Interests

The authors declare no conflicts of interest.

Chapter 4: Spatial and Survey Data Linkages

4.1 Linkage of Global Surface Water 1984-2015 raster dataset to a longitudinal household survey for integrative statistical analysis (*TIES 2023 abstract*)

Breanne K. Langlois¹, Erin Coughlan de Perez¹, Magaly Koch², Timothy Griffin¹, Elena N. Naumova¹

¹Tufts University Friedman School of Nutrition Science and Policy, Boston, MA, USA

²Center for Remote Sensing, Boston University, Boston, MA, USA

Abstract:

There is an abundance of continuous, publicly available data from satellite imagery that spans decades, but a lack of integrated datasets to evaluate how the changing environment shapes human outcomes. We identified an efficient method to link high-resolution raster data to a population survey using R programming and applied these steps to the Global Surface Water 1984-2015 (version 1) dataset (GSW) and the Indonesian Family Life Survey (IFLS). GSW captures the overall spatiotemporal distribution of Earth's surface water since 1984, obtained from ~3 million Landsat images at 30-meter pixel resolution. IFLS interviewed 15,921 households over 22 years from 1993-2015 with a district level of spatial aggregation. We downloaded GSW raster files and vector (polygon) data of Indonesian subnational boundaries. We read data into R using the 'raster' and 'sf' packages, then cropped and masked the GSW to Central Java. Using the 'exactextractr' package, we extracted and summarized GSW raster values by Central Java districts. We exported the results as a Stata dataset with rows representing districts and columns representing summary statistics. Finally, the dataset was merged to IFLS by district. The resulting, multidisciplinary dataset contains 6,743

observations (Central Java households over 5 waves) and detailed information on sociodemographic characteristics, assets, economic shocks, businesses, income, household consumption and expenditures, and surface water changes in the districts where households live. For each district, variables representing the sum of GSW raster cells covered by the polygon and sum, average, variety, standard deviation, min, max, and fractions of cells corresponding to each value of surface water occurrence, change, and inter-annual behavior were included in the dataset. We created a multidisciplinary dataset. Data linkages are challenging, but possible, and open the door to a range of statistical analyses about environmental impacts on food security, nutrition, and public health that were previously not possible.

Keywords: remote sensing data, multidisciplinary, data linkages

4.2 Technical Workshop (*presented at Tufts Nutrition Data Symposium 2023*)

The presentation in this section was part of a Workshop titled “Data Linkages for Multidisciplinary Research” presented at the Tufts Nutrition Data Symposium on March 9th, 2023. This workshop provided an introduction and overview of technical data analytic tools with a primary focus on methodology. The format was 50 minutes total, including 40 minutes of presentation time and 10 minutes of questions from the participants. My presentation followed another speaker.

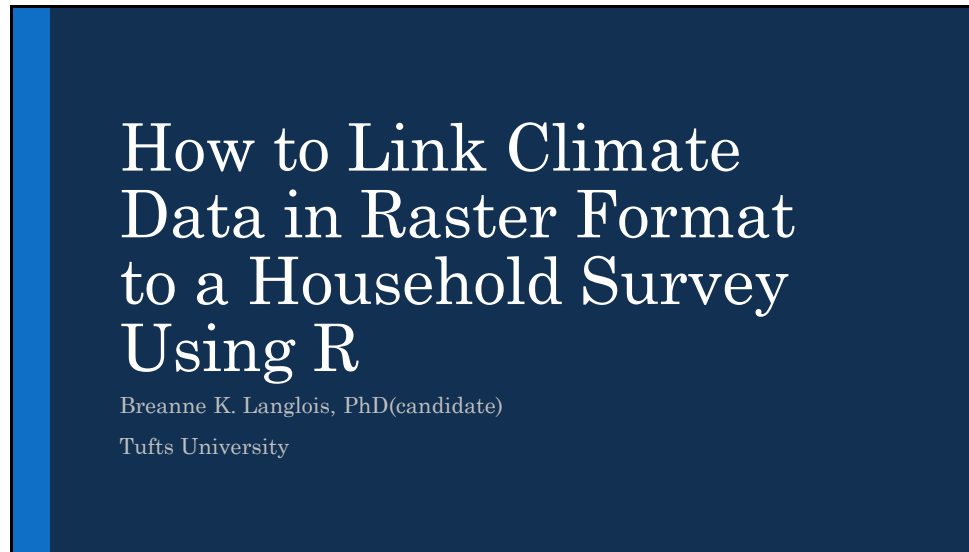


Figure 1. Title slide

In this workshop, I presented an example of how to link climate data to a household survey with some technical instruction. Participants learned important skills and resources to accomplish this task on their own.

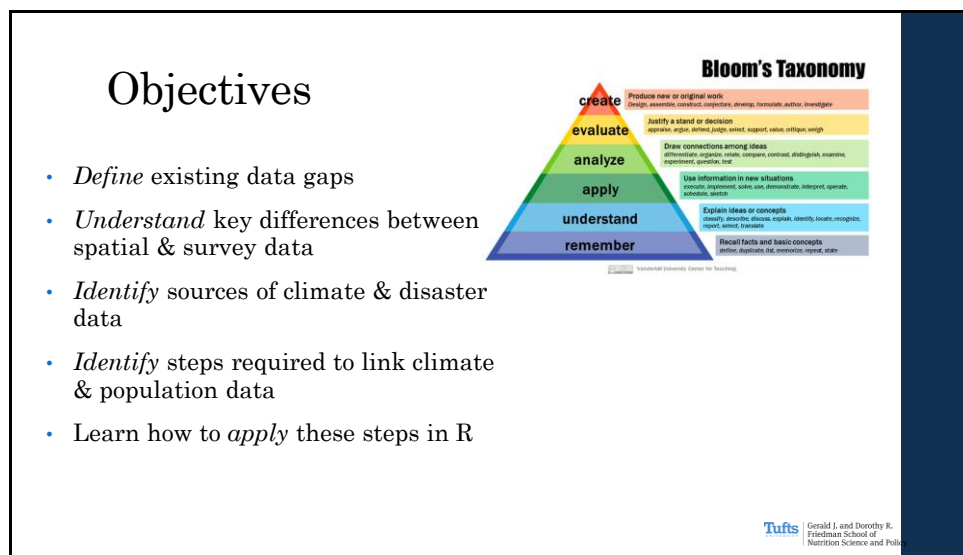


Figure 2. Workshop objectives

The learning objectives for the workshop are in Figure 2, which I created and structured according to Bloom's Taxonomy. First, I *reviewed* relevant information by defining existing data gaps related to studying how climate events impact nutrition and food security. Next, we aimed to *understand* important concepts by discussing key differences between spatial and survey data and identifying the necessary steps required to link climate data to a population survey. Finally, participants learned how to *apply* these steps using R programming.

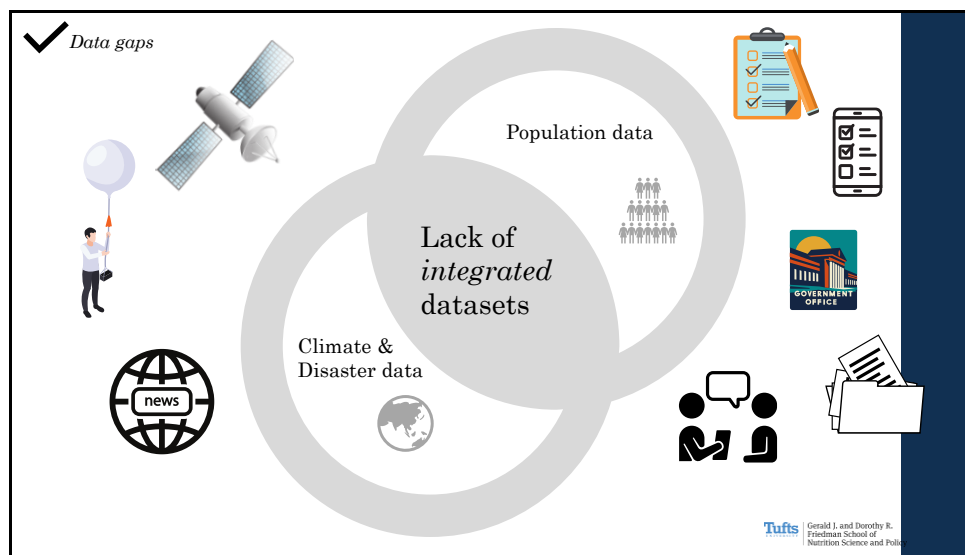


Figure 3. Data gaps

There is now an abundance of continuous data available in the public domain pertaining to the climate that spans decades. Most of the data is geospatial, gathered from satellite imagery which can be highly granular. Other sources of climate and disaster data are weather stations or weather balloons. Media reports provide another source of information for many disaster databases, though this type of data often lacks spatial granularity. A less common source of information about climate and disasters is surveys, providing reports about experiences with disasters over time.

There is also an abundance of data available from population surveys, including health, economic, and nutrition surveys. In contrast to geospatial data, population data usually requires direct interaction. In other words, a person or field team is needed to collect the data on the ground. These data are often collected via web or paper-based surveys or phone interviews. Other sources include government – such as the census or data from other agencies – or existing records (e.g. hospital data warehouses). While

population data can be longitudinal, it often does not have the same continuity over space and time that satellite imagery provides.

What's missing, however, is the gray area in between (Figure 3). There is a lack of datasets available containing information about both the climate and humans. As researchers, we must create these integrated, multidisciplinary datasets to evaluate questions about how the climate and disasters impact humans.

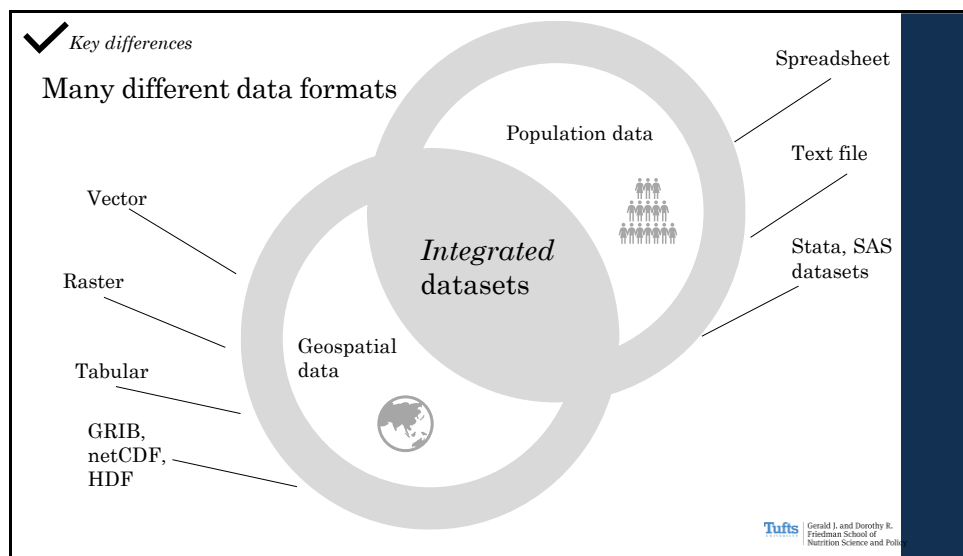


Figure 4. Key differences in data formats

There are many different data formats. For those with public health or nutrition backgrounds, we typically want a dataset in a common format that we can use for statistical analysis. The challenge is that these data are usually very difficult to connect because they often lack common attributes.

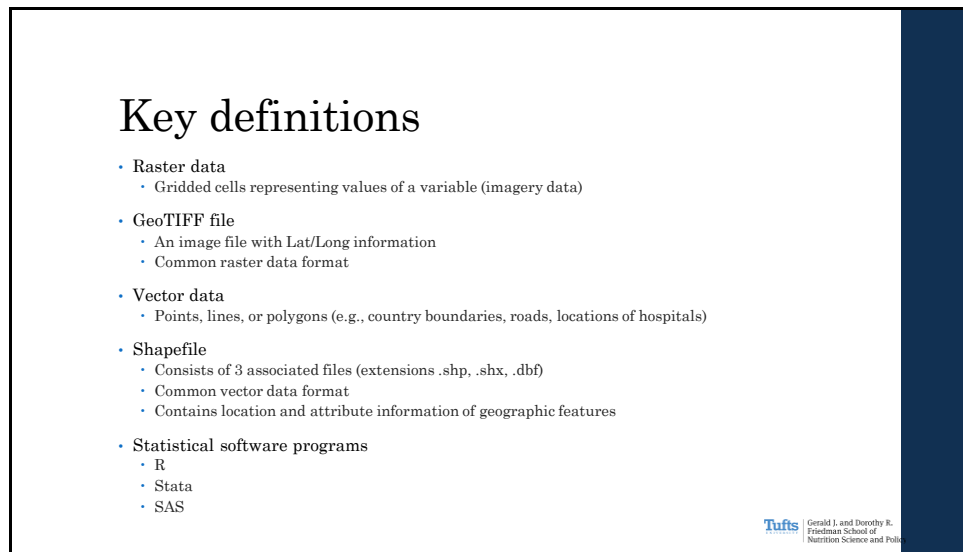


Figure 5. Key definitions

These are some key definitions that are important to understand. Geospatial data is created, shared, and stored in many different formats. The two primary data types are **raster** and **vector**.

Raster data are a series of grid cells where each cell has a value representing the feature being observed, depicted as an image, typically a GeoTIFF file. Raster data is best suited for continuous data, or information that does not have hard boundaries or locations. The data are viewed as a series of grid cells where each cell has a value representing the feature being observed. This is the common format of climate data from remote sensing – like aerial photography and satellite imagery – measuring things like temperature, precipitation, or other variables.

In contrast vector data is represented as either points, lines, or polygons, e.g. county boundaries, or locations of roads using lines. A shapefile is a common format for storing vector data.

For population or economic data is usually in the format of a statistical software program. Statistical software programs are tools designed for data management and statistical data analysis, each having a different user interface and programming

language. When it comes to population survey data, we usually see either Stata or SAS dataset format.

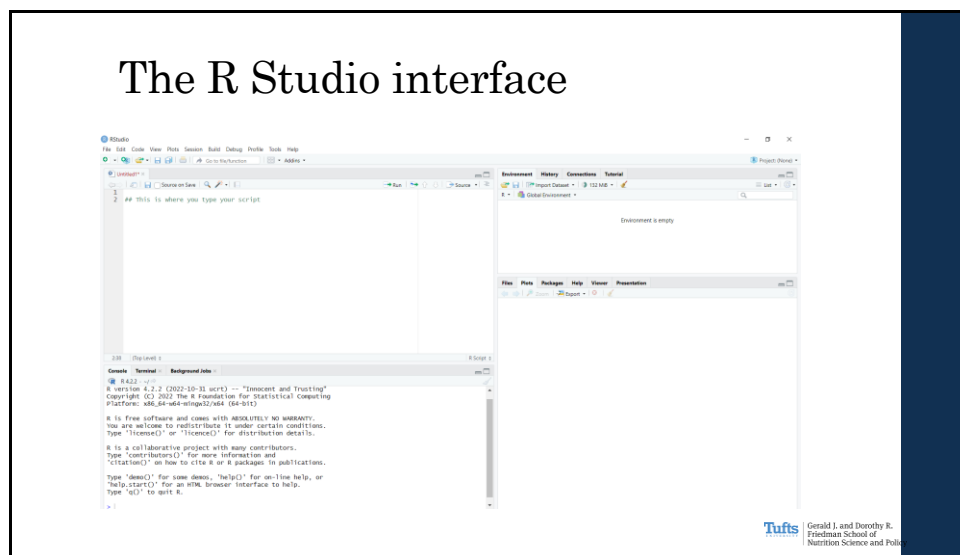


Figure 6. The R Studio interface

R is a common language and program for statistical computing. It is the only free, open-source statistical software. One of the main strengths of R is its versatility and ability to be extended through packages that the user can install through the Comprehensive R Archive Network (CRAN). R can cover a wide range of statistical and spatial analyses.

On the top left (shown in the slide above) is where the user writes and executes their script. The bottom left serves as the log and results window. The top right is where the user can view a listing of datasets and objects. The bottom right is where the plots appear once executed.

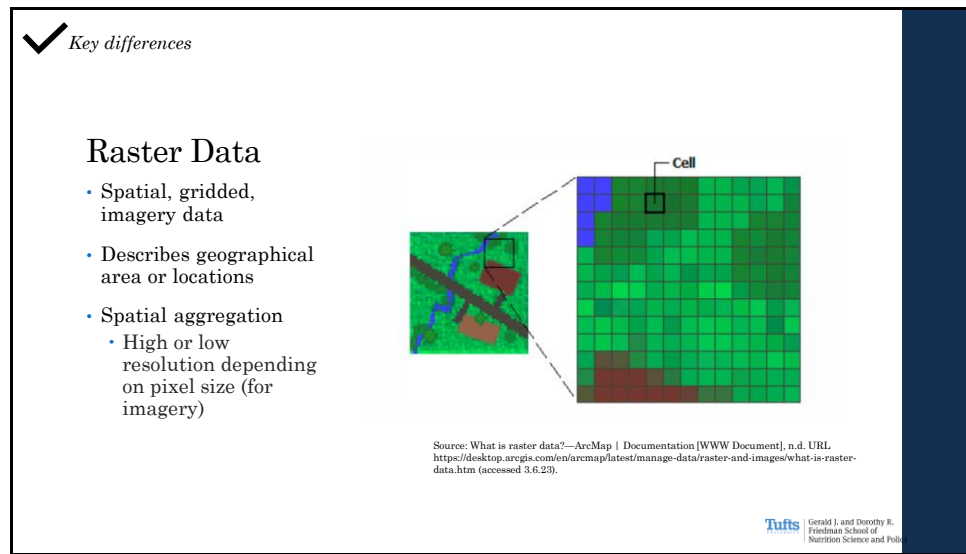


Figure 7. Description of raster data

I defined raster data previously. Here is an example of a raster dataset. You can see the gridded nature of data format. One key feature is that it can be highly granular.

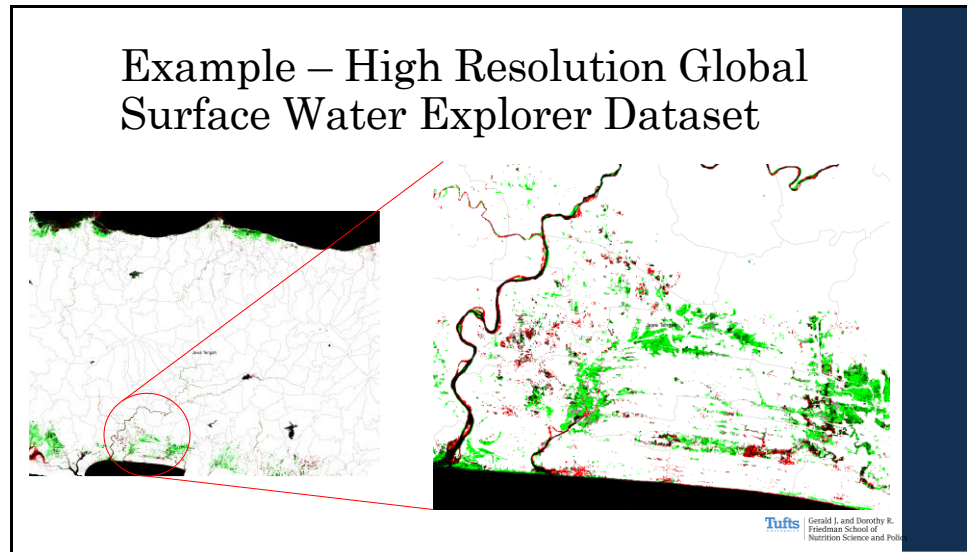


Figure 8. Example of high-resolution raster data

This is an example of raster a dataset. This is high-resolution data taken from the Global Surface Water Explorer. Shown here is Central Java, Indonesia, with a zoomed in view to the right. The faint gray outlines are the districts. The green and red are the colors representing the variable being shown which describes where surface water occurrence increased, decreased or remained the same between 1984-1999 and 2000-2021.

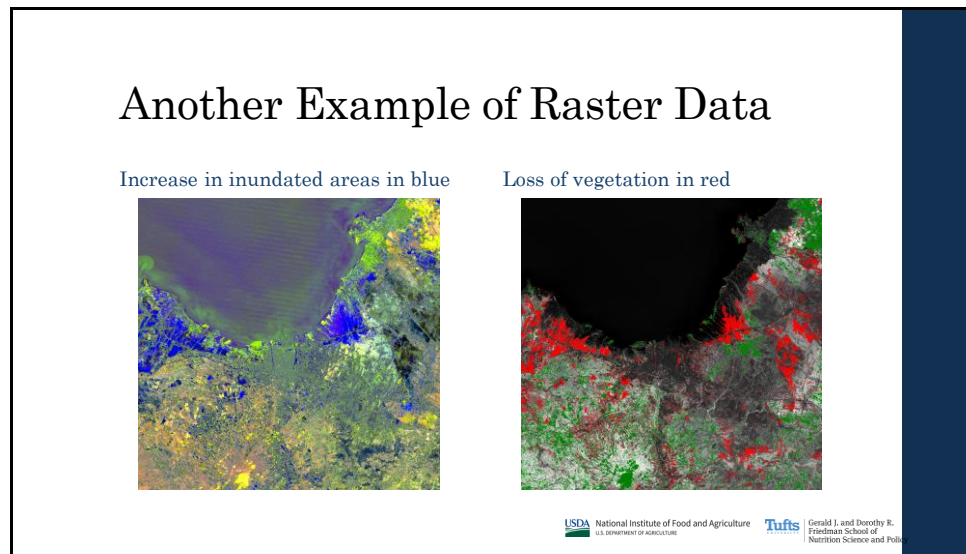


Figure 9. Example #2 of raster data

Shown here is another example of raster data. This was a change detection analysis done using ENVI software. The images show changes around Semarang City, Central Java from 2008-2019. The left highlights the large increase in inundated areas (using MNDWI -- Modified Normalized Difference Water Index) from 2008 to 2019, while the right depicts vegetation loss (using SVR – simple vegetation ratio).

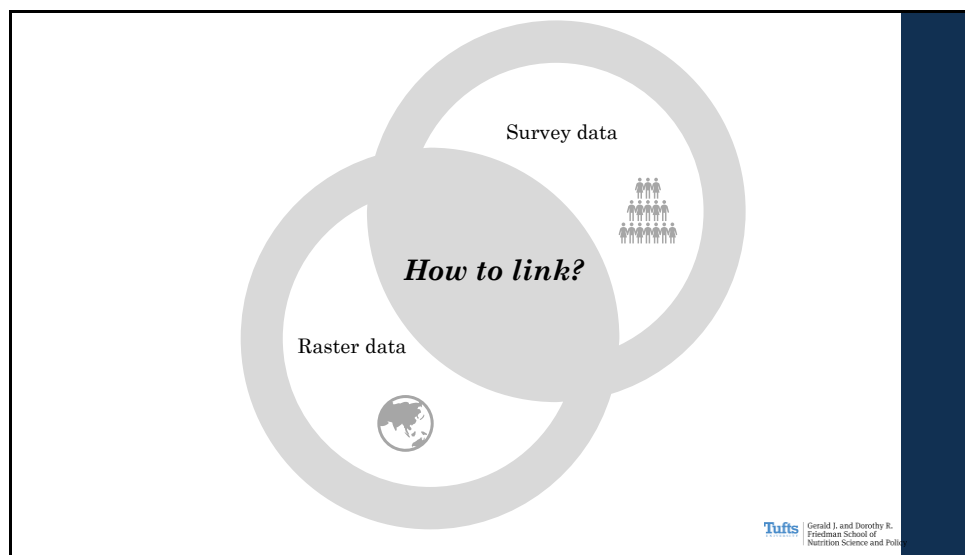


Figure 11. Key challenge

To link these two different types of data from different disciplines, you need common spatial and temporal variables. Usually, however, the different data sources do not match with regard to these attributes.

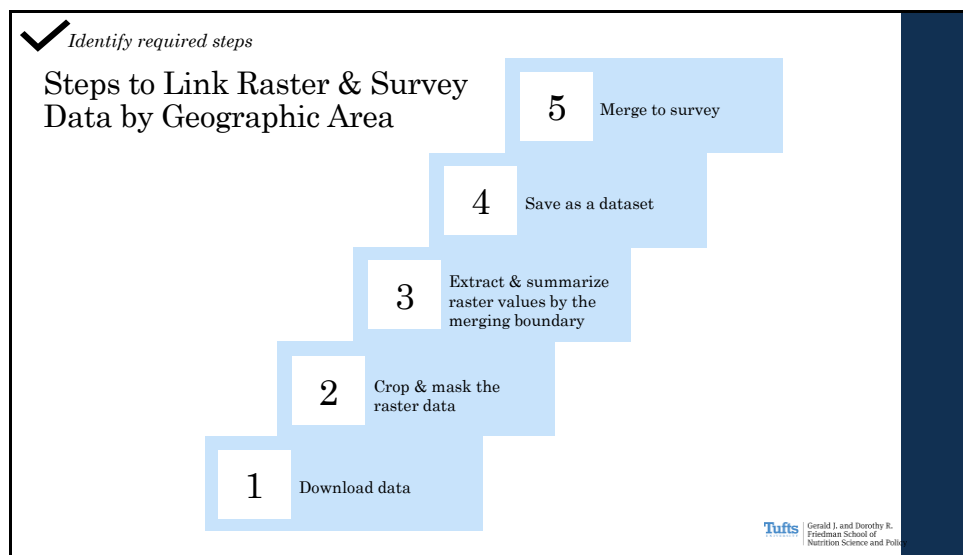


Figure 12. Required steps to link raster and survey data

These are the steps required to accomplish this type of data linking. The first step is to download the data. Both the raster data file and vector data representing the geographic boundaries are needed. The next step is masking and cropping the raster by the general area of interest (e.g., country or province level). Then, extract and summarize the raster values by the geographic boundary needed to merge the datasets (e.g. district or village). In other words, create a common attribute(s) so the spatial data can be linked to the survey. In the example that follows, the survey has households with the districts where they live. The raster data values therefore need to be summarized by districts. After summarizing the data by a common spatial boundary, export or save the results as a dataset where rows represent the geographic boundary and columns represent the summary statistics. Finally, the data can easily be merged into the survey by the common geographic boundary.

Example: Linking Global Surface Water Explorer Data to the Indonesian Family Life Survey



Figure 13. Data linking example

This example shows how high-resolution raster data from Global Surface Water Explorer, which describes the spatiotemporal distribution of surface water since 1984, can be linked to the Indonesian Family Life Survey, which is a longitudinal health and economic survey of the Indonesian population dating back to 1993. The Global Surface Water dataset does not vary over time, rather it captures the changes throughout the timespan. The Indonesian Family Life Survey varies by time (i.e. survey wave). District is the common attribute by which the two datasets can be linked.

Identify required steps

1

Download data

Raster data from
Global Surface Water Explorer

Download process

Individual 10°x10° files

The Global Surface Water data are available to download in files 10°x10° from the map shown below. Click on the t

Granule with top-left corner at 100E, 0N

Occurrence: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

Change: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

Seasonality 2021: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

Resurgence: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

Transients: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

Maximum extent: https://storage.googleapis.com/global-surface-water/tiles/100E_0N_100E_10N.tif

<https://global-surface-water.appspot.com/>

+

Vector (polygon) data of
Indonesian Administrative Boundaries

Indonesia administrative level 0 (country - regional), 1 (province - provincial), 2 (city, district, regency - kota, kabupaten), 3 (sub-district - kecamatan, district, and 4 village - desa, kelurahan, kampung, nagari, jambo or gampong) boundaries.

Writing and live service provision by [Informative Technology Outreach Services \(ITOS\)](#) with funding from USAID.

<https://data.humdata.org/dataset/cod-ab-idn>

Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Figure 14. Step 1 of the data linking process: Download data

As mentioned previously, both the raster data and a shapefile containing administrative boundaries for the area of interest are needed. To the left is an example of the data download process for the Global Surface Water dataset. The raster data is downloaded as ‘tif’ image tiles shown in the grid. On the right is an example of where you could obtain data on administrative boundaries. This is vector data, where the boundaries are represented by polygons.

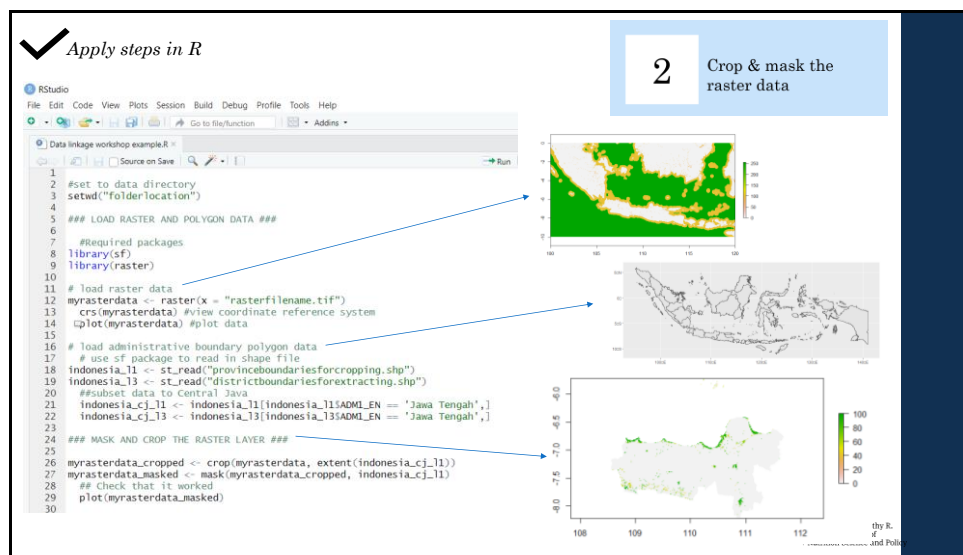


Figure 15. Step 2 of the data linking process: Crop and mask the raster

The spatial data can be loaded into R using the 'sf' and 'raster' packages, as shown here. The map can then be visualized using the plot function, shown on the right. The raster data values shown here range from 0-100. These are discrete values with 0 = Not Water and 100 = 100% recurrence. On lines 21 and 22, the vector data is subset to Central Java (the area of interest) using 2 levels. The first level is the province boundary for cropping the raster. The second level is the district boundary for summarizing the raster values. Next, the raster data is cropped to Central Java and masked to produce a raster map of just Central Java (lines 26-27). The masked output is shown in the plot at the bottom. This cuts down substantially on the processing time for the next step.

✓ *Apply steps in R*

R package 'exactextractr'

3 Extract & summarize raster values by the merging boundary

```

32
33 ### EXTRACT RASTER VALUES PER DISTRICT LEVEL POLYGON ###
34
35 #Required packages
36 library(exactextractr) #Fast Extraction from Raster Datasets using Polygons
37 #(https://cran.r-project.org/web/packages/exactextractr/exactextractr.pdf)
38
39 #return summary statistics for each district polygon
40 myrasterdata_polygon <- exact_extract(myrasterdata_masked,
41 indonesia_cj_l3, append_cols=TRUE, quantiles=.75,
42 c('count', 'sum', 'mean', 'min', 'max', 'stdev', 'frac'))
43

```

Tufts Gerald J. and Dorothy R. Friedman School of Nutrition Science and Policy

Figure 16. Step 3 of the data linking process: Extract and summarize raster values

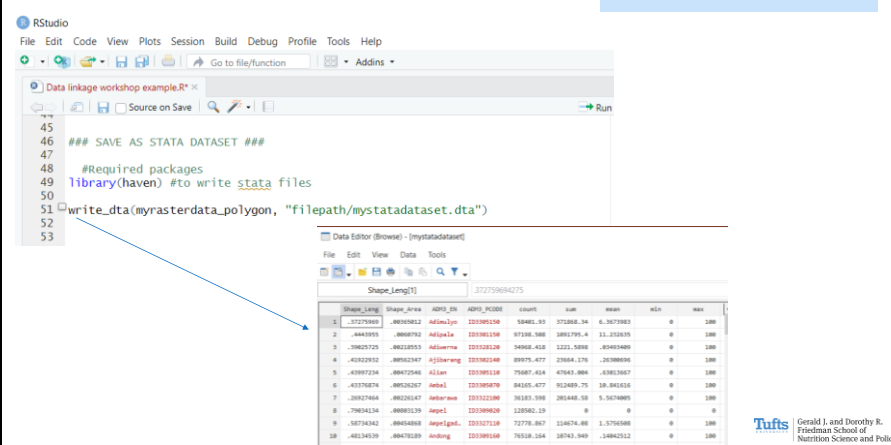
In this step, we extract and summarize the raster values by the common administrative boundary (district in this example). This can take a lot of processing time if using the traditional GIS interface with toolbox (e.g. raster to point conversion), especially with high-resolution data. With R programming, however, this task is simplified by using the 'exactextractr' package on the masked data from the previous step. Many summary statistics can be obtained with this package, including a count of the raster cells within each district, a sum of all the values, the average and many others.



Apply steps in R

4

Save as a dataset



The RStudio interface shows the following R code in the console:

```

45
46 ### SAVE AS STATA DATASET ###
47
48 #Required packages
49 library(haven) #to write stata files
50
51 write_dta(myrasterdata_polygon, "filepath/mystatadataset.dta")
52
53
  
```

Below the code, the 'Data Editor (Browse)' window displays a table with the following columns: Shape_Leng1, Shape_Area, ADP3_36, ADP3_PC300, count, sum, mean, min, and max. The table contains 10 rows of data representing different districts.

Shape_Leng1	Shape_Area	ADP3_36	ADP3_PC300	count	sum	mean	min	max
1	37275968	ADP3_36	ADP3_PC300	58481.93	571388.34	6.3673983	0	100
2	4445955	ADP3_36	ADP3_PC300	97338.388	3861795.4	11.212835	0	100
3	34802579	ADP3_36	ADP3_PC300	34862.428	1221.3098	0.0004989	0	100
4	43282932	ADP3_36	ADP3_PC300	69975.477	23884.176	0.3418856	0	100
5	43997234	ADP3_36	ADP3_PC300	75887.454	47643.884	0.3813627	0	100
6	43378274	ADP3_36	ADP3_PC300	84335.477	912489.75	10.841616	0	100
7	38927464	ADP3_36	ADP3_PC300	36185.398	281448.58	5.5674895	0	100
8	78042434	ADP3_36	ADP3_PC300	120582.19	0	0	0	0
9	58774342	ADP3_36	ADP3_PC300	12778.867	114674.88	1.5756588	0	100
10	48134539	ADP3_36	ADP3_PC300	76538.344	18743.949	1.4842512	0	100

Figure 17. Step 4 of the data linking process: Save raster as tabular dataset

From R, the resulting data can be saved to any format. In this example, the dataset is saved as Stata format – the same format as the survey data it will be linked to. The dataset shown at the bottom is the result of the previous step. Each row represents a district, and each column is a variable containing summary information of the raster values within that district. Next, the data can be merged to the survey.



Apply steps in R

5

Merge to survey



Do-file Editor - Untitled*

```

1
2
3
4
5
6
7
      "merge the 2 Stata datasets
      use mysurveydata.dta ///survey dataset structured by household
      merge m:1 ADM3_PCODE using "${working}\mystatadataset.dta" |
    
```

Data Editor (Browse) - [merged_dataset]

	hhid	if1s_wave	ADM3_PCODE	plve1_count	mean_occ	wrice
1	0000300	2	3303020	61432.449	-.55858624	28.98937
2	0000300	3	3303020	61432.449	-.55858624	14.24983
3	0000300	4	3303020	61432.449	-.55858624	15.78921
4	0000300	5	3303020	61432.449	-.55858624	16.7461
5	0000341	4	3303020	61432.449	-.55858624	14.39863
6	0000341	5	3303020	61432.449	-.55858624	16.66262
7	0000342	4	3303020	61432.449	-.55858624	34.28667

Figure 18. Step 5 of the data linking process: Merge raster data to survey

Merging the two datasets is now made simple and can be done by merging many (the survey data has many rows per district) to one (the spatial data has one row per district) by district. The resulting merged dataset contains both household level information from the survey and information about the changing environment in the districts where they live. This is now a multidisciplinary dataset, where a range of questions can be addressed that were previously not possible.

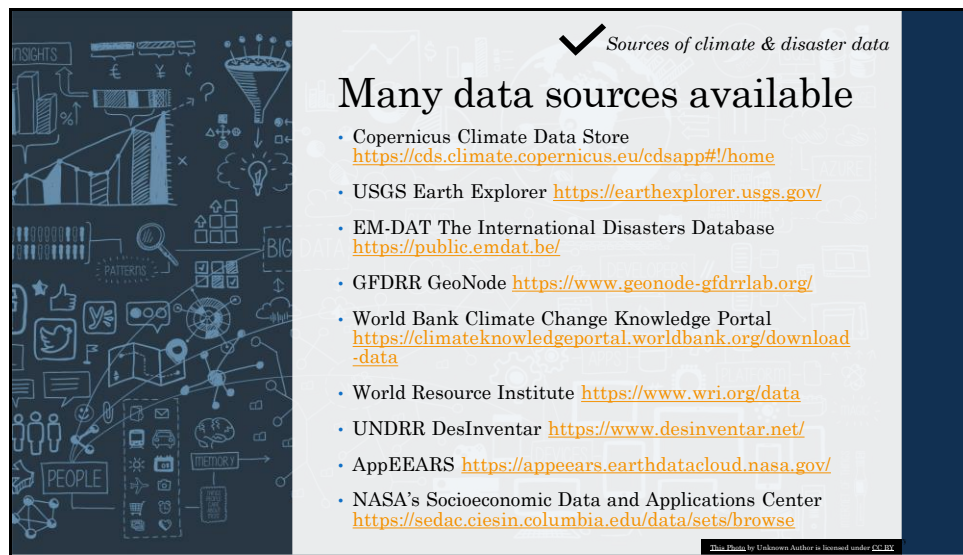


Figure 19. Available data sources

Lastly, there are many kinds of data products available, as shown here. There are disaster databases, climate data products derived from satellite imagery, spatial GIS layers, or some surveys that include self-reported information about experiences with disasters. This is not a comprehensive list but can be used as a further resource.

4.3 References

1. Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418-422. (doi:10.1038/nature20584)
2. Baston, D., ISciences, L. L. C., & Baston, M. D. (2022). Package 'exactextractr'. *terra*, 1, 17.

Chapter 5: Recurrent flooding and household food access in Central Java, Indonesia (*manuscript formatted for Lancet Planetary Health*)

Breanne K. Langlois¹; Aris Ismanto²; Leah Beaulac¹; Katherine Berry¹; Magaly Koch³; Timothy Griffin¹; Erin Coughlan de Perez¹; Elena Naumova¹

¹ Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

² Departemen Oseanografi, Universitas Diponegoro, Semarang, IDN

³ Center for Remote Sensing, Boston University, Boston, MA, USA

5.1 Abstract

Background: Recurring flooding in Central Java impacts food availability through direct losses to agricultural crops, but it is not well-understood how floods are affecting diet-related aspects of food security such as household food access and utilization. We explored the relationship between recurrent flooding and household food access over a 22-year span in Central Java. We examine linear and nonlinear relationships between recurrent flooding within districts and household diet quantity and quality and joint effects with indicators of adaptive capacity. **Methods:** We linked the Global Surface Water dataset to the Indonesian Family Life Survey by district and analyzed Central Java households across 5 survey waves spanning 1993 to 2015. We measured recurrent flooding as the fraction of district raster cells with episodic flooding over the period 1984-2015 (defined as GSW recurrence values between 1-99). We constructed food access indicators based on the Data4Diets framework using the available consumption and expenditure data: household food expenditure share and dietary diversity score. We used statistical and machine learning methods to assess the relationship – generalized linear mixed models and Random Forest regression models. **Findings:** We detected joint effects with flooding and adaptive capacity. Wealth and access to credit

were associated with improved household food expenditure share and diet diversity. The effect of wealth on food expenditure share was stronger in households in more flood affected districts, while access to credit was associated with reduced odds of diet diversity in more flood affected districts. Flooding had more predictive importance for food expenditure share than for diet diversity. **Interpretation:** Access to credit, a factor that ordinarily improves food access, may not be effective in flood-prone areas for already vulnerable households. Wealthier households may be better able to adapt in terms of food access. **Funding:** This research was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) Cooperative State Research, Education, and Extension Service Fellowship (grant number 2020-38420-30724), and the NSF International Research Experiences for Students (IRES) grant: US-Indonesian Research Experience for Students on Sustainable Adaptation of Coastal Areas to Environmental Change (grant number 1826939).

Keywords: flood; natural disasters; data usability; food security; food access

5.2 Introduction

It is known that climate events are impacting food security around the world through direct losses to agricultural crops, presenting a major challenge to food and nutrition security through many pathways¹⁻⁵. In the 6th Assessment Report, the Intergovernmental Panel on Climate Change (IPCC) estimates increased nutrition-related diseases and undernourishment because of ongoing climate changes⁵. Flooding, caused by significant variations in rainfall patterns and other factors, is threatening our food system^{1,2}.

Current literature indicates that while higher temperatures are associated with decreased diet diversity, increased rainfall has the opposite effect in some regions⁶. The current evidence base for flooding, however, is poor^{3,6}. Much of the research on

flooding and food security has focused on food production or availability, while other diet-related dimensions at the household level are not well-understood^{1,7}. Some studies have looked at diet diversity following floods^{8–11}. Some evidence of undernutrition at the household level suggests a link between extreme floods and stunting^{12–15}. The sparseness of these studies and their heterogeneity make causal inference difficult. While flood impacts on disease outcomes are better understood, more high-quality interdisciplinary studies are needed to understand the impacts to the multiple dimensions of food and nutrition security, diet, and nutritional status^{16–18}. Further, the cumulative effects of *recurrent* flooding over time are understudied, with many prior studies focused only single events^{8–11,15,19,20}. Use of existing, longitudinal data can play an important role in addressing these knowledge gaps when linked with remote sensing data.

Central Java, Indonesia is an important case study of how recurrent flooding impacts household-level food security. Due to a confluence of geological, environmental, and economic factors, Central Java faces recurrent fluvial and tidal flooding from tropical storms, intense rainfall, sea level rise, and land subsidence^{21,22}. Central Java is also predominantly agricultural, producing most of Indonesia's rice supply²³. Another important feature is the proclivity to stay and adapt in response to floods for various reasons^{24–27}. The impacts of flooding on household diet and nutrition in this area are not fully understood. The Indonesian Family Life Survey (IFLS) provides data on 1,949 Central Java households over 22 years from 1993-2015²⁸. The IFLS provides consumption and expenditure data which can be used to derive household food access indicators⁷.

In this paper, we link Global Surface Water (GSW) data – describing the long term spatiotemporal distribution of surface water and its changes – to the IFLS to explore the relationship between recurring flooding and household access to 1) sufficient quantities of food and 2) adequate nutrition in Central Java, Indonesia²⁹. We explore the form and shape of the relationship by examining nonlinearity in our models and whether differential effects exist based on indicators of adaptive capacity. As

flooding and climate-related hazards are expected to worsen, there is a need to understand the impacts on diet and nutrition to identify key areas for intervention to achieve sustainable development goals 2 and 13. Further, a more comprehensive understanding of the impacts of disasters, including flooding, on food and nutrition security is needed for climate resilience policies¹.

5.3 Methods

5.3.1 Data Sources and Survey Design

We used the Indonesian Family Life Survey and Global Surface Water datasets^{28,30}. Both are publicly available and free to users.

5.3.1a Indonesian Family Life Survey

The IFLS is a nationally representative, longitudinal survey designed to study various aspects of social and economic well-being of the Indonesian population. The survey includes data collected from individuals, households, and communities, allowing users to address a range of questions related to livelihoods, expenditures, migration, health, and social relationships. The IFLS included 5 survey waves, with the first conducted in 1993-94. The following 4 waves were conducted in 1997-98, 2000, 2007-08, 2014-15, respectively. We conducted a longitudinal study of Central Java households using all 5 waves of IFLS. Our dataset included modules from the household survey: household characteristics, consumption and expenditures, farm and non-farm business, household assets, and household economic shocks.

The IFLS was conducted by the RAND Corporation²⁸. All waves underwent ethical approval by the RAND Institutional Review Board (IRB) in the United States and the University of Gadjah Mada IRB in Indonesia prior to the collection of any data. The approval number given by the RAND IRB is s0064-06-01-CR01.

Sampling for all waves was based on Wave 1 (1993-94), which used 2-stage stratification of province and urban/rural location. Provinces were selected to capture the diversity of the population to maximize representativeness. For cost reasons, only 13 of the 27 provinces were sampled in 1993. In Wave 1, the goal was to achieve a final sample size of 7,000 households. Assuming 10% nonresponse, the sample size was inflated to 7,730. A total of 7,224 households completed interviews in Wave 1. This represented ~83% of the 1993 population.

IFLS had an excellent tracking rate. For waves 2-5, the survey attempted to re-interview “origin” households – i.e., households from Wave 1 that contained at least 1 member of the Wave 1 interviewed household. If a household had moved, the survey attempted to locate them. The survey also contacted “split-off” households. These were households containing a target member of an original household (i.e. a wave 1 household member that was an individual panel respondent, or was over the age of 26, or other criteria). In Waves 2 through 5, a total of $n=6,821$, $n=6,661$, $n=6,596$, and $n=6,432$ of the 7,224 origin households, respectively, were re-contacted. We analyzed the sample of 2,265 Central Java households (origin and split-off) included across all waves.

5.3.1b Global Surface Water Datasets

The GSW measures long-term changes in surface water, globally, from Landsat satellite imagery gathered over decades²⁹. GSW is a collection of datasets produced by the European Commission's Joint Research Centre under the Copernicus Program for a broad array of uses including water resource management, conservation, and food security. The GSW collection includes datasets of surface water occurrence, change, seasonality, recurrence, transitions and maximum extent. The long timespan (over 3 decades), use of millions of satellite images, and 30-meter pixel resolution make GSW data products of high quality.

We used version 1 of GSW covering the timespan from 1984-2015, as it corresponded with the IFLS study period. GSW datasets do not vary by time. Rather, they capture the total changes throughout the period. We used the ‘Recurrence’ dataset to estimate the degree of recurrent flooding over the timespan. This variable captures inter-annual change, describing the frequency with which water returned from year to year. A detailed description of GSW variables is available in the associated paper²⁹.

5.3.2 Dataset Preparation

5.3.2a IFLS dataset

We created a hierarchical multilevel dataset, structured by household-wave. First, we conducted a crosswalk of IFLS modules and variables to compare across waves, select variables of interest, and identify discontinuities (e.g. where the survey instrument changed). We identified the Stata files needed to build our dataset from the raw data provided by IFLS. We examined the dataset structures and reshaped them to represent unique households where needed. IFLS provided aggregated consumption and expenditure data for waves 1-4 but not wave 5, which must be constructed by the researcher. Using the Stata script provided by IFLS, we constructed the consumption and expenditure aggregates for wave 5. Next, we merged individual data files from each wave by household. Then, we concatenated each of the waves into one dataset. We subset it to Central Java, including all Central Java households regardless of their time in the study (i.e., whether they appeared in 1 or multiple waves). The final dataset contained 6,743 observations, 2,265 unique households from Central Java across 5 survey waves.

5.3.2b IFLS and GSW Data Linkage

We used R Statistical Software (v4.2.2; R Core Team 2022) to obtain district-level summaries of GSW raster data to link to IFLS³¹. We downloaded GSW ‘recurrence’ data as GeoTIFF files. Next, we obtained vector (polygon) data of Indonesian subnational administrative boundaries from the Humanitarian Data Exchange. We extracted and summarized the raster values by the district polygons using the ‘exactextractr’ package. This produced a dataset of summary statistics for each Central Java district. Summary statistics included the sum of raster cells covered by each district and the fractions of cells corresponding to each value within the district. The results were saved as a Stata dataset and merged to IFLS by district. We summed GSW ‘recurrence’ fractions with values between 1-99 to obtain our district level measure of recurrent flooding (detailed below).

5.3.3 Definition of Flood and Outcome Variables

5.3.3a Recurrent Flooding

We approximated *recurrent flooding* using the ‘recurrence’ layer from GSW covering the period from 1984-2015. This variable was defined as the fraction of district raster cells with episodic flooding over the period (GSW recurrence values between 1-99). GSW recurrence ranges from 0 to 100 discrete values where 0 is not water, 1 is 1% recurrence over the period, and 100 is 100% recurrence. Values of 100 represent seasonal or permanent water bodies, while values from 1 to 99 represent a range of episodic flooding. By summing the fractions of the districts with recurrence values between 1-99, we obtained a variable representing the total portion of the district that had episodic (i.e. non-seasonal, or unpredictable) flooding over the period. We assume that districts with a larger portion of episodic flooding over the period were affected by more floods.

The current body of literature examining recurrent flood events and household food security longitudinally is limited. Therefore, similar metrics from which to base our definition of recurrent flooding used in this study are lacking and our metric is therefore considered exploratory.

5.3.3b Outcomes

Our outcomes of interest are widely used indicators constructed from household consumption and expenditure data and based on the Data4Diets framework⁷. Our measure of diet quantity was household food expenditure share, expressed as the percentage of total spending spent on food. This variable was calculated by dividing the total household expenditure spent on food by the total household expenditure, multiplied by 100. It has been well-documented that the share of income spent on food increases with a household's vulnerability, known as Engel's law⁷. Accordingly, with rising income, non-food expenditure increases more than food expenditure, thus decreasing food expenditure share³²⁷.

Our diet quality measure was household dietary diversity score, expressed as an integer from 0-12 (low to high). This variable was calculated as the total number of food groups consumed by the household in the past week as follows: cereals, roots and tubers, vegetables, fruits, meat/poultry/offal, eggs, fish and seafood, pulses/legumes/nuts, milk and milk products, oils/fats, sugar/honey, miscellaneous. This selection of food groups is based on prior work by FANTA (Food and Nutrition Technical Assistance)³³. This serves as a useful indicator of diet quality as it is known that a more diverse diet is associated with a range of positive outcomes including improved anthropometry, caloric and protein adequacy, and income^{7,33}.

Together, both of these indicators serve as proxies of household food access, measuring improved household food consumption³³.

5.4 Statistical and Machine Learning Methods

We took a hybrid approach to our modelling. Our main interpretation is taken from statistical models, to explain the relationships in our data. However, we also used a machine learning method – Random Forest algorithm – to assess predictive accuracy and add context to our statistical findings.

Our selection of covariates was based on a conceptual framework from prior evidence relating these factors to food security (**Figure 1**)^{3,7,33}. These factors may also be determinants of living in more heavily flooded areas. We were unable to include a variable indicating whether the community had a disaster risk management plan as we did not have the data. We excluded level of education, as there was a high percentage (>20%) of missing values.

We used Stata 17 (Stata Statistical Software: College Station, TX, USA: Stata Corp LP) for all analyses. We constructed a wealth index to measure socioeconomic status using Principal Component Analysis, based on a methods guidance paper from the World Food Programme³⁴. We checked assumptions using Bartlett's test for sphericity and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy. Variables with low factor loadings or related in the opposite direction were removed. A listing of the variables that were included in the wealth index are shown in the descriptive table (**Table 1**).

Our dataset contained 3 levels of hierarchy: observations made at each cross-sectional survey wave (level 1), for each household (level 2), within each district (level 3). We included two random effects in our statistical models to account for the correlation by household and district, described by the following notation:

$$f(Y_{itsv}) = \beta_0 + \beta_1(f(X)_s) + \beta_2(f(Z)_{itsu}) + \beta_3K_i + b_s + b_{is} + \varepsilon_{its}$$

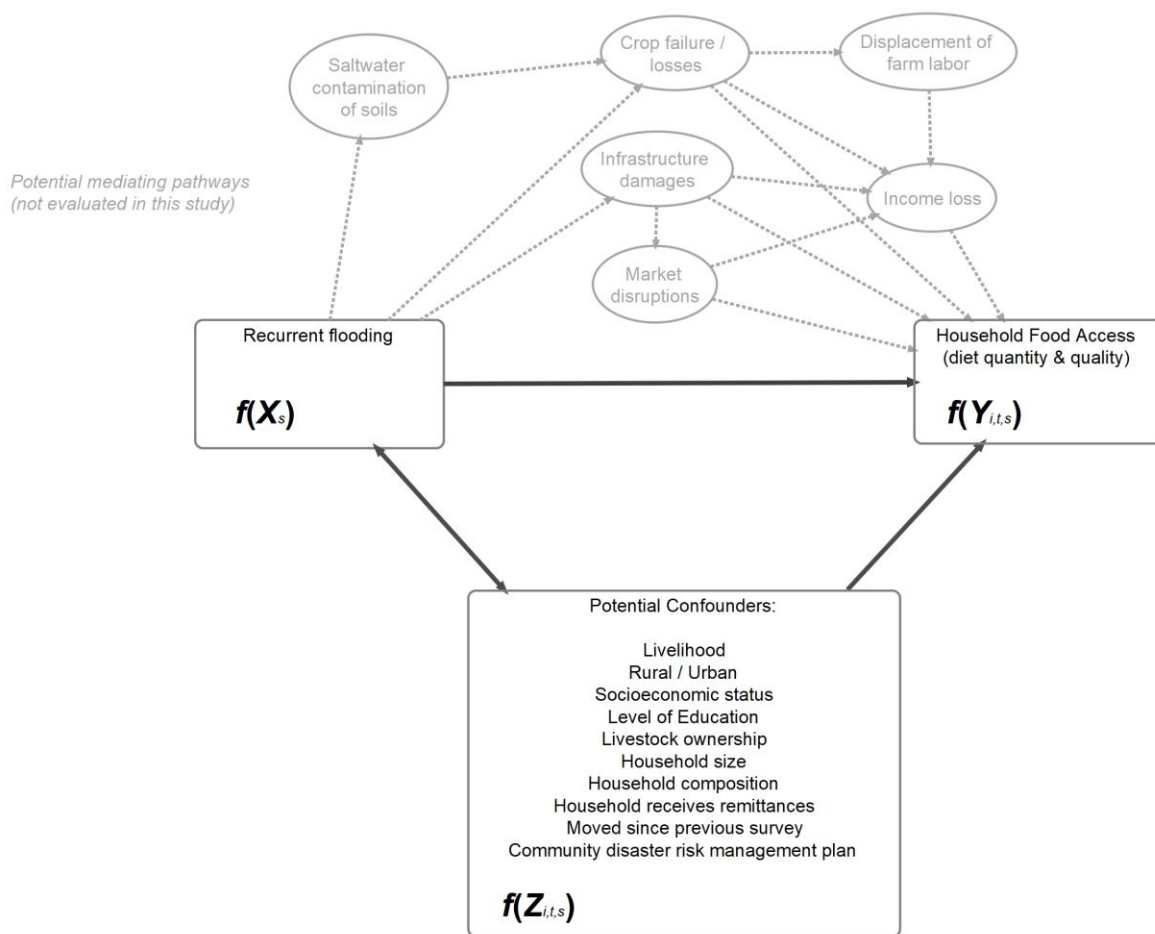
In the above equation, $f(Y_{itsv})$ represents a function of the outcome v for household i at time t within district s ; $\beta_1(f(X)_s)$ represents the coefficient for a function of the flood variable within district s ; $\beta_2(f(Z)_{itsv})$ represents the coefficient for a function of the confounding factors, u , displayed in **Figure 1** for household i at time t within district s . $\beta_3 K_i$ represents the coefficient for the survey wave for household i . The random effects b_s and b_{is} vary across districts and households within districts, respectively.

We fit linear mixed regression models and examined specification errors using the link test, multicollinearity of independent variables using variance inflation factor, and residual plots. We checked for outliers using Cook's distance after fixed effect models. For the dietary diversity score outcome, there was specification error and evidence of heteroscedasticity. We first tried transformations of the dependent variable which improved but did not remedy the heteroscedasticity. Therefore, we present ordered logistic mixed regression for the dietary diversity score outcome. We present a proportional odds model with mixed effects, although this assumption was not met for all variables. We explored nonlinearity using a purely functional approach, by fitting linear and quadratic terms of our continuous flood variable. We used Likelihood Ratio tests to determine the added value of the quadratic terms. We conducted complete case analysis as there was little missing data (<5%).

We used the following indicators to assess adaptive capacity, which were only available in waves 3 to 5: knowledge of where to borrow money, secured a loan in past 12 months, received any assistance in past 12 months, received cash assistance. We examined their joint effects with flooding in separate models.

We conducted Random Forest regression models for both outcomes. To determine the optimal number of iterations to include in the model, we examined scatter plots of out-of-bag error versus varying number of iterations. For both models, error converged at ~400 iterations. We present results in Variable Importance Plots.

Figure 1. Conceptual framework



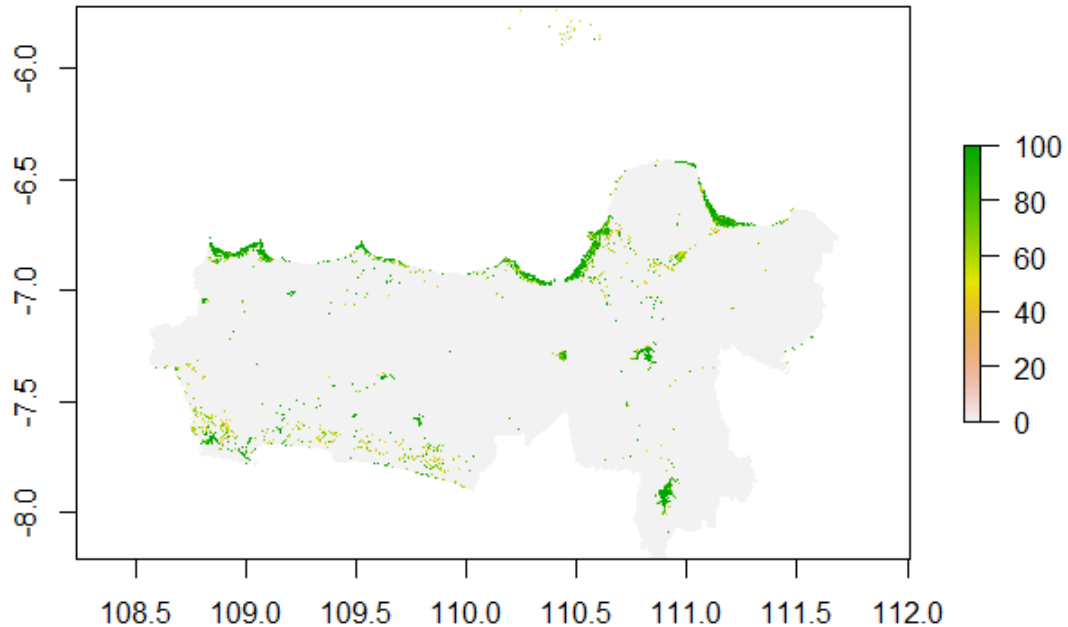
5.5 Results

Of 2,265 Central Java households, 856 were sampled in Wave 1, 998 in Wave 2, 1,325 in Wave 3, 1,616 in Wave 4, and 1,948 in Wave 5. The tracking rate was 97% (827 of 856 origin households in Wave 5). The rest of the sample were “split-off” households (new households containing a target member of an original household). The sample can

be characterized as about half rural, a third agricultural, mostly homeowners, living with moderate size yards and improved water sources and facilities (**Table 1**).

Recurrent flooding was present in most districts (79%) throughout the period, with 80% of households living in affected districts. On average, small portions of the districts were affected (median percentage of district raster cells with episodic flooding=0.5%). The most heavily affected areas in the top decile had larger portions of their districts affected (between 10-31%). Districts with the most recurrent flooding were: Kaliwungu, Kedung, Juwana, Pekalonganutara. All of these were coastal districts on the north side of Central Java. **Figure 2** shows recurrence values plotted on a map of Central Java. The distribution among households in our sample was highly right skewed (skewness=2.14, kurtosis=6.68).

Figure 2. Map of Central Java showing Global Surface Water Recurrence Values 1984-2015¹



¹Recurrence values range from 0-100, where 0 is not water, 1 is 1% recurrence over the period, and 100 is 100% recurrence (seasonal or permanent water bodies).

Our regression models included 95% and 97% of the total sample for food expenditure share and dietary diversity score, respectively (**Table 2**). We did not detect any linear relationship or threshold effect (i.e. nonlinearity) between recurrent flooding and either of the outcomes. Wealth, access to credit, and home ownership were associated with improvements in both outcomes. Livestock ownership, household size, and living in a rural area were associated with worsening food expenditure share but higher diet diversity (**Table 2**).

When we examined recurrent flooding jointly with measures of adaptive capacity, we found associations with wealth and access to credit. Wealth was associated with improved food expenditure share and this effect was strengthened in households in more flood prone districts. In contrast, access to credit was associated with higher odds of a more diverse diet but the direction of the effect changed to reduced odds in more flood prone districts (**Table 2**).

In random forest models, flooding had more predictive importance for food expenditure share than for diet diversity in relation to the other variables (**Figure 3**).

Figure 3. Variable Importance Plots

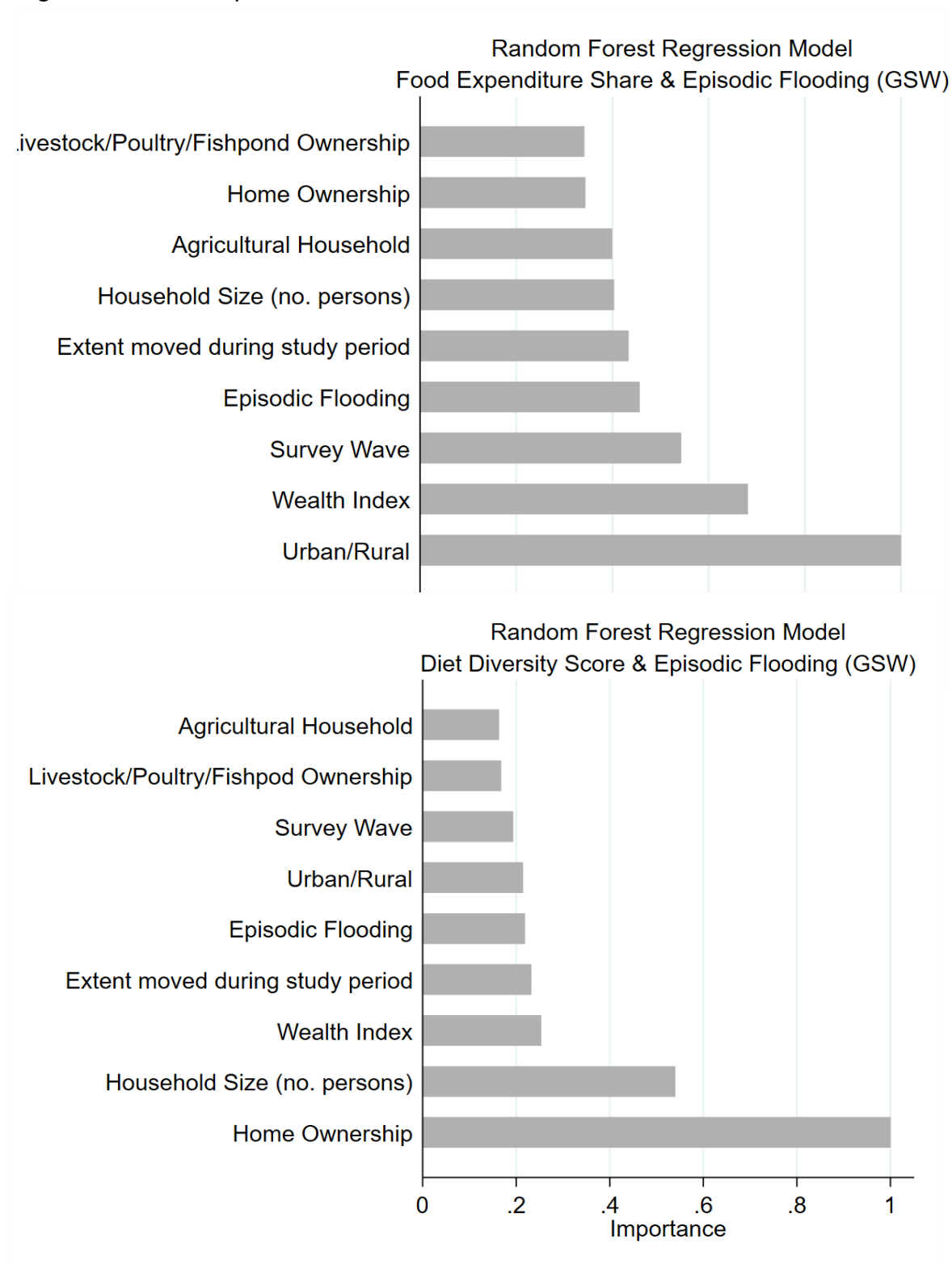


Table 1. Descriptive characteristics of Central Java households¹

Characteristic	Variable	First Wave	Last Wave
Sample size		2265	2265
Livelihood	Has land for farming ³	269 (22%)	613 (29%)
	Anyone was a farmer in the past year	755 (34%)	682 (32%)
	Anyone engaged in fishing (non-fishery) ³	10 (9%)	37 (6%)
Livestock ownership	Owns livestock/poultry/fishpond	619 (27%)	458 (20%)
Highest level of education ²	Elementary	507 (35%)	468 (23%)
	Junior High General	311 (21%)	463 (23%)
	Junior High Vocational	10 (1%)	5 (0%)
	Senior High General	187 (13%)	282 (14%)
	Senior High Vocational	192 (13%)	375 (19%)
	College	75 (5%)	104 (5%)
	University (Bachelor/Master/Dr)	171 (12%)	298 (15%)
Household composition	Household size	3.74 (1.89)	3.37 (1.66)
Housing characteristics	Rural	1201 (53%)	1129 (50%)
	Number of rooms ⁴	5.44 (2.53)	5.95 (2.50)
	House is self-owned	1575 (70%)	1569 (74%)
	Uses electricity ⁴	1969 (87%)	2106 (99%)

Flooring	Ceramic/marble/granite/stone or tiles or cement/bricks or lumber/board ⁴	1579 (70%)	1994 (88%)
Walls	Masonry (cement/prefabricated bricks) or lumber/board/plywood ⁴	1973 (87%)	2192 (97%)
Roof	Concrete or roof tiles/shingles ⁴	1947 (86%)	1958 (86%)
Drinking/cooking water	Improved source (piped, well, spring, rain)	1883 (84%)	1658 (78%)
Bath/laundry water	Improved source (piped, well, spring, rain) ⁴	2017 (89%)	2033 (96%)
Toilet type	Improved facility (own toilet, septic system) ⁴	1392 (62%)	1716 (81%)
Sewage disposal	Improved system (drainage ditch, permanent pit) ⁴	1536 (68%)	1583 (75%)
Garbage disposal	Collection by sanitation service ⁴	549 (24%)	650 (31%)
Assets (ownership by any household member)	House and land occupied by household	1596 (71%)	1573 (74%)
	Other house or building (including land) ⁴	151 (7%)	180 (8%)
	Vehicles (cars, boats, bicycles, motorbikes) ⁴	1168 (52%)	1481 (70%)
	Household appliances (radio, fridge, TV, etc.) ⁴	1775 (79%)	1960 (92%)
	Jewelry ⁴	1236 (55%)	1015 (48%)
	Savings ⁴	684 (30%)	728 (34%)
	Receivables ⁴	285 (13%)	235 (11%)
Access to credit	Know of a place to borrow money ³	1236 (88%)	1852 (82%)
	Ever borrowed money ³	42 (27%)	2 (11%)
	Secured a loan (past 12 months) ³	275 (22%)	608 (27%)
Household assistance	Received money from a community group ³	6 (2%)	0 (0%)

	Received cash assistance ³	149 (12%)	598 (27%)
	Received any assistance from government or NGO ³	455 (36%)	1295 (58%)
Expenditure	Monthly household expenditure	1211335.75 (1756339.13)	2818259.00 (2401557.50)
Interviewer observation of house	Moderate size yard ⁴	1344 (59%)	1326 (59%)
	Well-kept yard ⁴	1735 (77%)	1772 (78%)
	Adequate ventillation ⁴	1762 (78%)	1839 (81%)
	Stable under/next to house	447 (20%)	293 (13%)
	House surrounded by puddles	148 (7%)	161 (7%)
	Piles of trash around house	159 (7%)	133 (6%)
	Human and animal waste near house	133 (6%)	106 (5%)
Wealth index	Score	-0.28 (2.01)	0.60 (1.42)
	Lowest quintile	580 (26%)	182 (9%)
	Mid-low quintile	429 (19%)	348 (16%)
	Medium quintile	418 (19%)	478 (23%)
	Mid-high quintile	395 (18%)	578 (27%)
	Highest quintile	424 (19%)	535 (25%)
Community/village characteristics ²	Coastal	132 (10%)	126 (10%)
	Blocked water duct or obstructed river/gutter	320 (23%)	258 (20%)
	Standing pools of water (not marshes, lakes)	329 (24%)	397 (30%)

¹ Statistics presented are mean (SD) or n (%). Less than 5% missing for all variables unless otherwise indicated.

² >25% missing. Variables not included in modelling.

³ Variable not collected across all waves. Not included in full models.

⁴ Variable included in wealth index derived using principal component analysis, excluding household assets and characteristics with low factor loadings or related in the opposite direction.

Table 2. Regression results¹

	Food Expenditure Share (% of total spending)			Dietary Diversity Score (0 to 12)		
	Coef.	SE	P-value	OR	SE	P-value
<i>Recurrent flooding measure</i>						
Fraction of district with episodic flooding (1984-2015) ²	5.18	7.11	0.467	1.30	1.57	0.830
<i>Household characteristics</i>						
Anyone was a farmer in the past year	0.60	0.49	0.223	1.10	0.08	0.187
Owns livestock/poultry/fishpond	1.06	0.45	0.020	1.29	0.08	0.000
Household size	0.25	0.12	0.032	1.32	0.02	0.000
Rural	2.60	0.72	0.000	1.35	0.16	0.011
House is self-owned	-1.12	0.58	0.055	1.89	0.16	0.000
Wealth index	-2.89	0.16	0.000	1.27	0.03	0.000
Wealth index (quadratic effect)	-0.42	0.05	0.000			
Wealth index (joint effect with fraction of district flooded episodically)	-4.56	2.17	0.035			
Extent household moved during entire study period						
Did not move	Ref.			Ref.		
Moved within the same village	-0.43	0.75	0.570	1.09	0.13	0.503
Moved within the same district	0.96	0.97	0.324	1.24	0.20	0.178
Moved within the same regency	-0.98	0.87	0.257	0.88	0.13	0.372
Moved within the same province	-1.48	0.92	0.106	0.44	0.07	0.000
Moved to another province	-2.23	0.97	0.022	0.86	0.14	0.345
<i>Survey wave</i>						
Wave 1 - 1993	Ref.			Ref.		
Wave 2 - 1997	0.11	0.70	0.875	1.34	0.13	0.002
Wave 3 - 2000	4.11	0.69	0.000	1.94	0.19	0.000
Wave 4 - 2007	2.31	0.70	0.001	1.36	0.13	0.001

Wave 5 - 2014	-1.85	0.71	0.009	1.05	0.10	0.630
---------------	-------	------	-------	------	------	-------

Indicators of adaptive capacity (waves 3 to 5 only)⁴

Linear effects (model with no interactions)

Know of a place to borrow money	-1.83	0.81	0.023	1.47	0.18	0.001
Secured a loan in past 12 months	-0.63	0.52	0.226	1.45	0.11	0.000
Received any assistance from government or non-governmental organization in past 12 months ⁵	1.58	0.64	0.014	1.33	0.13	0.003
Received cash assistance	-0.19	0.63	0.767	0.77	0.07	0.006

Differential/joint effects with flooding

Know where to borrow & fraction of district flooded episodically	-5.84	11.80	0.620	0.01	0.02	0.010
Secured a loan & fraction of district flooded episodically	-1.03	8.57	0.904	0.15	0.20	0.147
Received any assistance & fraction of district flooded episodically	-1.29	7.39	0.861	4.16	4.48	0.184
Received cash assistance & fraction of district flooded episodically	8.77	9.22	0.342	3.10	4.14	0.398

¹ Complete case analysis, n=6,412 (95% of total sample) for food expenditure share outcome and n=6,513 (97% of total sample) for dietary diversity score outcome.

² Calculated from Global Surface Water Recurrence values. Describes fraction of district raster cells with recurrence values between 1-99. Likelihood Ratio test comparing linear and quadratic models was not significant. Quadratic term removed due to collinearity.

³ Joint effects not included due to non-significant Likelihood Ratio test.

⁴ These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining the linear effects of all four indicators of adaptive capacity and each of their joint effects with flooding. Each of the models included complete case analysis of n=4,579 observations (94% of total sample in waves 3 to 5).

⁵ In the form of money, rice, vegetables, sugar, fruit, meat, snacks, other food, cooking oil, or kerosene.

5.6 Discussion

In this study we explored a measure of flood recurrence to evaluate its impact on household, diet-related aspects of food security over time in Central Java. The flood variable in this study measured the portions of districts affected by episodic flooding during the period (1984-2015). Our findings indicate that wealthier households in more flood affected districts may have an advantage in terms of their ability to purchase food. Access to credit, however, which is independently associated with improved diet diversity, showed reduced diet diversity in more flooded affected districts. This could be because resources are diverted elsewhere in these areas.

We did not detect a linear, independent effect between recurring flooding and either of the outcomes. This could be due to measurement challenges related to the aggregation and scale of the variable or because the relationship has a different form and shape. We aggregated all GSW recurrence values from 0-99, as we did not have a prior basis from which to measure or define this variable. We tried log transforming this variable to see if it improved linearity, but it did not appear to. Since there were no substantive changes in conclusions drawn, we presented the un-transformed form for interpretability. The scale of the variable was at the district rather than household level. Identifying flood affected households is another measurement challenge. A future direction could be to explore disaggregation of this variable further. More work is needed to compare GSW data to other flood databases and validate these measures. Nevertheless, high-resolution data like GSW are promising and provide a good quality source to link to surveys versus other data that must be downscaled.

We explored other data products to measure flooding, including the Dartmouth Flood Observatory, which we successfully linked to the survey. However, this data was very coarse, cutting across multiple districts, and had to be aggregated down to the district level to link. This is problematic for analysis as there are likely large differences across districts in terms of flooded areas which may not be accurately depicted with such a variable. Using this variable at a larger scale may be more informative. The Global Flood Awareness System is planning to introduce a flood recurrence layer in a future

update. Such products will provide an important advantage over flood occurrences to understand impacts over time.

Previous efforts to measure recurrent flooding have varied, including direct interviews of affected households or indirect comparisons of affected and unaffected communities, or linkages of household and spatial data^{6,13,24,35}. But, while whole communities may be affected, individual households may be impacted differently. Identification of flood affected households remains a challenge. Direct questioning of households about how they are impacted by floods provides advantages and disadvantages. This information, however, is lacking over a long period of time. Thus, spatial and survey linkages provide the best opportunity to study these impacts, but these linkages are challenging. We presented and demonstrated a method for linking raster data to a household survey. But more disaggregated disaster data are needed¹.

In other similar flood-prone contexts, a common coping or adaptation strategy in response to flooding is to borrow money^{8,19}. We found this to be associated with reduced odds of improved dietary diversity. Evidence from Central Java suggests that people are staying in flood-prone areas for various reasons^{24,26,36}. During flood events, people are often unable to work and cook³⁶. We found that wealthier households in flood prone districts had improved food expenditure share. Taken together, these findings suggest that already vulnerable households may be even more disadvantaged in terms of food access when faced with flooding, while wealthier households may be better able to adapt.

There were some data quality issues with the food access variables used in this analysis. The reference period for the food recalls used to create household dietary diversity score was 1 week, which may have led to less accurate recall than a shorter period (typically 24-hours). The recall did not include a question about whether consumption during the period was normal. We tried creating other metrics including household average dietary energy acquisition or consumption, food consumption score,

household adequacy of fruit and vegetable consumption, household share of animal protein in total protein consumption, household share of dietary energy from macronutrients, household share of energy consumed from non-staples. However, these variables could not be constructed due to lack of granularity with the food list. Some variables were not included in our models due to high missingness, including education level and community level variables, which could have been important predictors. We did not explore reasons for missingness in this study. We wanted to include a variable of whether the community had a disaster risk management plan but were unable to obtain this data. We aggregated the flood values, which could have masked the relationships in our data. The use of this data for this purpose should be further explored.

We demonstrated how linkages can be accomplished with high resolution data about the environment and survey data about the population, creating new possibilities for research. Our findings indicate that access to credit, a factor that ordinarily improves food access, may not be effective in flood-prone areas for already vulnerable households.

Wealthier households may be better able to adapt in terms of food access. These relationships are far more complex, and further investigation will shed more light. Disaggregated disaster data products are needed to further this important area of study.

5.7 References

1. Food and Agriculture Organization of the United Nations. *The impact of disasters and crises on*

agriculture and food security: 2021. (2021).

2. Food and Agriculture Organization of the United Nations. Agriculture on the Proving Grounds: Damage and loss. <https://www.fao.org/resources/digital-reports/disasters-in-agriculture/en/>.
3. Fanzo, J., Davis, C., McLaren, R. & Choufani, J. The effect of climate change across food systems: Implications for nutrition outcomes. *Glob. Food Sec.* **18**, 12–19 (2018).
4. Salm, L., Nisbett, N., Cramer, L., Gillespie, S. & Thornton, P. How climate change interacts with inequity to affect nutrition. *Wiley Interdiscip. Rev. Clim. Chang.* **12**, e696 (2021).
5. Team, C. W. *IPCC, 2023: Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.*
6. Niles, M. T. *et al.* Climate impacts associated with reduced diet diversity in children across nineteen countries. *Environ. Res. Lett.* **16**, 15010 (2021).
7. Data4Diets: Food Security Indicators | INDDEx Project. <https://inddex.nutrition.tufts.edu/data4diets/indicators>.
8. Gabrysch, S. *et al.* Nutritional effects of flooding due to unseasonably early monsoon rainfall in Bangladesh: a cross-sectional study in an ongoing cluster-randomised trial. *Lancet Planet. Heal.* **2**, S3 (2018).
9. Oskorouchi, H. R. & Sousa-Poza, A. Floods, food security, and coping strategies: Evidence from Afghanistan. *Agric. Econ. (United Kingdom)* **52**, 123–140 (2021).
10. Doocy, S., Leidman, E., Aung, T. & Kirsch, T. Household economic and food security after the 2010 Pakistan floods. *Food Nutr. Bull.* **34**, 95–103 (2013).
11. Ambikapathi, R. *et al.* La Niña weather impacts dietary patterns and dietary diversity among children in the Peruvian Amazon. *Public Health Nutr.* **24**, 3477–3487 (2021).
12. Phalkey, R. K., Aranda-Jan, C., Marx, S., Höfle, B. & Sauerborn, R. *Systematic review of current efforts to quantify the impacts of climate change on undernutrition.* vol. 112 E4522–E4529 (National Academy of Sciences, 2015).

13. Rodriguez-Llanes, J. M., Ranjan-Dash, S., Degomme, O., Mukhopadhyay, A. & Guha-Sapir, D. Child malnutrition and recurrent flooding in rural eastern India: a community-based survey. *BMJ Open* **1**, e000109 (2011).
14. Gaire, S., Delbiso, T. D., Pandey, S. & Guha-Sapir, D. Impact of disasters on child stunting in Nepal. *Risk Manag. Healthc. Policy* **9**, 113 (2016).
15. Rodriguez-Llanes, J. M., Ranjan-Dash, S., Mukhopadhyay, A. & Guha-Sapir, D. Flood-Exposure is Associated with Higher Prevalence of Child Undernutrition in Rural Eastern India. *Int. J. Environ. Res. Public Health* **13**, 210 (2016).
16. Lieber, M., Chin-Hong, P., Kelly, K., Dandu, M. & Weiser, S. D. A systematic review and meta-analysis assessing the impact of droughts, flooding, and climate variability on malnutrition. *Glob. Public Health* **17**, 68–82 (2022).
17. Alderman, K., Turner, L. R. & Tong, S. Floods and human health: a systematic review. *Environ. Int.* **47**, 37–47 (2012).
18. Ahern, M., Kovats, R. S., Wilkinson, P., Few, R. & Matthies, F. Global Health Impacts of Floods: Epidemiologic Evidence. *Epidemiol. Rev.* **27**, 36–46 (2005).
19. Gepp, S. *et al.* Impact of unseasonable flooding on women's food security and mental health in rural Sylhet, Bangladesh: a longitudinal observational study. *Lancet Planet. Heal.* **6**, S14 (2022).
20. Parvez, M., Islam, M. R. & Dey, N. C. Household food insecurity after the early monsoon flash flood of 2017 among wetland (Haor) communities of northeastern Bangladesh: a cross-sectional study. *Food Energy Secur.* **11**, e326 (2022).
21. Marfai, M. A. *et al.* Natural hazards in Central Java Province, Indonesia: an overview. *Environ. Geol.* **56**, 335–351 (2008).
22. Marfai, M. A. & King, L. Tidal inundation mapping under enhanced land subsidence in Semarang, Central Java Indonesia. *Nat. Hazards* **44**, 93–109 (2008).
23. Lassa, J. A. *Emerging 'Agricultural Involution' in Indonesia: Impact of Natural Hazards and Climate Extremes on Agricultural Crops and Food System.* (2012).
24. Harwitasari, D. & van Ast, J. A. A. Climate change adaptation in practice: people's responses to tidal

- flooding in Semarang, Indonesia. *J. Flood Risk Manag.* **4**, 216–233 (2011).
25. Harwitasari, D. & van Ast, J. A. *Adaptation responses to tidal flooding in Semarang, Indonesia*. (Erasmus University, 2008).
 26. Langlois, B. K. *et al.* Household Flood Severity and Migration Extent in Central Java: Analysis of the Indonesian Family Life Survey. *Int. J. Environ. Res. Public Health* **20**, (2023).
 27. Abdullah, F., Marzuki, M., Yanis, M. & Ismail, N. Food security strategies toward flood hazards along the Cinendang river, Aceh Province. *IOP Conf. Ser. Earth Environ. Sci.* **667**, (2021).
 28. *Indonesian Family Life Survey (IFLS) Data and Documentation* | RAND.
 29. Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016).
 30. Global Surface Water Explorer. <https://global-surface-water.appspot.com/#features>.
 31. R Core Team (R Foundation for Statistical Computing). R: A Language and Environment for Statistical Computing. (2021).
 32. Lele, U. *et al.* Measuring Food and Nutrition Security: An Independent Technical Assessment and User's Guide for Existing Indicators. (2016).
 33. Swindale, A. & Bilinsky, P. *Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (Version 2)*. (2006).
 34. Hjelm, L., Mathiassen, A., Miller, D. & Wadhwa, A. VAM Guidance Paper: Creation of a Wealth Index. (2017).
 35. Baeribameng Yiran, G. A., Atubiga, J. A., Kusimi, J. M., Kwang, C. & Owusu, A. B. Adaptation to perennial flooding and food insecurity in the Sudan savannah agroecological zone of Ghana. *Environ. Res.* **214**, 114037 (2022).
 36. Marfai, M. A. *et al.* The impact of tidal flooding on a coastal community in Semarang, Indonesia. *Environmentalist* **28**, 237–248 (2008).

5.8 Description of all other exploratory modelling

I conducted other exploratory modelling comparing recurrent flooding from different perspectives -- large, extreme events versus small, gradual shifts in surface water – as captured by DFO and GSW. Results of these regression models are shown in **Appendix 7**. Few studies have measured recurrent flooding, representing an important knowledge gap. Use of existing flood databases for this purpose is necessary but is also a new area of exploration. Similar metrics from which to base a definition are lacking in the literature. These metrics are therefore exploratory but are an important step forward for how we can address questions relating to natural disasters and environmental changes and their impact on public health and nutrition through multidisciplinary data linkages.

The DFO historical archive captures floods that are considered large, causing significant damage, and that gather media attention dating back to 1985. In contrast, the GSW captures smaller scale changes in surface water across the landscape since 1984. More detail about these databases is provided in Chapter 3. One important distinction is the spatial aggregation. Events recorded by DFO affect large areas, cutting across multiple districts, which must be downscaled to link to the IFLS. The process of downscaling such data to a smaller, more disaggregated level is challenging and may represent a less accurate depiction of the landscape compared to high resolution data which can be more precisely summarized to a larger level of aggregation.

In section 5.3.2, I described the methods of linking high-resolution, pixelated GSW data to the IFLS. Below I detail the process of linking coarse, polygon DFO data to the IFLS.

5.8.1 Linking of DFO archive data to IFLS

DFO data was linked to IFLS through the following steps. First, I spatially joined the DFO flood data to Central Java administrative boundaries using ArcGIS. DFO provides a shapefile of its historical archive with polygons representing the areas affected for each flood event. This file was read into ArcGIS software with Central Java administrative boundaries corresponding to 2020 codes from the Indonesian government. The projection of the DFO layer was set to GCS_WGS_1984 to align with the administrative boundaries. The DFO layer was then subset to Indonesia using the 'select by attribute' feature and saved as a separate shapefile. The spatial join was done using the 'join' data management tools in Arc Toolbox. Administrative boundary data (level 4) was joined with the flood archive data to create a joined table containing all administrative boundaries that the flood events spanned.

The spatially joined layer was then exported as a shapefile and read into Stata software for further manipulation. IFLS provides a crosswalk file of geographic boundaries (i.e. the boundary codes and names from all waves matched to each other) to facilitate merging from Wave 1 in 1993 to Wave 5 in 2015. Using this file, I matched DFO and IFLS districts by obtaining the names of IFLS districts (administrative level 3) which were then linked to DFO districts (from the spatial join described previously). I then changed the level of observation to create summary variables – the number of flood events and the days of flooding – by district and year. Finally, the data was merged to IFLS by district and survey wave.

Merging DFO to IFLS was challenging due to shifts in population dynamics over the 1985-2015 timespan and changing geographic boundaries. While IFLS provided a crosswalk file of geographic boundaries, I was unable to obtain an administrative boundary shapefile corresponding to the 2015 codes for the DFO spatial join step. This

further complicated the process and required comparing and manually reviewing districts that did not match.

5.8.2 Exploration of variables to measure recurrent flooding

I explored four flood variables: 1) the total number of flood events from DFO, 2) the duration (in days) of flooding from DFO, 3) surface water recurrence from GSW, and 4) occurrence change intensity from GSW. I explored indicator, ordinal, and continuous or count forms of the variables due to the skewness of the distributions.

The DFO measures were count variables defined as the total number of large (including extreme) flood events causing significant agricultural or structural damages that occurred in the district since the previous survey. The duration of flooding was derived from the start and end date of these flood events. While the DFO variables vary over time, GSW variables measure the changes throughout the entire period 1984-2015. The GSW recurrence variable (described in section 5.3.3) is defined as the fraction of district raster cells with episodic flooding over the period. GSW recurrence ranges from 0 to 100 discrete values. Values of 0 are not water, while values of 100 represent seasonal or permanent water bodies, and values from 1 to 99 represent a range of episodic flooding. The GSW occurrence change intensity variable measures where water occurrence decreased, did not change, or increased throughout the period. This variable is defined as the fraction of district raster cells with an increase in surface water over the period. GSW occurrence change intensity ranges from 0 to 200 discrete values, where values between 101 and 200 represent areas that had an increase in surface water.

5.8.3 Statistical modelling and nonlinearity

The modelling approach was described in section 5.3.4. Briefly, I fit linear (and nonlinear) mixed-effects regression models for the household food expenditure share outcome and ordinal logistic mixed-effect regression models for the dietary diversity score outcome after noting specification error and evidence of heteroscedasticity in linear mixed-effect models. I examined joint effects with adaptive capacity variables: wealth, access to credit, securing a loan, receiving assistance, receiving cash assistance.

I explored nonlinearity using two approaches. First, by fitting quadratic terms of the continuous and count flood variables to the models. Secondly, I used a non-parametric approach, ignoring the functionality by modelling ordinal categories of the flood variable.

5.8.4 Findings

All exploratory models are presented in **Appendix 7** in **Tables S1-S6**. For all models, I conducted complete case analysis.

5.8.4a Household food expenditure share

The DFO variables showed no evidence of linear or nonlinear relationships to household food expenditure share, but joint effects were detected with wealth, having access to credit (i.e. knowing where to borrow money), and securing a loan (**Table S1**). Interestingly, wealth and access to credit were associated with decreased (i.e. improved) food expenditure share, but the direction of the coefficients changed with more days of flooding. In other words, in districts with increasing days of flooding, wealth was associated with a small but significant increase in food expenditure share. Similarly, there was a marginally significant differential effect with access to credit and

increasing days of flooding. Compared to households that had access to credit in districts with no days of flooding, those in districts with 2 to 7 or 34 or more days of flooding had a large increase in food expenditure share. A differential effect was also detected with securing a loan and increasing number of flood events. Households that secured a loan in districts with increasing number of floods had a small but significant increase in food expenditure share.

The GSW variables showed some evidence of linear relationships to increasing household food expenditure share (**Tables S2 and S3**). Living in districts with an increase in surface water throughout the period was associated with an increase in food expenditure share. Similarly, living in districts with the most episodic flooding was associated with increased food expenditure share. Joint effects were detected with both increase in surface water and episodic flooding. In districts with the largest increase in surface water throughout the period, wealth was associated with a small decrease in food expenditure share. Similarly, in districts with more episodic flooding, wealth was associated with a large decrease in food expenditure share. In contrast, joint effects with access to credit and any episodic flooding and receiving assistance and any increase in surface water were associated with increased food expenditure share.

5.8.4b Dietary diversity score

Linear and nonlinear effects were detected with the DFO variables and dietary diversity score (**Table S4**). Interestingly, the days of flooding and number of flood events showed increased odds of having a more diverse diet. The quadratic, nonlinear effect of number of flood events showed slightly reduced odds of a more diverse diet. This means that the number of floods was associated with improved diet diversity up to a threshold point where the relationship shifted to reduced diet diversity with more flooding. Joint effects were not detected.

Linear effects were also detected with the GSW indicator variables (**Tables S5 and S6**). Living in districts with any increase in surface water or any episodic flooding was associated with increased odds of a more diverse diet. Nonlinear effects were not detected. Joint effects were detected with wealth and access to credit. Wealth was associated with increased odds of a more diverse diet, but this effect shifted to decreased odds in districts with increases in surface water and any episodic flooding. Likewise, access to credit was associated with increased odds of a more diverse diet but shifted to reduced odds in districts with higher increase in surface water and more episodic flooding.

Chapter 6: Conclusions

This study provides important methods for creating multidisciplinary datasets that other researchers can replicate to assess questions which were previously not possible about how the environment shapes household food security. Several challenges exist in the measurement of floods, including identifying flood affected households, capturing the type of flood, and classifying severity. I explored the relationship between recurrent flooding and household food access from two perspectives using different databases over a span of 3 decades. Some relationships emerged from these models that can be explored further in future studies.

The primary variable approximated recurrent flooding by measuring small, gradual increases in flooding over the period using the Global Surface Water dataset (the fraction of districts flooded episodically between 1984-2015). Living in districts with the highest flooding throughout the period was associated with increased food expenditure share – or worsening ability to purchase food. Wealth and access to credit were independently associated with improvements in both outcomes – ability to purchase foods and diet diversity. When examined jointly, the protective effect of wealth on ability to purchase food was strengthened in more flood affected districts. In contrast, access to credit was associated with reduced odds of a more diverse diet – or poorer-quality diet – in more flood affected districts. These findings indicate that, when faced with gradual increases in flooding, wealthier households may be more advantaged in terms of their ability to purchase food while those that rely on credit may be reducing the quality of their diet in response.

I also explored large, extreme flooding events from the Dartmouth Flood Observatory. These models showed some different effects than the models using the Global Surface Water measure. Households in districts with more large flood events had *higher* odds of a more diverse diet – or better-quality diet. This could be because of the

rainfall associated with floods, which has been related to improved food security and diet diversity in other studies, or because of an increased water supply for irrigation. The impacts likely differed across districts with some areas negatively affected (e.g. low-lying or coastal) and others positively affected (e.g. at higher elevation), but these distinctions could not be determined with the district level measure. There was a nonlinear effect where more floods were associated with higher odds of a more diverse diet up to a threshold point, where the relationship plateaued. Wealth and access to credit were independently associated with improved ability to purchase food, but the direction changed with longer durations of flooding. Securing a loan was associated with worsening ability to purchase food in areas with more flooding.

In summary, household food access is likely impacted by recurrent flooding over time and these impacts may be both positive and negative depending on the scale (local versus national, small versus large flood), type of flood, and level of recurrence. These relationships are complex. Measures of flood recurrence and other characteristics are needed to disaggregate these effects and explore more deeply.

This study found households that access loans or credit and that live in areas with more flooding may have reduced food access. Wealth may be protective in areas with smaller, gradual increases in flooding, but not in areas with recurrent large flood events. We need more studies to understand what these effects mean or what effect sizes are meaningful. This research points to important future directions. Incorporating land use data will help uncover how different locales are affected. Gridded, global land cover products derived from satellite imagery are currently available and should be explored for this purpose. Another important direction is to further understand the complexity and pathways of these relationships, such as why households are securing loans in more flood affected districts. A deeper understanding of these relationships will guide long term climate change resilience planning.

Appendices

1. Usability of existing global and national data for flood related vulnerability assessment in Indonesia (*manuscript published in Science of the Total Environment*)

Usability of existing global and national data for flood related vulnerability assessment in Indonesia

Breanne K. Langlois^a, Elizabeth Marsh^a, Tyler Stotland^a, Ryan Simpson^a, Katherine Berry^a, David A. Carroll^a, Aris Ismanto^b, Magaly Koch^{a,c}, and Elena N. Naumova^a

^aFriedman School of Nutrition Science and Policy, Tufts University, 150 Harrison Ave., Boston, M.A., 02111, U.S.A.

^bDepartment of Oceanography, Faculty of Fisheries and Marine Science, Universitas Diponegoro, Semarang, Indonesia

^cCenter for Remote Sensing, Boston University, 675 Commonwealth Ave # 433, Boston, M.A., 02215, U.S.A.

Corresponding author: Breanne K Langlois, 150 Harrison Ave., Boston, MA, 02111, email: breanne.langlois@tufts.edu

Abstract: Public climatic data are rapidly growing in volume and complexity at a global and national scales but these data remain underutilized for vulnerability assessment. We aim to explore how flood records from Dartmouth Flood Observatory, a global flood monitoring database, can be linked with a national disaster database maintained by the Indonesian National Board for Disaster Management, to aid local vulnerability assessment in Indonesia. We focused on physical damage to structures and agricultural crops from flooding and examined spatiotemporal patterns of a vulnerability metric derived from principal component analysis. We identified the most vulnerable areas based on emerging hot spot analysis and detected sporadic hotspots (i.e. on again then off again) of flooding in Jakarta and West Java. Using our derived metric, we identified oscillating cold spots (i.e. a cold spot that was previously a hot spot) of vulnerability in Banten, Jakarta, West Java, and Central Java. The detection of nonhomogeneous spatiotemporal trends in flooding and vulnerability demonstrate potential usability of public climate data and help to outline directions for novel research.

Keywords: Agricultural Resilience, Flooding, Vulnerability, Food Security, Climate Change, Central Java

Introduction

According to the Intergovernmental Panel on Climate Change (IPCC), Indonesia is highly vulnerable to climate change (Füssel and Klein, 2006; Mahendra Dev, 2012; Porter L. Xie A.J. Challinor K. Cochrane S.M. Howden M.M. Iqbal D.B. Lobell and M.I. Travasso, 2014). The IPCC stated that *vulnerability* can be defined as “the degree to which a system is susceptible to, or unable to cope with, adverse effects of climate change, including climate variability and extremes” (Füssel and Klein, 2006; McCarthy, 2001; Porter L. Xie A.J. Challinor K. Cochrane S.M. Howden M.M. Iqbal D.B. Lobell and M.I. Travasso, 2014). This common conceptualization helps to derive an operational definition suitable to address specific aspects of climate extremes in a specific location with the further vision to determine universal metrics in the context of climate change related vulnerability (*A Framework for Climate Change Vulnerability Assessments*, n.d.; Douglas, 2009). It also demands to have a clear vision on type and quality of data and information needed to derive such universal metrics and explore the usability of public climatic data collected on global and local scales.

Indonesia, the world’s largest archipelagic state, is one of the most flood prone areas of the world due to tidal inundation, sea level rise, river overflow from

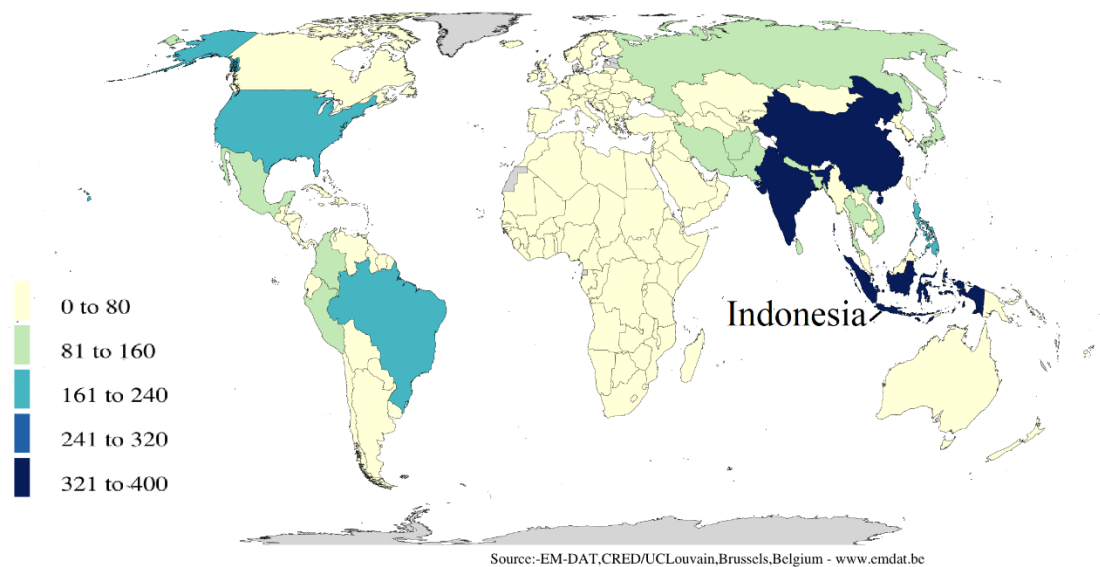
intense rainfall, and land subsidence – caused by naturally occurring young alluvial sediment, groundwater over-extraction, and urbanization (**Figure 1**) (Harwitasari and van Ast, 2011; Marfai and King, 2008, 2007). Indonesia experiences high exposure to climate variations and extremes and high sensitivity to such threats, as a large percentage of the population is employed in the agricultural sector (Füssel and Klein, 2006; Mahendra Dev, 2012; Porter L. Xie A.J. Challinor K. Cochrane S.M. Howden M.M. Iqbal D.B. Lobell and M.I. Travasso, 2014). These factors cause loss of land in coastal and agricultural areas, impacting fishponds and paddy fields (Lassa, 2012; Marfai, 2011; Marfai and King, 2008, 2007). The increase in extreme flooding events, caused by climate and geological factors, is a threat to food and nutrition security in the region, particularly among those living in rural areas who depend on agriculture and are already vulnerable to food insecurity (Douglas, 2009; Mahendra Dev, 2012). The long-term effects of flooding may be devastating for the poor in rural areas through damages to crops and decreased economic opportunities, creating a situation that may impact food consumption and nutrition (Mahendra Dev, 2012). This is expected to be a major challenge in years ahead (Mahendra Dev, 2012). Moreover, the World Bank estimates that the population exposed to extreme flooding in Indonesia will grow by over one million by 2035-2044 (*Climate Risk Profile: Indonesia*, 2021).

Assessing vulnerability of local communities in Indonesia is important to identify areas that may be prioritized for intervention or adaptation planning. But limited data availability can be a major challenge when it comes to vulnerability assessment (*A Framework for Climate Change Vulnerability Assessments*, n.d.). Vulnerability is a complex topic that can be examined through many different lenses. There are many approaches to climate vulnerability assessment in the literature that vary across disciplines (Moret, 2014; Nasiri et al., 2016; Oloyede et al., 2021). These include top-down vs bottom-up approaches, with participatory and quantitative methods (Moret, 2014; Oloyede et al., 2021). Approaches relying on primary data collection can be challenging and costly to conduct. Primary data collection may also introduce reporting biases. With such an abundance of freely accessible, routine data in the public domain, there is opportunity for interdisciplinary research related to the environment and populations (Lindersson et al., 2020). Over the past decade, the capability of identifying and tracking floods has improved and there are now many flood monitoring data portals in existence. Some studies have addressed data usability aspects for vulnerability assessment, but not of flood monitoring data which are underutilized for this purpose (Fakhruddin et al., 2022; Gall, 2015; Lindersson et al., 2020). Further, data linkages allow us to leverage strengths from multiple

data sources to address research questions that are not possible when these data are used in isolation. However, the methodological process of linking data can be challenging but is often overlooked and not well-discussed within the literature.

In this paper we aimed to demonstrate the utility of a well-established global flood monitoring database of high accessibility, alignment, velocity, and variety properties linked to a national disaster database for local flood-related vulnerability assessment in Indonesia. To accomplish this task, we: 1) devised a method to merge global flood monitoring and local disaster data; 2) created a metric of vulnerability from the merged datasets; 3) identified clusters and emerging hotspot locations of flood events and of the most vulnerable areas using the derived vulnerability metric. We demonstrated opportunities and challenges of how the constructs relating to vulnerability can be operationalized using existing, publicly accessible data to evaluate the impacts and consequences of flooding by linking global and national data sources.

Figure 1. Global occurrences from hydrological disasters, 1900 to 2022



Data sources

From May to July 2020, we searched for publicly available flood monitoring and disaster data that could be used to assess flooding and its impacts related to agricultural production in Indonesia. We assessed the applicability and usability of several data sources. We identified the Dartmouth Flood Observatory (DFO) and the Indonesian National Board for Disaster Management (BNPB) for the current analysis (“DIBI - Charts / Query Module,” n.d., “The Flood Observatory,” n.d.). We selected DFO and BNPB because they provided historical data, and they could be linked via geographic location and time. Further validation of these data sources for

this purpose could be done through mapping using other data sources but is beyond the scope of this paper.

Both data sources provide free and accessible data for download and offer different types of information pertaining to flood events. The DFO provides a global archive of large flood events dating back to 1985 using a compilation of media reports, remote sensing imagery from NASA, the Japanese Space Agency, and the European Space Agency. These data are intended for use by researchers and humanitarian aid workers. The archive provides the following information about large flood events: country, dates, glide number (if applicable), locations, area affected, duration, number of deaths and displaced, main cause, severity. The DFO defined severity in three categories: 1. large floods that caused structural or agricultural damage (occurring once every ~10 to 20 years), 2. very large floods affecting a geographic area greater than 5,000 square km (occurring once every ~20 to 100 years), and 3. extreme floods (with more than a 100-year recurrence interval). A full description of the variables is available here:

<https://floodobservatory.colorado.edu/Archives/ArchiveNotes.html>.

The BNPB provides more localized information about flood events in Indonesia dating back to 1822. This data source has a query tool where users can select the disaster type, province, and other information of interest pertaining to the events.

Available information about flood events includes date, number of deaths, evacuated, indirectly affected, people injured, total damages to crops, roads, houses, and number of houses destroyed and inundated. Exact locations of flood events are not provided – the data are aggregated at the district level.

Methods

3.1 Vulnerability framework

We applied a framework for vulnerability of agricultural systems to flooding developed by Hans-Martin Füssel to DFO and BNPB records (Füssel and Klein, 2006; Füssel, 2007). Using this framework, we designed a *current, internal, and biophysical* vulnerability metric for a local assessment and operationalize the constructs of *exposure* and *sensitivity* defined by the IPCC (**Figure 2**) (Füssel and Klein, 2006; Füssel, 2007). The DFO and BNPB data do not provide sufficient information about adaptive capacity. Thus, for the purposes of this paper, we measure vulnerability through the impacts and damages caused by the flooding events.

3.2 Data preparation and merging

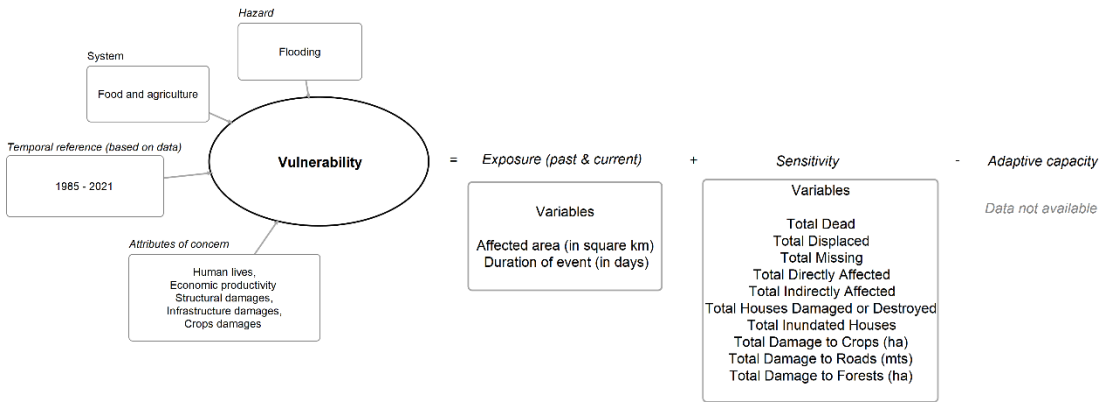
We summarized and analyzed all data using R statistical package and ArcMap, with some data management and manipulations done using Stata (ESRI, 2022; R

Core Team, 2021; StataCorp, 2017). We obtained data from the DFO and the BNPB in January of 2021. We downloaded DFO data as a shape file that included polygons pertaining to the total area affected (in km²) for each flood event. The data were then sub-set to keep only the flood events pertaining to Indonesia. We did a spatial join with Indonesian administrative boundary data to obtain the sub-districts that each flood event spanned, then exported the resulting attribute table to be imported by a statistical software program. For the BNPB data, we selected flood disasters dating back to 1985 using the query tool, since that is the earliest flood event recorded in the DFO.

We merged DFO and BNPB data by date and district (i.e. if the date of a BNPB entry fell within the range of a start and end date of a flood event in the DFO and if the district in the BNPB entry matched one of the districts affected by the flood event in the DFO). Because BNPB data corresponded to a date and district and the DFO recorded large events spanning multiple districts over multiple days, there were several cases where multiple BNPB entries matched with a single DFO event. We assumed that if the dates and districts matched, the entries from the BNPB database did refer to the same events in the DFO database. Ninety-two percent (9,630) of flood events from the BNPB did not merge with the DFO data, likely because they

were smaller events not recorded by DFO. We selected variables related to flooding exposure and the damages (i.e. sensitivity) for further analysis (**Table 1**).

Figure 2. Conceptual framework for vulnerability of agricultural system to flooding^{1,2}



¹Adapted from "Füssel HM. Vulnerability: A generally applicable conceptual framework for climate change research. Glob Environ Chang. 2007 May 1;17(2):155–67"
²Temporal reference of 1985-2021 selected based on availability of data.

3.3 Data analysis

Our analysis consisted of Principal Component Analysis to derive a vulnerability metric describing the flood-related damages and Emerging Hot Spot Analysis to examine the spatiotemporal patterns of floods and related vulnerability.

3.3.1 Constructing a vulnerability metric using Principal Component Analysis

Prior to the analysis, we examined 12 selected variables for skewness and either log transformed or treated them as binary (presence versus absence)

indicators if there were a high percentage with a value of zero (**Table 1**). We then conducted two principal component analyses (PCA). The first PCA included all variables related to both flooding exposure and the related damages displayed in **Figure 2** and **Table 1**. Then, to improve the metric, we conducted a second PCA including only the top 5 variables that contributed most highly to the first principal component (marked in **Table 1**). The first principal component from the PCA was our final metric used for mapping.

3.3.2 Emerging Hot Spot Analysis

To examine the spatiotemporal trends and dependencies of flooding events and vulnerability, we conducted Emerging Hot Spot Analysis in ArcGIS-Pro (ESRI, n.d.). We aimed to identify specific patterns classified by ArcGIS (ESRI, n.d.). We detected two common patterns: sporadic hot spot and oscillating cold spot. ArcGIS defines sporadic hot spots as “a location that is an on-again then off-again hot spot. Less than ninety percent of the time-step intervals have been statistically significant hot spots and none of the time-step intervals have been statistically significant cold spots” (ESRI, n.d.). Oscillating cold spots are defined as “a statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-

step intervals have been statistically significant cold spots” (ESRI, n.d.). A full list of the different patterns with detailed descriptions is available elsewhere (ESRI, n.d.).

The analysis was done in a few steps (ESRI, n.d.). First, we converted the DFO polygon data (representing the total area affected for each flood) into a point layer by identifying the polygon centroids. Then we did Moran’s I cluster analysis of the point layer to identify spatial clusters of flooding events. Then, we conducted the emerging hot spot analysis using the point layer of flood events. An emerging hot spot analysis relies on a space-time cube created by the user, thus, we created the space-time cube using the “Create Space Time Cube by Aggregating Points” tool in ArcGIS Pro, which used a default distance interval of 66,965 meters and a default time step interval of 2 years. We applied this space-time cube setting to create a fishnet grid over the spatial area (in this case all of Indonesia) with a neighborhood distance of 100km for 20 spatial neighborhood with the neighborhood time step set as 4 (which is multiplied by the 2 years of the space time cube, so analysis is done for every 8 years).

To examine spatiotemporal patterns in damages caused by flooding, we conducted another emerging hot spot analysis using the vulnerability metric we created from PCA. The space-time cube parameters were the same as the first analysis and hot spot parameters had 10 spatial neighbors and 3 neighborhood

timesteps (which is multiplied by the 2 years of the space time cube, so analysis is done for every 6 years).

To visualize areas most vulnerable to flooding over the past ~20 years, we created two maps, one showing exposure to flooding and one showing the flood related damage based on the emerging hot spot analyses.

Results

4.1 Spatiotemporal distribution of floods

The DFO recorded 216 flood events that occurred in Indonesia between 1985 and January 2021. Of the 216 events recorded in the DFO, 114 (53%) linked with data from the BNPB. Overall, most of the floods were large (82%), while (10%) were very large events, and (8%) were extreme. Most of the flood events were caused by heavy, torrential, or monsoonal rain (97%). On average, the floods lasted 4 days with the shortest duration being 1 day and the longest duration being 52 days. **Table 1** shows descriptive characteristics of the floods throughout the period.

Table 1. Description of large flood events and related damages recorded by the Dartmouth Flood Observatory, 1985-2020 (n=216)¹

	Data source	% With count of 0	Mean±sd	Min	Median	Max	Skewness	Principal Component 1 loadings
<i>Flood exposure and severity</i>								
Area affected (sq km) ²	DFO	N/A	29029±41518	135	13786	296259	2.97	
Natural log transformation			9±2	5	10	13	-0.40	0.06
Duration of event (days)	DFO	N/A	7±8	1	4	52	2.88	
Natural log transformation			1±0.9	0	1	4	0.35	0.19
<i>Flood-related damages on lives, community, and infrastructure</i>								
Total Dead	DFO	36%	22±70	0	3	800	7.68	0.09
Total Displaced	DFO	20%	20068±61054	0	3000	500000	5.50	0.09
Total Missing	BNPB	83%	2±11	0	0	140	10.42	0.30
Total Directly Affected ³	BNPB	64%	9383±56806	0	0	758809	11.35	0.37
Total Indirectly Affected ³	BNPB	67%	9538±47885	0	0	595955	9.40	0.36
Total Houses Damaged or Destroyed ³	BNPB	65%	794±4090	0	0	41110	7.47	0.38
Total Inundated Houses ³	BNPB	70%	4981±26065	0	0	315326	9.32	0.34
Total Damage to Crops (ha) ³	BNPB	73%	3242±19537	0	0	250000	10.13	0.35
Total Damage to Roads (mts)	BNPB	84%	62±823	0	0	12093	14.54	0.32
Total Damage to Forests (ha)	BNPB	87%	765±8937	0	0	138	12.31	0.31

¹ Information about flood-related damages obtained by linking Dartmouth Flood Observatory (DFO) data with the Indonesia National Board for Disaster Management (BNPB) data.

² Refers to the extent of geographic regions affected by flooding, and not the actual flooded areas.

³ Included in the final vulnerability metric used for mapping.

Emerging hot spot analysis showed West and Central Java had statistically significant sporadic hot spots of flooding events (**Figure 3**). Most of the time periods in this analysis (every 8 years) have been statistically significant hot spots, some have been not significant, and none have been cold spots.

4.2 Vulnerability

In the first PCA including all variables in **Figure 2** and **Table 1**, principal component 1 explained only 35% of the total variation in the data, followed by 11% in principal component 2. Indicators of having directly or indirectly affected people, caused damage to houses, inundated houses, or caused damage to crops were the strongest correlated to the first principal component which were included in another PCA to derive the vulnerability metric (**Table 1**). In this analysis that included only these 5 variables, the first component explained 62% of the total variation in the data, followed by 12% in the second component (**Figure 4**). Principal Component 1 from this analysis served as the vulnerability metric. By ranking the first component, a total of 48% of flood events were characterized as having low damages, while 25% had severe damages.

Emerging hot spot analysis identified oscillating cold spots for the derived vulnerability metric, principal component 1, in Banten, Jakarta, West Java, and

Central Java (**Figure 5**). This indicates that while this area was trending towards more damaging floods, in the last 6 years of the study period, flood damage was less bad compared to previously observed (ESRI, n.d.).

Figure 3. Sporadic hot spots of large flood events detected in emerging hot spot analysis

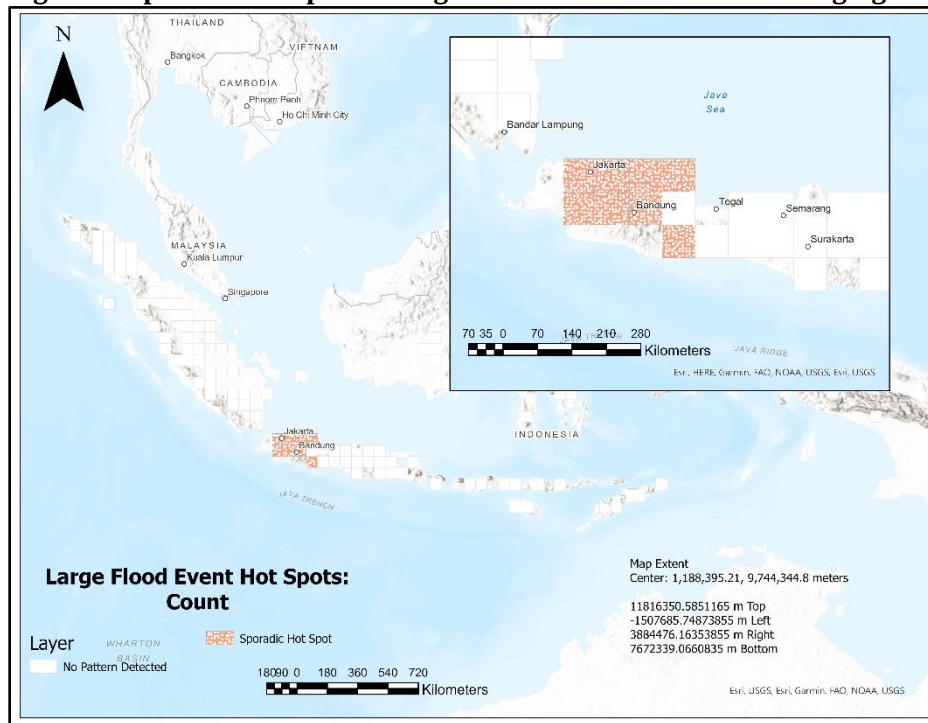
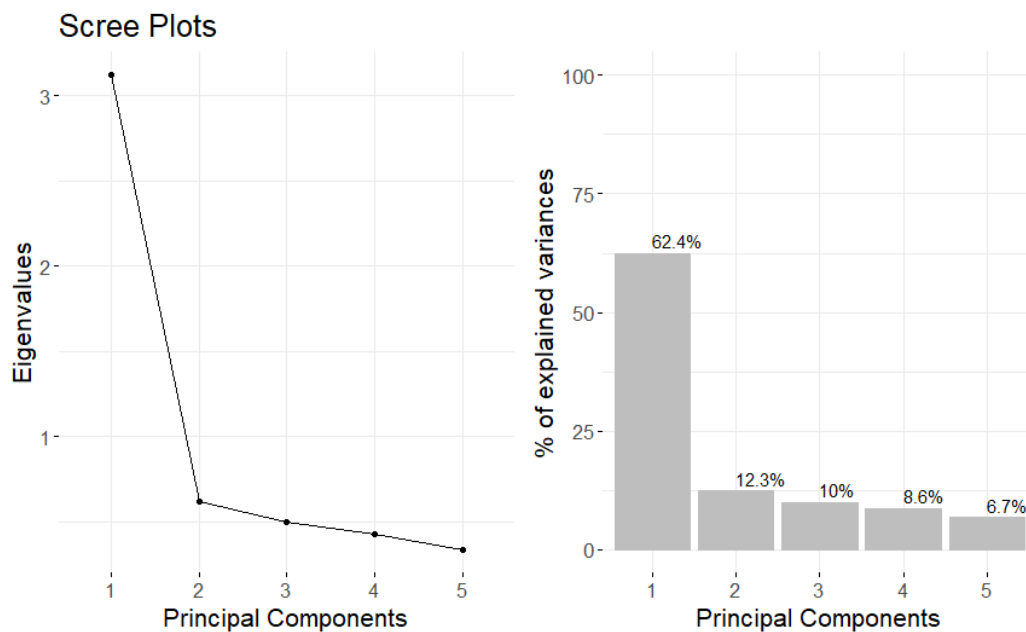
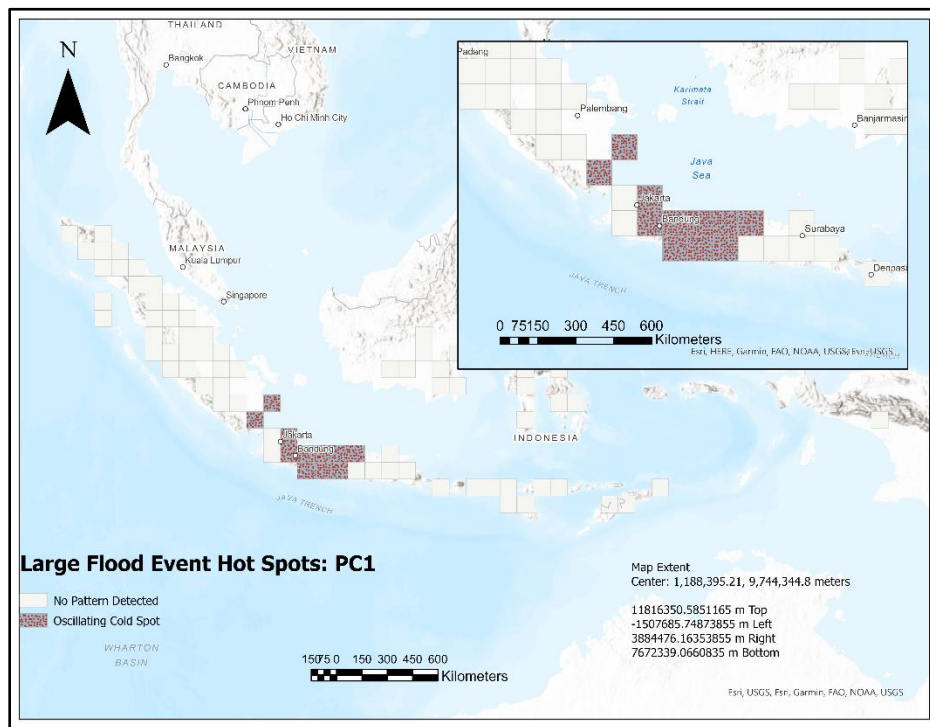


Figure 4. Principal component analysis including indicator variables for having affected people and causing damages to structures and crops¹



²Vulnerability metric derived from principal component 1 that includes indicators of having directly or indirectly affected people, caused damage to houses, inundated houses, or caused damage to crops

Figure 5. Oscillating cold spots of flood-related damages detected in emerging hot spot analysis¹



³Vulnerability metric derived from principal component 1 that includes indicators of having directly or indirectly affected people, caused damage to houses, inundated houses, or caused damage to crops

Discussion

In this exploratory study, we demonstrated the usability of public global and national data for flood related vulnerability assessment in Indonesia. We showed how existing data from separate sources – one global and one national -- can be linked to better understand the impacts of large flood events on human lives, communities, and infrastructure. By linking these two databases, we were able to obtain detailed information relating to flooding exposure and sensitivity needed for vulnerability assessment. We derived a spatial vulnerability metric that described the damage caused by flooding. This analysis would not have been possible without linking the two datasets. Our findings indicate that the

provinces of Banten, Jakarta, West Java, and Central Java are most vulnerable to flooding and flood related damages.

The study of vulnerability to climate impacts has been of interest to researchers for several decades, though there is no unified conceptualization or framework that is applicable to all contexts (Dow, 1992; Füssel, 2007; Mustafa, 1998). There is no standardized metric when it comes to climate vulnerability assessment and data availability is often the most important factor when choosing a method (*A Framework for Climate Change Vulnerability Assessments*, n.d.; Dow, 1992; Füssel, 2007; Hinkel, 2011; Mustafa, 1998). Heterogeneity of indicators of vulnerability has been discussed in earlier research and presents a challenge when it comes to policy recommendations (Eriksen and Kelly, 2007).

Prior research has shown that lack of adaptation mechanisms, typically found in smaller cities and rural areas, leads to greater vulnerability compared to larger urban areas (Handayani et al., 2017). The impact of flooding disasters likely extends beyond the immediate physical damages, but this is not well-understood yet due to lack of data to evaluate these impacts.

Our metric did not include a measure of adaptive capacity, which highlights an important data gap. There is currently a lack of existing data that is suitable for a comprehensive, low-cost vulnerability assessment, including information

on both hazard exposure and the human impacts. This was confirmed in a recent review of freely available, geospatial datasets (Lindersson et al., 2020). Authors found that, in contrast to data on hazards, availability of data on vulnerability aspects are not well developed (Lindersson et al., 2020). In recent years, there have been many improvements to the quality of hazard and exposure data (Birkmann et al., 2022). There are many hazard exposure assessments using spatial GIS-based methods that may aid disaster mitigation efforts (Bathrellos et al., 2017; Kazakis et al., 2015; Mafi-Gholami et al., 2020; Tellman et al., 2021). While these are important improvements, much work is needed to improve upon vulnerability assessment (Birkmann et al., 2022).

Studies examining vulnerability often use multiple data sources, including socioeconomic data, to construct integrated indices (Jurgilevich et al., 2017; Weis et al., 2016). However, obtaining socioeconomic data is challenging and usually relies on participatory methods (Jurgilevich et al., 2017). Poor availability of socioeconomic data has made projections of future vulnerability difficult (Jurgilevich et al., 2017). Recently, however, the Socioeconomic Data and Applications Center (SEDAC) of NASA's Earth Observing System Data and Information System (EOSDIS) announced the release of the first ever subnational global poverty grid ("News | New SEDAC Data Releases Include First-Ever

Global Poverty Grid," n.d.). This provides great opportunity for vulnerability research.

Deriving the data necessary to conduct climate vulnerability assessments can be challenging, as incorporating the different components of vulnerability such as adaptive capacity may require multiple data sources aggregated at different spatial scales. Availability of a comprehensive data source for this purpose that includes in-depth information about flooding exposures, food and agricultural related impacts, and adaptive capacity at a granular or local spatial scale is lacking. Many vulnerability assessments therefore rely on data at the national or district level which does not capture the true variability across the landscape and leads to loss of information (Eriksen and Kelly, 2007; Handayani et al., 2017). Other approaches use primary data collection and assess vulnerability through self-report (Abid et al., 2016; Moret, 2014). The data used in this paper included indicators of flooding exposure and sensitivity. Additional data relating to adaptive capacity may be obtained from survey data sources. A future direction is to determine how these linkages can be made to study vulnerability. Additional sensitivity analyses of data used in this study could be done to determine the specific time periods of emerging hotspots. Shorter time steps

could yield additional information and tease apart the patterns we observed with the flood events.

Further ground truthing is needed to determine the accuracy of these data and for validation of the metric. Validation of vulnerability indices is an important topic for improving their usability for research, but has been given little attention (Birkmann et al., 2022). Other data sources may aid in ground truthing, including MODIS NRT Global Flood Product, NASA Worldview, and the Global Surface Water Explorer.

The main limitation of this study was the coarseness of the data. While DFO contains details about flood affected areas with some information about damages, it is based mainly on news reports with hand drawn polygons. The available data are useful for identifying which general areas are more affected by large flood events, yet insufficient for determining which specific communities are affected. We operationalized a framework of vulnerability in a local context using existing data, but more spatially disaggregated data at the village level are needed to determine which communities are most affected. In addition, capturing the biological in addition to the physical impacts of flooding could be useful for assessing the effectiveness of adaptation efforts, including public health outcomes.

Principal Component Analysis, which we used to derive our vulnerability metric, is a common approach that other similar studies have used but it relies on linear relationships (Sam et al., 2021; Uddin et al., 2019). It's possible there is nonlinearity in the relationships between the variables, but this was not evaluated with Principal Component Analysis. Composite indices may oversimplify and further research is needed for how to measure vulnerability (Wiréhn et al., 2015).

While we examined 216 flood events over the study period, these floods are likely to depict primarily large-scale events and underestimate the effects. We conducted Moran's I cluster analysis (assessing spatial dependence) of the point layer of flooding events. While other alternatives exist, such as SatScan by Kulldorf et al which detects spatial outliers, Moran's I was more appropriate for our task of examining whether spatial dependencies change over time. The fishnet used for analysis had some noncontiguous areas, which may have impacted the analysis towards the null. While the two emerging hot spot analyses agree on location of increasing frequency and oscillating but significant levels of damage, it is surprising that they are not both hot spots. A sensitivity analysis of the emerging hot spot analysis could be further developed; however,

it is more plausible that due to the principal component only having a small range of values (-3 to 3) less significant changes were identified.

Our exploration of usability of two public datasets encourages novel approaches and opens new directions on the interoperability and complementarity of routinely collected data and calls for broader collaborations.

Conclusions

This exploratory study indicated complex patterns of flood occurrences in Indonesia, with Banten, Jakarta, West Java, and Central Java being vulnerable to flood related damages. Future studies of vulnerability and adaptive capacity should facilitate data linkages, assess the impact of compound events and the nature of sporadic and oscillating patterns at the granular level to better understand how different communities are impacted.

Acknowledgements

This research was supported by Tufts Institute of the Environment, Environmental Research Fellowship; the National Science Foundation US-Indonesian Research Experience for Students on Sustainable Adaptation of Coastal Areas to Environmental Change (grant number 1826939); and the United States Department of Agriculture (USDA) National Institute of Food and

Agriculture (NIFA) Cooperative State Research, Education, and Extension

Service Fellowship (grant number 2020-38420-30724). The authors would like to thank colleagues at the center for Coastal Disaster Mitigation and Rehabilitation Studies, Universitas Diponegoro in Semarang City, Central Java, Indonesia, for their continued guidance and critical feedback on presentations and manuscript drafts.

References

A Framework for Climate Change Vulnerability Assessments, n.d.

Abid, M., Schilling, J., Scheffran, J., Zulfiqar, F., 2016. Climate change vulnerability, adaptation and risk perceptions at farm level in Punjab, Pakistan. *Sci. Total Environ.* 547, 447–460.
<https://doi.org/10.1016/j.scitotenv.2015.11.125>

Bathrellos, G.D., Skilodimou, H.D., Chousianitis, K., Youssef, A.M., Pradhan, B., 2017. Suitability estimation for urban development using multi-hazard assessment map. *Sci. Total Environ.* 575, 119–134.
<https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.10.025>

Birkmann, J., Jamshed, A., McMillan, J.M., Feldmeyer, D., Totin, E., Solecki, W., Ibrahim, Z.Z., Roberts, D., Kerr, R.B., Poertner, H.-O., Pelling, M., Djalante, R., Garschagen, M., Leal Filho, W., Guha-Sapir, D., Alegría, A., 2022. Understanding human vulnerability to climate change: A global perspective on index validation for adaptation planning. *Sci. Total Environ.* 803, 150065.
<https://doi.org/10.1016/j.scitotenv.2021.150065>

Climate Risk Profile: Indonesia, 2021.

DIBI - Charts / Query Module [WWW Document], n.d. URL
<https://dibi.bnpb.go.id/DesInventar/main.jsp?countrycode=id&lang=EN>
 (accessed 1.28.21).

Douglas, I., 2009. Climate change, flooding and food security in south Asia. *Food Secur.* 1, 127–136.

Dow, K., 1992. Exploring differences in our common future(s): the meaning of vulnerability to global environmental change. *Geoforum* 23, 417–436.
[https://doi.org/10.1016/0016-7185\(92\)90052-6](https://doi.org/10.1016/0016-7185(92)90052-6)

Eriksen, S.H., Kelly, P.M., 2007. Developing credible vulnerability indicators for climate adaptation policy assessment. *Mitig. Adapt. Strateg. Glob. Chang.*
<https://doi.org/10.1007/s11027-006-3460-6>

ESRI, 2022. ArcGIS Pro Release 10.3.

ESRI, n.d. ArcGIS Pro | Documentation: How Emerging Hot Spot Analysis

works.

- Fakhruddin, B., Kirsch-Wood, J., Niyogi, D., Guoqing, L., Murray, V., Frolova, N., 2022. Harnessing risk-informed data for disaster and climate resilience. *Prog. Disaster Sci.* 16, 100254. <https://doi.org/https://doi.org/10.1016/j.pdisas.2022.100254>
- Füssel, H.-M., Klein, R.J.T., 2006. Climate change vulnerability assessments: an evolution of conceptual thinking. *Clim. Change* 75, 301–329.
- Füssel, H.M., 2007. Vulnerability: A generally applicable conceptual framework for climate change research. *Glob. Environ. Chang.* 17, 155–167. <https://doi.org/10.1016/j.gloenvcha.2006.05.002>
- Gall, M., 2015. The suitability of disaster loss databases to measure loss and damage from climate change. *Int J Glob. Warm.* 8, 170–190.
- Handayani, W., Rudiarto, I., Setyono, J.S., Chigbu, U.E., Sukmawati, A.M. awanah, 2017. Vulnerability assessment: A comparison of three different city sizes in the coastal area of Central Java, Indonesia. *Adv. Clim. Chang. Res.* 8, 286–296. <https://doi.org/10.1016/j.accre.2017.11.002>
- Harwitasari, D., van Ast, J.A.A., 2011. Climate change adaptation in practice: people's responses to tidal flooding in Semarang, Indonesia. *J. Flood Risk Manag.* 4, 216–233. <https://doi.org/10.1111/j.1753-318X.2011.01104.x>
- Hinkel, J., 2011. “Indicators of vulnerability and adaptive capacity”: Towards a clarification of the science-policy interface. *Glob. Environ. Chang.* 21, 198–208. <https://doi.org/10.1016/j.gloenvcha.2010.08.002>
- Jurgilevich, A., Räsänen, A., Groundstroem, F., Juhola, S., 2017. A systematic review of dynamics in climate risk and vulnerability assessments. *Environ. Res. Lett.* 12, 13002. <https://doi.org/10.1088/1748-9326/aa5508>
- Kazakis, N., Kougiass, I., Patsialis, T., 2015. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope–Evros region, Greece. *Sci. Total Environ.* 538, 555–563. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2015.08.055>
- Lassa, J.A., 2012. Emerging “Agricultural Involution” in Indonesia: Impact of Natural Hazards and Climate Extremes on Agricultural Crops and Food

System.

- Lindersson, S., Brandimarte, L., Mård, J., Di Baldassarre, G., 2020. A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies. *WIREs Water* 7, e1424. <https://doi.org/https://doi.org/10.1002/wat2.1424>
- Mafi-Gholami, D., Jaafari, A., Zenner, E.K., Nouri Kamari, A., Tien Bui, D., 2020. Spatial modeling of exposure of mangrove ecosystems to multiple environmental hazards. *Sci. Total Environ.* 740, 140167. <https://doi.org/10.1016/j.scitotenv.2020.140167>
- Mahendra Dev, S., 2012. Climate change, rural livelihoods and agriculture (focus on food security) in Asia-Pacific region.
- Marfai, M., 2011. Impact of coastal inundation on ecology and agricultural land use case study in central Java, Indonesia. *Quaest. Geogr.* 30, 19–32.
- Marfai, M.A., King, L., 2008. Tidal inundation mapping under enhanced land subsidence in Semarang, Central Java Indonesia. *Nat. Hazards* 44, 93–109.
- Marfai, M.A., King, L., 2007. Potential vulnerability implications of coastal inundation due to sea level rise for the coastal zone of Semarang city, Indonesia. *Environ. Geol.* 54, 1235–1245. <https://doi.org/10.1007/s00254-007-0906-4>
- McCarthy, 2001. Climate change 2001 : impacts, adaptation, and vulnerability : contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change /. Cambridge University Press, Cambridge, UK ; New York :
- Moret, W., 2014. Vulnerability Assessment Methodologies: A Review of the Literature.
- Mustafa, D., 1998. Structural Causes of Vulnerability to Flood Hazard in Pakistan. *Econ. Geogr.* 74, 289. <https://doi.org/10.2307/144378>
- Nasiri, H., Mohd Yusof, M.J., Mohammad Ali, T.A., 2016. An overview to flood vulnerability assessment methods. *Sustain. Water Resour. Manag.* 2, 331–336. <https://doi.org/10.1007/s40899-016-0051-x>

News | New SEDAC Data Releases Include First-Ever Global Poverty Grid

[WWW Document], n.d. URL

<https://sedac.ciesin.columbia.edu/news/browse> (accessed 1.18.23).

Oloyede, M.O., Benson, N.U., Williams, A.B., 2021. Climate change and coastal vulnerability assessment methods: A review. *IOP Conf. Ser. Earth Environ. Sci.* 665, 12069. <https://doi.org/10.1088/1755-1315/665/1/012069>

Porter L. Xie A.J. Challinor K. Cochrane S.M. Howden M.M. Iqbal D.B. Lobell and M.I. Travasso, J.R., 2014. Food security and food production systems. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

R Core Team, 2021. R: A Language and Environment for Statistical Computing.

Sam, A.S., Abbas, A., Surendran Padmaja, S., Raghavan Sathyan, A., Vijayan, D., Kächele, H., Kumar, R., Müller, K., 2021. Flood vulnerability and food security in eastern India: A threat to the achievement of the Sustainable Development Goals. *Int. J. Disaster Risk Reduct.* 66, 102589. <https://doi.org/10.1016/j.IJDRR.2021.102589>

StataCorp, 2017. Stata Statistical Software: Release 15.

Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A., Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. *Nat.* 2021 5967870 596, 80–86. <https://doi.org/10.1038/s41586-021-03695-w>

The Flood Observatory [WWW Document], n.d. URL

<https://floodobservatory.colorado.edu/index.html> (accessed 3.7.22).

Uddin, M.N., Islam, A.K.M.S., Saha, D., Islam, G.M.T., Md. Golam Rabbani, F., Akter, R., Haque, S., Adhikary, S., Bala, S.K., 2019. Mapping of climate vulnerability of the coastal region of Bangladesh using principal component analysis. *Appl. Geogr.* 102, 47–57. <https://doi.org/http://dx.doi.org/10.1016/j.apgeog.2018.12.011>

Weis, S.W.M., Agostini, V.N., Roth, L.M., Gilmer, B., Schill, S.R., Knowles, J.E.,

Blyther, R., 2016. Assessing vulnerability: an integrated approach for mapping adaptive capacity, sensitivity, and exposure. *Clim. Change*. <https://doi.org/10.1007/s10584-016-1642-0>

Wiréhn, L., Danielsson, Å., Neset, T.S.S., 2015. Assessment of composite index methods for agricultural vulnerability to climate change. *J. Environ. Manage.* 156, 70–80. <https://doi.org/10.1016/j.jenvman.2015.03.020>

2. Usability of the Indonesian Family Life Survey (IFLS) To Examine Recurrent Flooding and Household Food Access in Central Java, Indonesia (*oral presentation at Nutrition 2022 conference*)

Abstract #: OR03-04-22



Usability of the Indonesian Family Life Survey (IFLS) to Examine Recurrent Flooding and Household Food Access in Central Java, Indonesia

Breanne K Langlois, PhD(c)

Friedman School of Nutrition Science & Policy

Tufts University



American Society for Nutrition
Excellence in Nutrition Research and Practice
www.nutrition.org

Where the Best in Science & Health Meet

June 14-16, 2022 | #NutritionLiveOnline

Co-authors

- Oyedolapo Anyanwu, Leah Beaulac, Katherine Berry, Aliya Magnuson, Aris Ismanto, Timothy Griffin, Erin Coughlan de Perez, Magaly Koch, Elena N. Naumova



June 14-16, 2022 | #NutritionLiveOnline



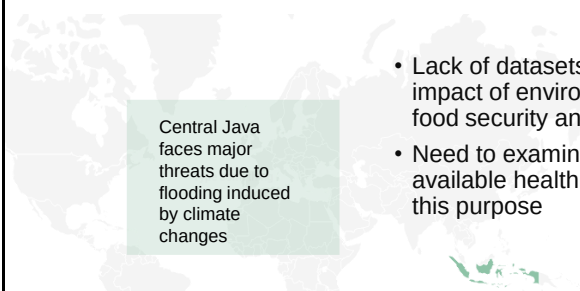
Disclosures

No conflicts to disclose

June 14-16, 2022 | #NutritionLiveOnline

NUTRITION  2022

Background



Central Java
faces major
threats due to
flooding induced
by climate
changes

- Lack of datasets designed for studying the impact of environmental exposures on household food security and health outcomes
- Need to examine usability aspects of publicly available health / nutrition survey datasets for this purpose

June 14-16, 2022 | #NutritionLiveOnline

NUTRITION  2022

IFLS design

- Nationally representative, longitudinal economic and health survey tracking Indonesian households from 1993 to 2015
- Data can be analyzed longitudinally or cross-sectionally
- Individual, household, and community level data
- Has a sub-district (*Kecamatan*) spatial aggregation

How to measure
household
exposure to floods
&
household food
access?

June 14-16, 2022 | #NutritionLiveOnline

NUTRITION **LIVE ONLINE** 2022

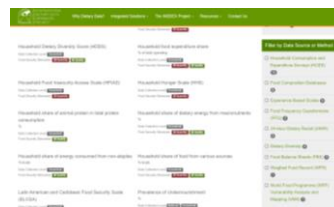
Methods

- Review of IFLS household survey modules and documentation for *availability, alignment, variety, continuity, and compatibility* of information to measure exposure to floods and food access metrics
- Crosswalk of survey modules and variables to compare across waves
- Constructed multilevel dataset by transposing, merging, and appending waves 1-5

The Data4Diets framework

	Quality	Quantity	Preferability	Safety	Stability	Sustainability
National, Market (Available)						
Market, Household, Individual (Accessible)						
Household, Individual (Utilizable)						

Source: Adapted from Data4Diets, HCEDE Project (2019) [https://doi.org/10.1016/j.data.2019.01.001]

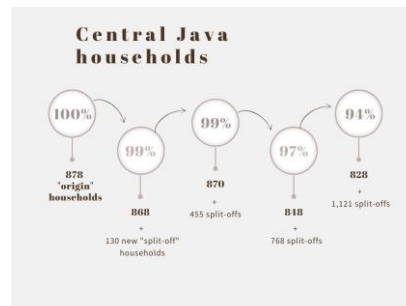


June 14-16, 2022 | #NutritionLiveOnline

NUTRITION **LIVE ONLINE** 2022

Tracking of households over 5 cross-sectional waves

- 15,921 interviewed households over 22 years, including 1,949 in Central Java
- Of the 15,921 households
 - 6,044 are “origin” households appearing in all waves
 - 9,877 are “split-off” households formed from a target member of a previous wave

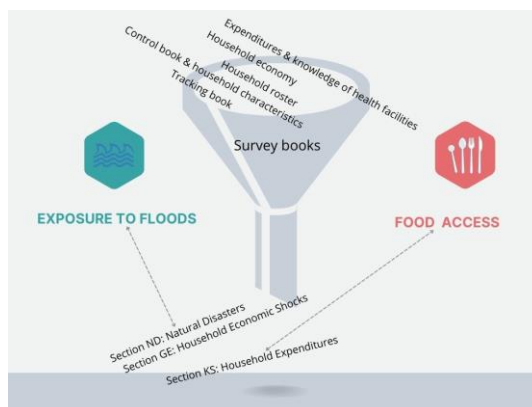


June 14-16, 2022 | #NutritionLiveOnline

NUTRITION LIVE ONLINE **2022**

Composition of household survey

✓ Availability



June 14-16, 2022 | #NutritionLiveOnline

NUTRITION LIVE ONLINE **2022**

Feasibility of provided data files

✓ Continuity

✓ Alignment

Relevant Survey Module	Continuity across waves	Number of data files	Data format	Data structure	User processing required
Section GE: Household Economic Shocks	1 → 2 → 3 → 4 → 5	1	Stata, SAS	Hardship type	Transpose
Section ND: Natural Disasters	1 → 2 → 3 → 4 → 5	2	Stata, SAS	Household & disaster type	Transpose
Section KS: Household Expenditures	1 → 2 → 3 → 4 → 5	6	Stata, SAS	Household, food expenditure item, non-food expenditure item, & food item	Transpose, construct monthly aggregates for wave5

June 14-16, 2022 | #NutritionLiveOnline

NUTRITION  2022

Measurement of key variables

✓ Variety

✓ Compatibility

Household exposure to flood over 10-year span

- ✓ Indicator of flood occurrence
- ✓ Indicator of severely impacted by flood
- ✓ Frequency of floods
- ✓ Derived flood severity defined as:
 - No flood occurrence
 - Flood occurrence with no severe impact
 - Flood occurrence with severe impacts

Food access indicators over 20-year span

- ✓ Food consumption score,
- ✓ Household average dietary energy acquisition or consumption,
- ✓ Household adequacy of fruit and vegetable consumption,
- ✓ Household food expenditure share
- ✓ Household dietary diversity score

June 14-16, 2022 | #NutritionLiveOnline

NUTRITION  2022

Spatial survey data linkage

- Spatially join flood monitoring data 1988-2015 with Central Java administrative boundaries using GIS to obtain spatial area that flood events spanned
- Create summary variable by district year
- Merge with IFLS longitudinal dataset by district and survey year

June 14-16, 2022 | #NutritionLiveOnline



Conclusion

- IFLS has an excellent tracking record and can identify flood affected households, but does not provide a good measure of flood recurrence
- Food access indicators can be constructed using consumption and expenditure data
- Spatial / survey linkages are possible and can address the data gap, but surveys should include modules of environmental exposures
- Need for a validated disaster severity index
- Data usability evaluations needed for purposes other than intended

June 14-16, 2022 | #NutritionLiveOnline



Next steps

- Examine usability of flood monitoring data for public health and nutrition research
- Thorough examination of completeness of IFLS indicators
- Describe the spatiotemporal distribution of floods across Central Java using dynamic mapping and validate with IFLS flood data
- Link flood monitoring data to IFLS

3. Template of usability characteristics for comparing flood monitoring portals

Data portal:	Link:	Date
reviewed:		

1. *Accessibility* refers to the aim and mission of the host organization collecting and publishing the data source(s) used in your dataset. Please consider:

- a. Is the data publicly accessible or derived from a private source?
- b. Who is the host organization and why did/do they collect the data?
- c. What does the host organization seek to gain by offering these data for analysis?
- d. How are the data accessed? Can they be easily downloaded? Or requires username/login or contacting the organization to obtain?

2. *Availability* refers to the type of information presented in the dataset and the comparability of the data provided. Please consider:

- a. What data products are available?
- b. Provides definition of flood?
- c. Offers what type of information about floods?
- d. Distinguishes main cause and type of flood?
- e. How are these variables defined? Can they be compared easily over time?

3. *Data source(s)*:

- a. Where/how does the portal obtain the data?
- b. What satellite sensors are the data derived from?

- c. Does the portal include other non-satellite data sources (e.g. news reports or government data?)

4. Data format(s)

- a. what format are the data provided in?
- b. Are multiple formats available?

5. Alignment *refers to how the dataset is structured and referenced with respect to time, geographic location, and other contexts. Please consider:*

- a. What time units, date, or time periods are data reported for?
- b. What is the spatial domain covered by the data?

6. Volume *refers to the quantity of observations and variables within the data available for analysis. Please consider:*

- a. What is the size of the dataset (number of observations and variables)?

7. Velocity *refers to the frequency at which data are updated, collected, reported, and made available for analysis. Please consider:*

- a. Do gaps exist between time points of collected data?
- b. How regularly does the data source(s) monitor and document changes?

8. Variety *refers to the data types in a dataset, general format of the variables, and possible manipulations needed to obtain the desirable outcomes. Please consider:*

- a. Are variables of interest presented as continuous, discrete, ordered, categorical, censored, or dichotomous forms?
- b. Are any data transformations or calculations necessary for your analysis?

9. Completeness *refers to the amount and potential patterns of missing information within your dataset. Please consider:*

- a. How are missing values reported? Are data imputed if no values exist?
- b. What is the effective sample size once accounting for missing data points?
- c. How might missing data influence your analysis?

10. Has user tutorial on website/requires tutorial?

11. Has visualizations that represent the data?

- a. *visualizations show temporal and spatial dimensions?*

4. Webform Survey – User Feasibility Aspects of Flood Monitoring & Data Portals

Flood Monitoring Data Portals - External User Assessment

CONSENT TO PARTICIPATE - PLEASE READ CAREFULLY

HELLO, I AM A PHD CANDIDATE AT THE FRIEDMAN SCHOOL. I AM CONDUCTING A RESEARCH STUDY COMPARING
USABILITY, QUALITY, AND FEASIBILITY ASPECTS OF 5 FLOOD MONITORING DATA PORTALS FROM THE PERSPECTIVE OF
PUBLIC HEALTH AND NUTRITION RESEARCHERS. AS PART OF THIS RESEARCH, I WOULD LIKE YOUR EXPERT OPINION
ABOUT EACH OF THESE DATA PORTALS USING A FORM I HAVE CREATED. IT WILL NOT TAKE MORE THAN 2 HOURS OF
YOUR TIME. DECISION TO PARTICIPATE IS COMPLETELY VOLUNTARY AND THERE ARE NO CONSEQUENCES TO SAYING
NO. THERE ARE NO ANTICIPATED RISKS TO PARTICIPATION. IF THERE ARE ANY FIELDS ON THE FORM YOU DO NOT WANT
TO ASSESS, YOU MAY LEAVE IT BLANK. IF AT ANY TIME YOU WOULD LIKE TO STOP PARTICIPATING, YOU MAY DO SO BY
CLOSING THE WEB FORM. YOUR RESPONSES WILL BE RECORDED AND MAY BE USED IN PUBLICATIONS OR
PRESENTATIONS. NO PERSONALLY IDENTIFIABLE INFORMATION WILL BE COLLECTED. YOUR EMAIL ADDRESS WILL NOT
BE LINKED TO THE SURVEY. FOR QUESTIONS OR CONCERNS ABOUT THE RESEARCH STUDY OR PROCEDURES, OR IF YOU

NEED TO NOTIFY SOMEONE OF A COMPLAINT, PLEASE CONTACT THE TUFTS RESEARCH TEAM.

BREANNE LANGLOIS

TUFTS UNIVERSITY

FRIEDMAN SCHOOL OF NUTRITION SCIENCE AND POLICY

EMAIL: BREANNE.LANGLOIS@TUFTS.EDU

PHONE NUMBER: 603-303-1404

IF YOU HAVE QUESTIONS OR CONCERNS ABOUT YOUR RIGHTS AS A RESEARCH PARTICIPANT, OR IF YOU WOULD LIKE

TO DISCUSS THE STUDY WITH SOMEONE OUTSIDE OF THE RESEARCH TEAM, CONTACT THE TUFTS SBER IRB.

TUFTS UNIVERSITY

SOCIAL BEHAVIORAL & EDUCATIONAL RESEARCH

INSTITUTIONAL REVIEW BOARD (SBER IRB)

75 KNEELAND STREET, 6TH FLOOR | BOSTON, MA 02111

TELEPHONE: 617-627-8804

EMAIL: SBER@TUFTS.EDU

WEBSITE: [HTTP://VICEPROVOST.TUFTS.EDU/SBERIRB/](http://VICEPROVOST.TUFTS.EDU/SBERIRB/)

BY SELECTING THE "I CONSENT" BUTTON BELOW, YOU ARE CONSENTING TO PARTICIPATE IN THIS RESEARCH STUDY. *

☐ I consent to participate

☐ I do not consent to participate

THANK YOU FOR PARTICIPATING! PLEASE REVIEW EACH OF THE 5 FLOOD MONITORING PORTALS LISTED IN THE DROP DOWN MENU BELOW AND COMPLETE THIS FORM SEPARATELY FOR EACH ONE. TO START, SELECT A PORTAL FROM THE DROP DOWN MENU AND COMPLETE THE CORRESPONDING FORM. WHEN FINISHED, SUBMIT THE FORM BY CLICKING THE SUBMIT BUTTON AT THE BOTTOM OF THE FORM. CLICK ON THE LINK TO THE WEBFORM PROVIDED IN THE EMAIL TO REVIEW ANOTHER PORTAL AND SUBMIT ANOTHER FORM. YOU MAY COMPLETE ALL 5 REVIEWS IN 1 SESSION OR AT DIFFERENT TIMES.

NAME OF FLOOD MONITORING DATA PORTAL

[Please select which flood monitoring portal you are reviewing and complete the corresponding survey. Please submit 1 form for each flood monitoring portal. Do not spend more than ~20 minutes for each one.]

- ☐ Dartmouth Flood Observatory
 ☐ Global Flood Monitoring System
- ☐ JRC's Global Flood Awareness System (GloFAS)
- ☐ Global Surface Water Explorer
 ☐ NASA Worldview

CLICK BELOW TO ACCESS DARTMOUTH FLOOD OBSERVATORY

<https://floodobservatory.colorado.edu>

CLICK BELOW TO ACCESS GLOBAL FLOOD MONITORING SYSTEM

<http://flood.umd.edu>

CLICK BELOW TO ACCESS JRC'S GLOBAL FLOOD AWARENESS SYSTEM (GLOFAS)

<https://www.globalfloods.eu>

COPY AND PASTE THIS LINK INTO A BROWSER TO ACCESS GLOBAL SURFACE WATER EXPLORER: [HTTPS://GLOBAL- SURFACE-WATER.APPSPOT.COM/](https://GLOBAL-SURFACE-WATER.APPSPOT.COM/)

CLICK BELOW TO ACCESS NASA WORLDVIEW

<https://worldview.earthdata.nasa.gov>

GOALS & SCOPE

OVERALL GOALS ARE CLEARLY STATED.

Strongly disagree Disagree Agree Strongly agree

THE AVAILABLE DATA PRODUCTS ARE CLEARLY LISTED WITH A SUFFICIENT DESCRIPTION.

Strongly disagree Disagree Agree Strongly agree

OVERALL, IT WAS QUICK AND EASY TO UNDERSTAND WHAT DATA PRODUCTS ARE OFFERED.

Strongly disagree Disagree Agree Strongly agree

DATA QUALITY

THE DATA SOURCES AND METADATA ARE CLEARLY DESCRIBED.			
Strongly disagree	Disagree	Agree	Strongly agree
THE LEVEL OF AVAILABLE GEOSPATIAL AGGREGATIONS IS CLEAR: E.G. FLOOD EXTENT, POSTAL CODE, VILLAGE/CITY/TOWN, STATE/PROVINCE/REGION, AND COUNTRY.			
Strongly disagree	Disagree	Agree	Strongly agree
THE LEVEL OF AVAILABLE TEMPORAL AGGREGATIONS IS CLEAR: E.G. DAY, WEEK, MONTH, QUARTER, YEAR, ETC.			
Strongly disagree	Disagree	Agree	Strongly agree
HISTORICAL DATA ARE PROVIDED (1980S TO PRESENT).			
Strongly disagree	Disagree	Agree	Strongly agree
THE DATA CAN BE CLASSIFIED INTO SUB-NATIONAL ADMINISTRATIVE BOUNDARIES (E.G. PROVINCE, DISTRICT, OR VILLAGE).			
Strongly disagree	Disagree	Agree	Strongly agree
THE DATA CAN BE CLASSIFIED INTO SUB-YEARLY TEMPORAL UNITS (E.G. MONTH, WEEK, OR DAY).			
Strongly disagree	Disagree	Agree	Strongly agree

A DEFINITION OF FLOOD OR CLASSIFICATION OF HOW A FLOODED AREA WAS IDENTIFIED IS CLEARLY LISTED.			
Strongly disagree	Disagree	Agree	Strongly agree
A DEFINITION OR CLASSIFICATION OF FLOOD EXPOSURE IS PROVIDED (E.G. HOW A PERSON, HOUSEHOLD, OR COMMUNITY COULD BE CLASSIFIED AS BEING EXPOSED TO FLOOD ACROSS SPACE AND TIME).			
Strongly disagree	Disagree	Agree	Strongly agree
THE AVAILABLE DATA CLASSIFIES TYPE OF FLOOD (E.G. RIVER, TIDAL, OR FLASH FLOOD).			
Strongly disagree	Disagree	Agree	Strongly agree
THE AVAILABLE DATA CLASSIFIES OR PROVIDES A MEASURE OF FLOOD RECURRENCE FOR A GIVEN SPATIO-TEMPORAL AGGREGATION.			
Strongly disagree	Disagree	Agree	Strongly agree
THE AVAILABLE DATA INCLUDES OTHER INFORMATION ABOUT FLOOD (SUCH AS DURATION, DEPTH, DAMAGES)			
Strongly disagree	Disagree	Agree	Strongly agree
IF DATA IS NOT AVAILABLE FOR A GIVEN VARIABLE, REASONS ARE CLEARLY STATED.			
Strongly disagree	Disagree	Agree	Strongly agree

VISUALIZATION QUALITY

THE PORTAL PROVIDES ADEQUATE AND EFFECTIVE VISUALIZATIONS TO DEMONSTRATE THE DATA DISTRIBUTION, TIME TREND, QUALITY, QUANTITY AND RELATIONSHIP OF THE DATA.

Strongly disagree

Disagree

Agree

Strongly agree

THE VISUALIZATIONS CLEARLY COMMUNICATE THE DATA.

Strongly disagree

Disagree

Agree

Strongly agree

PLATFORM ACCESSIBILITY

THE PORTAL IS USER-FRIENDLY AND ALLOWS USERS TO EASILY NAVIGATE AND EXPLORE DATA AND VISUALIZATIONS.

Strongly disagree

Disagree

Agree

Strongly agree

THE PORTAL INCLUDES A TUTORIAL OR USER GUIDE (E.G., INFORMATIVE POP-UPS, A TUTORIAL VIDEO).

Strongly disagree

Disagree

Agree

Strongly agree

THE DATA ARE EASILY ACCESSIBLE.

Strongly disagree

Disagree

Agree

Strongly agree

MULTIPLE DATA FORMATS ARE AVAILABLE TO FACILITATE A RANGE OF USES IN DIFFERENT PROGRAMS (STATISTICAL PACKAGES AND SPATIAL).

Strongly disagree

Disagree

Agree

Strongly agree

LINKS ARE UPATED AND PROVIDE UPDATED INFORMATION.

Strongly disagree

Disagree

Agree

Strongly agree

CONTACTS, COMMUNICATION, & CONFLITS OF INTEREST

THE PORTAL INCLUDES CONTACT INFORMATION FOR QUESTIONS, SUGGESTIONS AND FEEDBACK.

Strongly disagree

Disagree

Agree

Strongly agree

THE HOST ORGANIZATION AND PURPOSE OF PROVIDING THE DATA IS CLEARLY STATED.

Strongly disagree

Disagree

Agree

Strongly agree

FUNDING SOURCES AND ROLES OF FUNDERS ARE CLEARLY STATED.

Strongly disagree

Disagree

Agree

Strongly agree

NOTES

ADDITIONAL NOTES

*NOTE: The form included a 5th category for 'not clear / difficult to assess'. Not shown due to formatting adjustments. The webform can also be viewed at this link:

<https://ee.kobotoolbox.org/preview/vPdukFls>.

5. Data and Code Sharing on Github

The following is a list of items that will be made available on my personal Github page at <https://github.com/BreeLanglois>:

- R script for dynamic map
- Dynamic map (animated GIF)
- R script for extracting GSW raster values into administrative polygons
- Stata script for linking flood data to IFLS
- Stata script for creating the longitudinal IFLS dataset
- Stata script for creating the wealth index using PCA
- Dataset used for analysis

6. Comparison of IFLS modules

Wave	Year	Sampling Info.	Survey	Book (survey instrument file(s))	Section	Stata dataset name	Folder	Dataset structure	Unique ID	Linking Vars.	Reshape
1	1993	IFLS is a longitudinal survey - sampling scheme for wave 1 is the primary determinant of the sample in subsequent waves. IFLS1 sampling scheme stratified on provinces, then randomly sampled within provinces. Includes 13 of Indonesia's 26 provinces containing	Household	BOOK CA Anthropometric Record (IFLS1\DRU1195\DRU1195_2\DRU1195.2.Book.CA.pdf)	Section CA (Anthropometric Measurements)	bukcca1.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
						bukcca2.dta	hh93dta	individual	pidlink	case or hhid (case = hhid), pidlink (person id)	
				BOOK K Control Book	Section SC (Sampling Information and Enumeration Record)	bukks1.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
				BOOK I Household Roster and Characteristics (IFLS1\DRU1195\DRU1195_2\DRU1195.2.Book1.pdf)	Section KR (Household Characteristics)	buk1kr1.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
					Section KS (Consumption) - USE DERIVED HOUSEHOLD CONSUMPTION EXPENDITURE AGGREGATES	buk1ks1.dta	hh93dta	household food item consumed	case, item	case or hhid (case = hhid)	y
						buk1ks2a.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
						buk1ks2b.dta	hh93dta	household non-food item	case, nf_item1	case or hhid (case = hhid)	y
						buk1ks3a.dta	hh93dta	household non-food item	case, nf_item2	case or hhid (case = hhid)	y
						buk1ks3b.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
						buk1ks4.dta	hh93dta	household food/non-food item type of purchase	case, item_m	case or hhid (case = hhid)	y
						pce93nom.dta	R:\Indonesia Family Life Survey\consumption_expenditure aggregates	household	hhid93	hhid93	
				BOOK II Household Economy (IFLS1\DRU1195\DRU1195_2\DRU1195.2.Book2.pdf)	Section UT (Farm Business)	buk2ut1.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
						buk2ut2.dta	hh93dta	household farm business asset	case, farmasst	case or hhid (case = hhid)	y

	ng 83% of the population. Within each of the 13 provinces, enumeration areas (EAs) were randomly chosen. Within a selected EA, field teams randomly selected households based upon 1993 SUSEN AS listings obtained from regional BPS office. For IFLS1 a total of 7,730 households were sampled to obtain a final sample			Section NT (Non-Farm Business)	buk2n t1.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
					buk2n t2.dta	hh93dta	household non-farm business asset	case, nfirmasst	case or hhid (case = hhid)	y
				Section HR (Household Assets)	buk2h r1.dta	hh93dta	household asset type	case, assettyp	case or hhid (case = hhid)	y
					buk2h r2.dta	hh93dta	household	case or hhid	case or hhid (case = hhid)	
				Section GE (Household Economic Shocks)	buk2g e1.dta	hh93dta	household hardship type	case, typehard	case or hhid (case = hhid)	y
		Community	Book 1	Section D (Agriculture and Industry)	bukid 01.dta	cf93dta				
					bukid 02a.dta	cf93dta				
				Section E (History and Climate)	bukie 02a.dta	cf93dta				
					bukie 02b.dta	cf93dta	community event year	be10	commid93	y
			Book 2 (IFLS1\DRU1195\DRU1195_3\DRU1195.3.book2.pdf)	Statistics of village and direct observation	bkii_c f.dta	cf93dta	community		commid93	

		size goal of 7,000 completed households. In fact, interviews were conducted with 7,224 households in IFLS1. IFLS 1 interview was conducted with some but not all household members.									
2	1997	The household questionnaire in IFLS2 was organized like its IFLS1 counterpart and repeated many of the same questions to allow comparisons across	Household	Book K: Control Book and Household Roster (\IFLS2\vol1_7\volume3\hh97bk.pdf)	SC Sampling and enumeration record	bk_sc.dta	\hh97dta\hh97bk	household	hhid97	hhid97	
					KRK Housing characteristics (In IFLS1, part of book I)	bk_krk.dta	\hh97dta\hh97bk	household	hhid97	hhid97	
				Book 1: Household Expenditures and Knowledge of Health Facilities (\IFLS2\vol1_7\volume3\hh97b1.pdf)	KS Household expenditures / Consumption (Corresponds to IFLS1, book I, but was shortened in IFLS2 to reduce respondent burden. Non-food expenses are for households in IFLS2, for households or individuals in IFLS1.) - USE DERIVED HOUSEHOLD CONSUMPTION EXPENDITURE AGGREGATES	b1_ks0.dta	\hh97dta\hh97b1	household	hhid97	hhid97	
						b1_ks1.dta	\hh97dta\hh97b1	household food exp item	hhid97, ks1type	hhid97	y
						b1_ks2.dta	\hh97dta\hh97b1	household non-food exp item	hhid97, ks2type	hhid97	y
						b1_ks3.dta	\hh97dta\hh97b1	household non-food exp item	hhid97, ks3type	hhid97	y
						b1_ks4.dta	\hh97dta\hh97b1	household food item	hhid97, ks4type	hhid97	y

		waves. The IFLS1 question naire containe d many retrospe ctive question s coverin g past events. In IFLS2, full retrospe ctives were asked of new respond ents. For most sections , panel respond ents were only asked to update the informat ion, starting approxi mately five years before the 1997 interview, so there is one year of				pce97 nom.d ta	R:\Indonesia Family Life Survey\consumption_e xpenditure aggregates	household	hhid97	hhid97	
			Book 2: Household Economy (\IFLS2\vol1_7\volume3\h h97b2.pdf)	KR Housing characteristics (in IFLS1, part of book I)	b2_kr. dta	\hh97dta\hh97b2	household	hhid97	hhid97		
				UT Farm business	b2_ut 1.dta	\hh97dta\hh97b2	household	hhid97	hhid97		
					b2_ut 2.dta	\hh97dta\hh97b2	household farm business asset type	hhid97, uttype	hhid97	y	
				NT Nonfarm business	b2_nt 1.dta	\hh97dta\hh97b2	household	hhid97	hhid97		
					b2_nt 2.dta	\hh97dta\hh97b2	household non- farm business asset	hhid97, nttype	hhid97	y	
				HR Household assets	b2_hr 1.dta	\hh97dta\hh97b2	household asset type	hhid97, hrtype	hhid97	y	
					b2_hr 2.dta	\hh97dta\hh97b2	household	hhid97	hhid97		
				GE Household economic shocks	b2_ge .dta	\hh97dta\hh97b2	household hardship type	hhid97, getype	hhid97	y	
	Communi ty		Book 1 (\IFLS2\vol1_7\volume4\c f97bk1\)	D Agriculture and industry	bk1_d 1.dta	cf97dta	irrigation	commid97, d1type	commid9 7	y	
					bk1_d 2.dta	cf97dta	activity	commid97, d2type	commid9 7	y	
					bk1_d 3.dta	cf97dta	crop	commid97, d3type	commid9 7	y	
					bk1_d 4.dta	cf97dta	factory	commid97, d4type	commid9 7	y	
					bk1_d 5.dta	cf97dta	cottage industry	commid97, d5type	commid9 7	y	
				E Community history and climate	bk1_e 1.dta	cf97dta	name change	commid97, e1type	commid9 7		
			bk1_e 2.dta		cf97dta	major event	commid97, e2type	commid9 7	y		
			Book 2: Community statistical records (\IFLS2\vol1_7\volume4\c f97bk2\)	LK Basic information	bk2.dt a	cf97dta	community	commid97	commid9 7		
				S Statistics	bk2.dt a	cf97dta	community	commid97	commid9 7		
				OL Direct observation (e.g., cleanliness, prosperity, social cohesion)	bk2.dt a	cf97dta	community	commid97	commid9 7		

		overlap between IFLS1 and IFLS2 data. IFLS2 included 6,742 origin households, plus an additional 878 new split off households for a total of 7620 households.									
3	2000	6,758 origin households, 1,031 split off households from IFLS 2 (751) and 2+ (280), 2,646 new split offs from IFLS 3, total 10,435 households	Household	Book K: Control Book and Household Roster (\\IFLS3\\hh00_all_doc_02_BOOK K_HH_.pdf.pdf)	SC Sampling and enumeration record	bk_sc.dta	hh00_all_dta	household	hhid00	hhid00	
					KRK Housing characteristics (In IFLS1, part of book 1)	bk_krk.dta	hh00_all_dta	household	hhid00	hhid00	
				Book 1: Household Expenditures and Knowledge of Health Facilities (\\IFLS3\\hh00_all_doc_03_BOOK 1_HH_.pdf.pdf)	KS Household expenditures / Consumption - USE DERIVED HOUSEHOLD CONSUMPTION EXPENDITURE AGGREGATES	b1_ks0.dta	hh00_all_dta	household	hhid00	hhid00	
						b1_ks1.dta	hh00_all_dta	household food exp item	hhid00, ks1type	hhid00	y
						b1_ks2.dta	hh00_all_dta	household non-food exp item	hhid00, ks2type	hhid00	y
						b1_ks3.dta	hh00_all_dta	household non-food exp item	hhid00, ks3type	hhid00	y
						b1_ks4.dta	hh00_all_dta	household food item	hhid00, ks4type	hhid00	y
						pce00nom.dta	R:\\Indonesia Family Life Survey\\consumption_expenditure aggregates	household	hhid00	hhid00	

4	2	0	6,429 origin	Household	Book 2: Household Economy (IFLS3\hh00_all_doc\04_BOOK 2_HH_.pdf)	KR Housing characteristics (in IFLS1, part of book 1)	b2_kr.dta	hh00_all_dta	household	hhid00	hhid00	
					UT Farm business		b2_ut1.dta	hh00_all_dta	household	hhid00	hhid00	
							b2_ut2.dta	hh00_all_dta	household farm business asset type	hhid00, uttype	hhid00	y
					NT Nonfarm business		b2_nt1.dta	hh00_all_dta	household	hhid00	hhid00	
							b2_nt2.dta	hh00_all_dta	household non-farm business asset	hhid00, nt_num	hhid00	y
					HR Household assets		b2_hr1.dta	hh00_all_dta	household asset type	hhid00, hrtype	hhid00	y
							b2_hr2.dta	hh00_all_dta	household asset type	hhid00, hr2type	hhid00	y
					GE Household economic shocks		b2_ge.dta	hh00_all_dta	household hardship type	hhid00, getype	hhid00	y
				Community	Book 1 (don't see the community survey instrument in the documentation)	D Agriculture and industry	bk1_d1.dta	cf00_all_dta	irrigation	commid00, d1type	commid00	y
							bk1_d2.dta	cf00_all_dta	activity	commid00, d2type	commid00	y
							bk1_d3.dta	cf00_all_dta	crop	commid00, d3type, d19type	commid00	y
							bk1_d4.dta	cf00_all_dta	factory	J VARIABLE MISSING IN DATASET -- CANNOT RESHAPE	commid00	y
							bk1_d5.dta	cf00_all_dta	cottage industry	commid00, d5type	commid00	y
						E Community history and climate	bk1_e1.dta	cf00_all_dta	name change	commid00, e1type	commid00	y
							bk1_e2.dta	cf00_all_dta	major event	commid00, e2type	commid00	y
					Book 2: Community statistical records (don't see the community survey instrument in the documentation))	LK Basic information	bk2.dta	cf00_all_dta	community	commid00	commid00	
						S Statistics	bk2.dta	cf00_all_dta	community	commid00	commid00	
						OL Direct observation (e.g., cleanliness, prosperity, social cohesion)	bk2.dta	cf00_all_dta	community	commid00	commid00	
					Book K: Control Book and Household Roster	SC Sampling and enumeration record	bk_sc.dta	hh07_all_dta	household	hhid07	hhid07	

07	households, 7,107 split off households from IFLS 2 (704) and 2+ (261) and 3 (2,109), 4,033 new split offs from IFLS 4, total 13,536 households		(\IFLS4\hh07_all_doc\hh07_english_pqx\HH BOOK K FINAL.pdf)	KRK Housing characteristics (In IFLS1, part of book I)	bk_kr.k.dta	hh07_all_dta	household	hhid07	hhid07	
			Book 1: Household Expenditures and Knowledge of Health Facilities (\IFLS4\hh07_all_doc\hh07_english_pqx\HH BOOK 1 FINAL.pdf)	KS Household expenditures / Consumption - USE DERIVED HOUSEHOLD CONSUMPTION EXPENDITURE AGGREGATES	b1_ks0.dta	hh07_all_dta	household	hhid07	hhid07	
					b1_ks1.dta	hh07_all_dta	household food exp item	hhid07, ks1type	hhid07	y
					b1_ks2.dta	hh07_all_dta	household non-food exp item	hhid07, ks2type	hhid07	y
					b1_ks3.dta	hh07_all_dta	household non-food exp item	hhid07, ks3type	hhid07	y
					b1_ks4.dta	hh07_all_dta	household food item	hhid07, ks4type	hhid07	y
					pce07nom.dta	R:\Indonesia Family Life Survey\consumption_expenditure aggregates	household	hhid07	hhid07	
			Book 2: Household Economy (\IFLS4\hh07_all_doc\hh07_english_pqx\HH BOOK 2 FINAL.pdf)	KR Housing characteristics (in IFLS1, part of book I)	b2_kr.dta	hh07_all_dta	household	hhid07	hhid07	
				UT Farm business (differs from the earlier waves)	b2_ut1.dta	hh07_all_dta	household	hhid07	hhid07	
					b2_ut2.dta	hh07_all_dta	household most recent rice crop type	hhid07, ut2type	hhid07	y
					b2_ut3.dta	hh07_all_dta	type of hardship	hhid07, ut3type	hhid07	y
					b2_ut4.dta	hh07_all_dta	household farm business asset type	hhid07, uttype	hhid07	y
				NT Nonfarm business	b2_nt1.dta	hh07_all_dta	household	hhid	hhid07	
					b2_nt2.dta	hh07_all_dta	household non-farm business asset	hhid, nt_num	hhid07	y
				HR Household assets	b2_hr1.dta	hh07_all_dta	household asset type	hhid, hrtype	hhid07	y
					b2_hr2.dta	hh07_all_dta	household asset type	hhid, hr2type	hhid07	y
				ND Natural Disasters (not included in waves prior to wave 4)	b2_nd1.dta	hh07_all_dta	household	hhid07	hhid07	
					b2_nd2.dta	hh07_all_dta	disaster type	hhid07, ndtype	hhid07	y
		Community	Book 1 (don't see the community survey instrument in the documentation)	D Agriculture and industry (differs slightly from earlier waves)	bk1_d1.dta	cf07_all_dta	irrigation	commid07, d1type	commid07	y
					bk1_d1a.dta	cf07_all_dta	rice field types	commid07, d8atype	commid07	y

5	2014	6,044 origin households, 9,877 split off households from IFLS 2 (607) and 2+ (207) and 3 (1,787) and 4 (3,273), 4,003 new split offs from IFLS 5, total 15,921	Household			bk1_d2.dta	cf07_all_dta	activity	commid07, d2type	commid07	y
						bk1_d3.dta	cf07_all_dta	crop	commid07, d3type	commid07	y
						bk1_d4.dta	cf07_all_dta	factory	commid07, d4type	commid07	y
				E Community history and climate		bk1_e1.dta	cf07_all_dta	name change	commid07, e1type	commid07	y
						bk1_e2.dta	cf07_all_dta	major event	commid07, e10evt	commid07	y
					F Natural disasters (not included in waves prior to wave 4)	bk1_f1.dta	cf07_all_dta	natural disaster	commid07, f1type	commid07	y
						bk1_f2.dta	cf07_all_dta	damage from natural disaster	commid07, f2type	commid07	y
				Book 2: Community statistical records (don't see the community survey instrument in the documentation))	LK Basic information	bk2.dta	cf07_all_dta	community	commid07	commid07	
					S Statistics	bk2.dta	cf07_all_dta	community	commid07	commid07	
					OL Direct observation (e.g., cleanliness, prosperity, social cohesion)	bk2.dta	cf07_all_dta	community	commid07	commid07	
				Book K: Control Book and Household Roster (IFLS5\IFLS5_all_doc (1)\IFLS5_HH_PQX_Public_Eng\04 IFLS5 HH BOOK K.pdf)	SC Sampling and enumeration record	bk_sc1.dta	hh14_all_dta	household	hhid14	hhid14	
					KRK Housing characteristics (In IFLS1, part of book I)	bk_krk.dta	hh14_all_dta	household	hhid14	hhid14	
				Book 1: Household Expenditures and Knowledge of Health Facilities (IFLS5\IFLS5_all_doc (1)\IFLS5_HH_PQX_Public_Eng\05 IFLS5 HH BOOK 1.pdf)	KS Household expenditures / Consumption - NEED TO CREATE HOUSEHOLD CONSUMPTION EXPENDITURE AGGREGATES TO CORRESPOND TO PREVIOUS WAVES	b1_ks0.dta	hh14_all_dta	household	hhid14	hhid14	
						b1_ks1.dta	hh14_all_dta	household food exp item	hhid14, ks1type	hhid14	y
						b1_ks2.dta	hh14_all_dta	household non-food exp item	hhid14, ks2type	hhid14	y
						b1_ks3.dta	hh14_all_dta	household non-food exp item	hhid14, ks3type	hhid14	y
						b1_ks4.dta	hh14_all_dta	household food item	hhid14, ks4type	hhid14	y
				Book 2: Household Economy (IFLS5\IFLS5_all_doc (1)\IFLS5_HH_PQX_Public_Eng\06 IFLS5 HH BOOK 2.pdf)	KR Housing characteristics (in IFLS1, part of book I)	b2_kr.dta	hh14_all_dta	household	hhid14	hhid14	
					UT Farm business (differs from the earlier waves)	b2_ut1.dta	hh14_all_dta	household	hhid14	hhid14	
						b2_ut2.dta	hh14_all_dta	household paddy harvest	hhid14	hhid14	
						b2_ut3.dta	hh14_all_dta	type of hardship	hhid14, ut3type	hhid14	y

		households				b2_ut 4.dta	hh14_all_dta	household farm business asset type	hhid14, uttype	hhid14	y
					NT Nonfarm business	b2_nt 1.dta	hh14_all_dta	household	hhid14	hhid14	
						b2_nt 2.dta	hh14_all_dta	household non- farm business asset	hhid14, nt_num	hhid14	y
					HR Household assets	b2_hr 1.dta	hh14_all_dta	household asset type	hhid, hrtype	hhid14	y
						b2_hr 2.dta	hh14_all_dta	household asset type	hhid, hr2type	hhid14	y
					ND Natural Disasters (not included in waves prior to wave 4)	b2_nd 1.dta	hh14_all_dta	household	hhid14	hhid14	
						b2_nd 2.dta	hh14_all_dta	disaster type	hhid14, ndtype	hhid14	y
			Communi- nity	Book 1 (don't see the community survey instrument in the documentation)	D Agriculture and industry (differs slightly from earlier waves)	bk1_d 1.dta	cf14_all_dta	irrigation	commid14, d1type	commid1 4	y
						bk1_d 1a.dta	cf14_all_dta	rice field types	commid14, d8atype	commid1 4	y
						bk1_d 2.dta	cf14_all_dta	activity	commid14, d2type	commid1 4	y
						bk1_d 3.dta	cf14_all_dta	crop	commid14, d3type	commid1 4	y
						bk1_d 4.dta	cf14_all_dta	factory	commid14, d4type	commid1 4	y
				E Community history and climate		bk1_e 1.dta	cf14_all_dta	name change	commid14	commid1 4	y
						bk1_e 2.dta	cf14_all_dta	major event	commid14, e10evt	commid1 4	y
					F Natural disasters (not included in waves prior to wave 4)	bk1_f 1.dta	cf14_all_dta	natural disaster	commid14, f1type	commid1 4	y
						bk1_f 2.dta	cf14_all_dta	damage from natural disaster	commid14, f2type	commid1 4	y
				Book 2: Community statistical records (don't see the community survey instrument in the documentation))	LK Basic information	bk2.dta	cf14_all_dta	community	commid14	commid1 4	
					S Statistics	bk2.dta	cf14_all_dta	community	commid14	commid1 4	
					OL Direct observation (e.g., cleanliness, prosperity, social cohesion)	bk2.dta	cf14_all_dta	community	commid14	commid1 4	

None & increasing wealth	Ref.								
2 to 7 & increasing wealth	0.88	0.45	0.053						
8 to 18 & increasing wealth	1.03	0.52	0.048						
19 to 33 & increasing wealth	0.48	0.47	0.302						
34 or more & increasing wealth	1.24	0.47	0.008						
Joint effect with total days of flooding w/in district (count)				0.01	0.01	0.063			
Joint effect with total number of floods w/in district (count)							0.11	0.09	0.221
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	-0.47	0.75	0.533	-0.47	0.75	0.530	-0.45	0.75	0.553
Moved within the same district	0.89	0.97	0.360	0.88	0.97	0.367	0.90	0.97	0.355
Moved within the same regency	-0.91	0.86	0.290	-0.95	0.86	0.270	-0.95	0.86	0.272
Moved within the same province	-1.53	0.91	0.094	-1.48	0.91	0.106	-1.49	0.91	0.102
Moved to another province	-2.28	0.97	0.019	-2.21	0.97	0.023	-2.20	0.97	0.024
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	-1.04	1.17	0.371	-0.96	1.14	0.398	-0.15	0.73	0.842
Wave 3 - 2000	2.79	1.20	0.020	3.03	1.20	0.012	3.70	0.88	0.000
Wave 4 - 2007	1.72	1.04	0.098	1.46	0.95	0.126	1.98	0.79	0.012
Wave 5 - 2014	-2.63	0.96	0.006	-2.67	0.96	0.005	-2.15	0.81	0.008
Indicators of adaptive capacity (waves 3 to 5 only)³									
<i>Linear effects (model with no interactions)</i>									
Know of a place to borrow money	-1.83	0.81	0.024	-1.83	0.81	0.024	-1.83	0.81	0.024
Secured a loan in past 12 months	-0.63	0.52	0.223	-0.63	0.52	0.227	-0.63	0.52	0.226
Received any assistance from government or non-governmental organization in past 12 months ⁴	1.53	0.65	0.014	1.58	0.64	0.014	1.59	0.65	0.013
Received cash assistance	-0.18	0.63	0.770	-0.18	0.63	0.770	-0.16	0.63	0.799

Differential/joint effects with flooding

Know where to borrow & total days of flooding w/in district
(ordinal)

None & knowing where to borrow money	Ref.		
2 to 7 & knowing where to borrow money	6.64	2.96	0.025
8 to 18 & knowing where to borrow money	4.13	3.13	0.187
19 to 33 & knowing where to borrow money	2.89	2.98	0.331
34 or more & knowing where to borrow money	6.12	3.50	0.081

Know where to borrow & total days of flooding w/in district
(count)

0.01 0.05 0.857

Know where to borrow & total number of floods w/in district (count)

0.44 0.86 0.607

Secured a loan & total days of flooding w/in district
(ordinal)

None & secured a loan	Ref.		
2 to 7 & secured a loan	2.05	1.61	0.202
8 to 18 & secured a loan	-0.04	1.83	0.984
19 to 33 & secured a loan	3.74	1.52	0.014
34 or more & secured a loan	1.96	1.92	0.306

Secured a loan & total days of flooding w/in district (count)

0.05 0.03 0.153

Secured a loan & total number of floods w/in district (count)

0.98 0.45 0.031

Received any assistance & total days of flooding w/in district
(ordinal)

None & received any assistance	Ref.		
2 to 7 & received any assistance	-2.36	1.83	0.198
8 to 18 & received any assistance	-0.50	1.90	0.792
19 to 33 & received any assistance	0.62	1.59	0.697
34 or more & received any assistance	0.57	2.02	0.780

Received any assistance & total days of flooding w/in district (count)

0.03 0.04 0.449

Received any assistance & total number of floods w/in district
(count)

0.36 0.44 0.413

Indicator of any increase in surface water w/in district	2.27	0.88	0.010						
Fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814				Ref.					
.0002815 to .0026219				0.50	1.02	0.62			
.0026220 to .0233712				1.78	1.07	0.10			
.0233713 to .3454737				0.91	1.08	0.40			
Fraction of district raster cells w/increase in surface water (continuous) ²							2.76	9.08	0.761
<i>Household characteristics</i>									
Anyone was a farmer in the past year	0.54	0.49	0.270	0.62	0.49	0.207	0.59	0.49	0.232
Owens livestock/poultry/fishpond	1.04	0.45	0.021	1.07	0.45	0.018	1.06	0.45	0.020
Household size	0.25	0.12	0.032	0.25	0.12	0.034	0.25	0.12	0.032
Rural	2.78	0.72	0.000	2.66	0.72	0.000	2.60	0.72	0.000
House is self-owned	-1.19	0.58	0.041	-1.14	0.58	0.050	-1.11	0.58	0.058
Wealth index	-2.82	0.25	0.000	-2.77	0.22	0.000	-2.94	0.16	0.000
Quadratic effect	-0.42	0.05	0.000	-0.42	0.05	0.000	-0.42	0.05	0.000
Joint effect with any increase in surface water w/in district (indicator)	-0.28	0.27	0.299						
Joint effect with fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & increasing wealth				Ref.					
.0002815 to .0026219 & increasing wealth				-0.25	0.32	0.428			
.0026220 to .0233712 & increasing wealth				-0.14	0.34	0.678			
.0233713 to .3454737 & increasing wealth				-0.87	0.34	0.011			
Joint effect with fraction of district raster cells w/increase in surface water (continuous)							-4.42	2.78	0.112
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	-0.46	0.75	0.539	-0.42	0.75	0.573	-0.43	0.75	0.568
Moved within the same district	0.77	0.97	0.429	1.03	0.97	0.292	0.95	0.97	0.330
Moved within the same regency	-1.06	0.87	0.219	-1.00	0.86	0.245	-0.97	0.86	0.262

Moved within the same province	-1.46	0.91	0.110	-1.47	0.92	0.109	-1.49	0.91	0.104
Moved to another province	-2.29	0.97	0.019	-2.24	0.97	0.022	-2.22	0.97	0.022
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	0.07	0.69	0.919	0.14	0.70	0.846	0.09	0.70	0.894
Wave 3 - 2000	4.06	0.69	0.000	4.14	0.69	0.000	4.09	0.69	0.000
Wave 4 - 2007	2.27	0.70	0.001	2.31	0.70	0.001	2.29	0.70	0.001
Wave 5 - 2014	-1.87	0.71	0.008	-1.84	0.71	0.009	-1.86	0.71	0.008
Indicators of adaptive capacity (waves 3 to 5 only)³									
<i>Linear effects (model with no interactions)</i>									
Know of a place to borrow money	-1.82	0.81	0.024	-1.82	0.81	0.024	-1.84	0.81	0.023
Secured a loan in past 12 months	-0.61	0.52	0.243	-0.64	0.52	0.222	-0.63	0.52	0.224
Received any assistance from government or non-governmental organization in past 12 months ⁴	1.56	0.64	0.016	1.57	0.64	0.015	1.59	0.64	0.014
Received cash assistance	-0.22	0.63	0.731	-0.18	0.63	0.770	-0.19	0.63	0.768
<i>Differential/joint effects with flooding</i>									
Know where to borrow & indicator of any increase in surface water w/in district	2.61	1.69	0.123						
Know where to borrow & fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & knowing where to borrow money				Ref.					
.0002815 to .0026219 & knowing where to borrow money				2.92	2.28	0.200			
.0026220 to .0233712 & knowing where to borrow money				3.50	2.27	0.123			
.0233713 to .3454737 & knowing where to borrow money				0.10	1.97	0.961			
Know where to borrow & fraction of district raster cells w/increase in surface water (continuous)							-8.26	15.1	0.584
Secured a loan & indicator of any increase in surface water w/in district	0.94	1.11	0.398						

Secured a loan & fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & secured a loan	Ref.								
.0002815 to .0026219 & secured a loan	0.87	1.40	0.536						
.0026220 to .0233712 & secured a loan	-1.01	1.38	0.465						
.0233713 to .3454737 & secured a loan	-0.21	1.39	0.878						
Secured a loan & fraction of district raster cells w/increase in surface water (continuous)									
							-3.57	11.24	0.751
Received any assistance & indicator of any increase in surface water w/in district	2.45	0.99	0.014						
Received any assistance & fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & received any assistance	Ref.								
.0002815 to .0026219 & received any assistance	0.80	1.24	0.515						
.0026220 to .0233712 & received any assistance	2.04	1.25	0.102						
.0233713 to .3454737 & received any assistance	-0.18	1.21	0.880						
Received any assistance & fraction of district raster cells w/increase in surface water (continuous)									
							-2.36	9.63	0.807
Received cash & indicator of any increase in surface water w/in district	2.27	1.24	0.068						
Received cash & fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & received cash	Ref.								
.0002815 to .0026219 & received cash	0.97	1.46	0.508						
.0026220 to .0233712 & received cash	2.01	1.47	0.171						
.0233713 to .3454737 & received cash	2.15	1.55	0.167						
Received cash & fraction of district raster cells w/increase in surface water (continuous)									
							9.59	11.61	0.409

¹Complete case analysis, n=6,412 across all models (95% of total sample).

²Likelihood Ratio test comparing linear and quadratic models was not significant, indicating the model with the quadratic term did not fit the data better than the linear model.

³These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining the linear effects of all four indicators of adaptive capacity and each of their joint effects with flooding. Each of the models included complete case analysis of n=4,579 observations (94% of total sample in waves 3 to 5).

⁴In the form of money, rice, vegetables, sugar, fruit, meat, snacks, other food, cooking oil, or kerosene.

S3. Models assessing episodic flooding from GSW on household food expenditure share.¹

Variable	Model with episodic flooding w/in district as an indicator variable			Model with episodic flooding w/in district as an ordinal variable			Model with episodic flooding w/in district as a continuous variable		
	Coef.	SE	P-value	Coef.	SE	P-value	Coef.	SE	P-value
<i>Effect of flooding</i>									
Indicator of any episodic flooding w/in district	0.98	1.01	0.329						
Fraction of district raster cells w/episodic flooding (ordinal)									
0 to .0003986				Ref.					
.0003987 to .0040443				0.70	1.06	0.505			
.0040444 to .0475874				1.18	1.04	0.254			
.0475875 to .3115916				2.14	1.12	0.057			
Fraction of district raster cells w/episodic flooding (continuous) ²							5.18	7.11	0.467
<i>Household characteristics</i>									
Anyone was a farmer in the past year	0.56	0.49	0.259	0.56	0.49	0.256	0.60	0.49	0.223

Owns livestock/poultry/fishpond	1.04	0.45	0.021	1.06	0.45	0.020	1.06	0.45	0.020
Household size	0.25	0.12	0.032	0.25	0.12	0.033	0.25	0.12	0.032
Rural	2.66	0.72	0.000	2.58	0.72	0.000	2.60	0.72	0.000
House is self-owned	-1.13	0.58	0.053	-1.16	0.58	0.047	-1.12	0.58	0.055
Wealth index	-2.91	0.29	0.000	-2.70	0.21	0.000	-2.89	0.16	0.000
Quadratic effect	-0.43	0.05	0.000	-0.42	0.05	0.000	-0.42	0.05	0.000
Joint effect with any episodic flooding w/in district (indicator)	-0.15	0.30	0.621						
Joint effect with fraction of district raster cells w/episodic flooding (ordinal)									
0 to .0003986 & increasing wealth				Ref.					
.0003987 to .0040443 & increasing wealth				-0.64	0.33	0.055			
.0040444 to .0475874 & increasing wealth				-0.17	0.32	0.594			
.0475875 to .3115916 & increasing wealth				-0.91	0.34	0.008			
Joint effect with fraction of district raster cells w/episodic flooding (continuous)							-4.56	2.17	0.035
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	-0.45	0.75	0.546	-0.47	0.75	0.536	-0.43	0.75	0.570
Moved within the same district	0.85	0.97	0.383	0.83	0.98	0.394	0.96	0.97	0.324
Moved within the same regency	-0.99	0.87	0.252	-1.04	0.87	0.232	-0.98	0.87	0.257
Moved within the same province	-1.49	0.92	0.105	-1.44	0.92	0.117	-1.48	0.92	0.106
Moved to another province	-2.23	0.98	0.022	-2.31	0.98	0.018	-2.23	0.97	0.022
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	0.07	0.70	0.924	0.12	0.70	0.861	0.11	0.70	0.875
Wave 3 - 2000	4.05	0.69	0.000	4.11	0.69	0.000	4.11	0.69	0.000
Wave 4 - 2007	2.26	0.70	0.001	2.28	0.70	0.001	2.31	0.70	0.001
Wave 5 - 2014	-1.89	0.71	0.007	-1.85	0.71	0.009	-1.85	0.71	0.009

Indicators of adaptive capacity (waves 3 to 5 only)³*Linear effects (model with no interactions)*

Know of a place to borrow money	-1.82	0.81	0.024	-1.81	0.81	0.025	-1.83	0.81	0.023
Secured a loan in past 12 months	-0.62	0.52	0.231	-0.61	0.52	0.240	-0.63	0.52	0.226
Received any assistance from government or non-governmental organization in past 12 months ⁴	1.57	0.64	0.015	1.57	0.64	0.015	1.58	0.64	0.014
Received cash assistance	-0.20	0.63	0.749	-0.20	0.63	0.757	-0.19	0.63	0.767

Differential/joint effects with flooding

Know where to borrow & indicator of any episodic flooding w/in district	5.43	1.86	0.004						
Know where to borrow & fraction of district raster cells w/episodic flooding (ordinal)									
0 to .0003986 & knowing where to borrow money				Ref.					
.0003987 to .0040443 & knowing where to borrow money				1.06	2.46	0.668			
.0040444 to .0475874 & knowing where to borrow money				3.41	2.25	0.129			
.0475875 to .3115916 & knowing where to borrow money				0.09	1.93	0.961			
Know where to borrow & fraction of district raster cells w/episodic flooding (continuous)							-5.84	11.80	0.620
Secured a loan & indicator of any episodic flooding w/in district	1.61	1.25	0.198						
Secured a loan & fraction of district raster cells w/episodic flooding (ordinal)									
0 to .0003986 & secured a loan				Ref.					
.0003987 to .0040443 & secured a loan				0.74	1.42	0.602			
.0040444 to .0475874 & secured a loan				-0.72	1.34	0.589			
.0475875 to .3115916 & secured a loan				-0.23	1.43	0.870			

Secured a loan & fraction of district raster cells w/episodic flooding (continuous)				-1.03	8.57	0.904
Received any assistance & indicator of any episodic flooding w/in district	1.99	1.09	0.068			
Received any assistance & fraction of district raster cells w/episodic flooding (ordinal)						
0 to .0003986 & received any assistance				Ref.		
.0003987 to .0040443 & received any assistance				1.58	1.26	0.210
.0040444 to .0475874 & received any assistance				2.00	1.22	0.100
.0475875 to .3115916 & received any assistance				0.20	1.22	0.869
Received any assistance & fraction of district raster cells w/episodic flooding (continuous)				-1.29	7.39	0.861
Received cash & indicator of any episodic flooding w/in district	2.13	1.37	0.120			
Received cash & fraction of district raster cells w/episodic flooding (ordinal)						
0 to .0003986 & received cash				Ref.		
.0003987 to .0040443 & received cash				1.40	1.47	0.340
.0040444 to .0475874 & received cash				2.26	1.47	0.125
.0475875 to .3115916 & received cash				2.30	1.54	0.137
Received cash & fraction of district raster cells w/episodic flooding (continuous)				8.77	9.22	0.342

¹Complete case analysis, n=6,412 across all models (95% of total sample).

²Likelihood Ratio test comparing linear and quadratic models was not significant, indicating the model with the quadratic term did not fit the data better than the linear model.

³These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining the linear effects of all four indicators of adaptive capacity and each of their joint effects with flooding. Each of the models included complete case analysis of n=4,579 observations (94% of total sample in waves 3 to 5)

⁴In the form of money, rice, vegetables, sugar, fruit, meat, snacks, other food, cooking oil, or kerosene.

S4. Models assessing total days of flooding and total number of floods from DFO on household dietary diversity score.¹

Variable	Model with total days of flooding as ordinal variable			Model with total days of flooding as count variable			Model with total number of floods count variable		
	OR	SE	P-value	OR	SE	P-value	OR	SE	P-value
<i>Effect of flooding</i>									
Total days of flooding w/in district (ordinal)									
None	Ref.								
2 to 7	1.33	0.17	0.028						
8 to 18	1.44	0.20	0.008						
19 to 33	1.21	0.16	0.137						
34 or more	1.58	0.23	0.002						
Total days of flooding w/in district (count) ²				1.006	0.003	0.020			
Total number of floods w/in district (count)							1.29	0.11	0.003
Quadratic effect							0.96	0.02	0.021
<i>Household characteristics</i>									
Anyone was a farmer in the past year	1.10	0.08	0.186	1.10	0.08	0.190	1.10	0.08	0.192
Owns livestock/poultry/fishpond	1.28	0.08	0.000	1.27	0.08	0.00	1.28	0.08	0.000
Household size	1.31	0.02	0.000	1.32	0.02	0.00	1.31	0.02	0.000
Rural	1.33	0.15	0.014	1.32	0.15	0.02	1.30	0.15	0.024
House is self-owned	1.90	0.16	0.000	1.90	0.16	0.00	1.90	0.16	0.000
Wealth index ³	1.27	0.03	0.000	1.31	0.03	0.000	1.27	0.03	0.000
Joint effect with total days of flooding w/in district (count)				1.00	0.00	0.048			
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	1.10	0.14	0.446	1.09	0.13	0.488	1.10	0.14	0.457
Moved within the same district	1.25	0.20	0.165	1.25	0.20	0.173	1.25	0.20	0.174

Moved within the same regency	0.89	0.13	0.388	0.88	0.13	0.374	0.89	0.13	0.398
Moved within the same province	0.44	0.07	0.000	0.43	0.07	0.000	0.44	0.07	0.000
Moved to another province	0.85	0.13	0.294	0.86	0.13	0.326	0.85	0.13	0.298
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	1.55	0.24	0.005	1.93	0.31	0.000	1.33	0.14	0.005
Wave 3 - 2000	2.37	0.39	0.000	2.87	0.49	0.000	2.17	0.27	0.000
Wave 4 - 2007	1.71	0.24	0.000	1.78	0.24	0.000	1.46	0.16	0.001
Wave 5 - 2014	1.21	0.16	0.135	1.36	0.18	0.021	1.10	0.13	0.420
Indicators of adaptive capacity (waves 3 to 5 only)⁴									
<i>Linear effects (model with no interactions)</i>									
Know of a place to borrow money	1.51	0.18	0.001	1.46	0.17	0.002	1.47	0.18	0.001
Secured a loan in past 12 months	1.46	0.12	0.000	1.45	0.12	0.000	1.45	0.11	0.000
Received any assistance from government or non-governmental organization in past 12 months ⁵	1.31	0.13	0.005	1.33	0.13	0.004	1.32	0.13	0.004
Received cash assistance	0.77	0.07	0.006	0.77	0.07	0.005	0.76	0.07	0.003
<i>Differential/joint effects with flooding</i>									
Know where to borrow & total days of flooding w/in district (ordinal)									
None & knowing where to borrow money	Ref.								
2 to 7 & knowing where to borrow money	1.27	0.55	0.587						
8 to 18 & knowing where to borrow money	0.88	0.40	0.776						
19 to 33 & knowing where to borrow money	1.34	0.58	0.504						
34 or more & knowing where to borrow money	0.64	0.33	0.383						
Know where to borrow & total days of flooding w/in district (count)				0.99	0.01	0.341			
Know where to borrow & total number of floods w/in district (count)							0.84	0.11	0.167
Secured a loan & total days of flooding w/in district (ordinal)									
None & secured a loan	Ref.								

2 to 7 & secured a loan	1.21	0.30	0.449						
8 to 18 & secured a loan	0.81	0.22	0.439						
19 to 33 & secured a loan	0.99	0.23	0.970						
34 or more & secured a loan	0.91	0.26	0.734						
Secured a loan & total days of flooding w/in district (count)				1.00	0.00	0.737			
Secured a loan & total number of floods w/in district (count)							1.03	0.07	0.664
Received any assistance & total days of flooding w/in district (ordinal)									
None & received any assistance	Ref.								
2 to 7 & received any assistance	0.71	0.19	0.195						
8 to 18 & received any assistance	1.45	0.41	0.189						
19 to 33 & received any assistance	0.89	0.21	0.625						
34 or more & received any assistance	1.42	0.43	0.244						
Received any assistance & total days of flooding w/in district (count)				1.01	0.01	0.341			
Received any assistance & total number of floods w/in district (count)							0.99	0.07	0.825
Received cash & total days of flooding w/in district (ordinal)									
None & received cash	Ref.								
2 to 7 & received cash	1.07	0.41	0.861						
8 to 18 & received cash	1.11	0.34	0.732						
19 to 33 & received cash	0.98	0.26	0.935						
34 or more & received cash	0.85	0.27	0.609						
Received cash & total days of flooding w/in district (count)				1.00	0.01	0.370			
Received cash & total number of floods w/in district (count)							0.97	0.07	0.664

¹Complete case analysis, n=6,513 across all models (97% of total sample).

²Likelihood Ratio test comparing linear and quadratic models was not significant, indicating the model with the quadratic term did not fit the data better than the linear model.

³Joint effect not included in models with total days of flooding as ordinal variable and total number of floods as count variable due to non-significant Likelihood Ratio test and multicollinearity.

⁴These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining

the linear effects of all four indicators of adaptive capacity and each of their joint effects with flooding. Each of the models included complete case analysis of n=4,579 observations (94% of total sample in waves 3 to 5).

⁵In the form of money, rice, vegetables, sugar, fruit, meat, snacks, other food, cooking oil, or kerosene.

S5. Models assessing increase in surface water from GSW on household dietary diversity score.¹

Variable	Model with increase in surface water w/in district as indicator variable			Model with increase in surface water w/in district as ordinal variable			Model with increase in surface water w/in district as continuous variable		
	OR	SE	P-value	OR	SE	P-value	OR	SE	P-value
<i>Effect of flooding</i>									
Indicator of any increase in surface water w/in district	1.58	0.24	0.002						
Fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814				Ref.					
.0002815 to .0026219				1.27	0.22	0.17			
.0026220 to .0233712				1.15	0.21	0.45			
.0233713 to .3454737				1.16	0.21	0.426			
Fraction of district raster cells w/increase in surface water (continuous) ²							0.69	1.04	0.805
<i>Household characteristics</i>									
Anyone was a farmer in the past year	1.10	0.08	0.190	1.10	0.08	0.186	1.10	0.08	0.188
Owens livestock/poultry/fishpond	1.29	0.08	0.00	1.29	0.08	0.000	1.29	0.08	0.000
Household size	1.32	0.02	0.00	1.32	0.02	0.000	1.32	0.02	0.000
Rural	1.39	0.16	0.00	1.36	0.16	0.008	1.35	0.16	0.010
House is self-owned	1.87	0.16	0.00	1.88	0.16	0.000	1.89	0.16	0.000
Wealth index ³	1.33	0.05	0.00	1.27	0.03	0.00	1.27	0.03	0.000

Joint effect with any increase in surface water w/in district (indicator)	0.93	0.04	0.10						
Joint effect with any episodic flooding w/in district (indicator)									
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	1.09	0.13	0.501	1.09	0.13	0.503	1.09	0.13	0.499
Moved within the same district	1.22	0.20	0.224	1.24	0.20	0.185	1.25	0.20	0.177
Moved within the same regency	0.86	0.12	0.303	0.88	0.12	0.359	0.88	0.13	0.376
Moved within the same province	0.44	0.07	0.000	0.43	0.07	0.000	0.44	0.07	0.000
Moved to another province	0.85	0.13	0.305	0.86	0.13	0.335	0.86	0.14	0.345
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	1.34	0.13	0.002	1.34	0.13	0.002	1.34	0.13	0.002
Wave 3 - 2000	1.94	0.19	0.000	1.94	0.19	0.000	1.94	0.19	0.000
Wave 4 - 2007	1.37	0.13	0.001	1.37	0.13	0.001	1.36	0.13	0.001
Wave 5 - 2014	1.05	0.10	0.604	1.05	0.10	0.611	1.05	0.10	0.635
Indicators of adaptive capacity (waves 3 to 5 only)⁴									
<i>Linear effects (model with no interactions)</i>									
Know of a place to borrow money	1.47	0.18	0.001	1.47	0.18	0.001	1.47	0.18	0.001
Secured a loan in past 12 months	1.45	0.12	0.000	1.45	0.12	0.000	1.45	0.11	0.000
Received any assistance from government or non-governmental organization in past 12 months ⁵	1.33	0.13	0.003	1.33	0.13	0.003	1.33	0.13	0.003
Received cash assistance	0.77	0.07	0.005	0.77	0.07	0.006	0.77	0.07	0.006
<i>Differential/joint effects with flooding</i>									
Know where to borrow & indicator of any increase in surface water w/in district	0.92	0.23	0.753						
Know where to borrow & fraction of district raster cells w/increase in surface water (ordinal)									
0 to .0002814 & knowing where to borrow money				Ref.					
.0002815 to .0026219 & knowing where to borrow money				1.86	0.64	0.071			
.0026220 to .0233712 & knowing where to borrow money				0.67	0.22	0.227			

.0233713 to .3454737 & knowing where to borrow money	0.53	0.16	0.031			
Know where to borrow & fraction of district raster cells w/increase in surface water (continuous)				0.00	0.01	0.012
Secured a loan & indicator of any increase in surface water w/in district	1.09	0.18	0.613			
Secured a loan & fraction of district raster cells w/increase in surface water (ordinal)						
0 to .0002814 & secured a loan	Ref.					
.0002815 to .0026219 & secured a loan	1.35	0.29	0.157			
.0026220 to .0233712 & secured a loan	1.38	0.29	0.128			
.0233713 to .3454737 & secured a loan	0.88	0.19	0.542			
Secured a loan & fraction of district raster cells w/increase in surface water (continuous)				0.09	0.16	0.159
Received any assistance & indicator of any increase in surface water w/in district	1.13	0.17	0.390			
Received any assistance & fraction of district raster cells w/increase in surface water (ordinal)						
0 to .0002814 & received any assistance	Ref.					
.0002815 to .0026219 & received any assistance	1.17	0.22	0.387			
.0026220 to .0233712 & received any assistance	1.26	0.23	0.208			
.0233713 to .3454737 & received any assistance	1.13	0.20	0.505			
Received any assistance & fraction of district raster cells w/increase in surface water (continuous)				11.51	16.00	0.079
Received cash & indicator of any increase in surface water w/in district	1.15	0.21	0.446			
Received cash & fraction of district raster cells w/increase in surface water (ordinal)						
0 to .0002814 & received cash	Ref.					
.0002815 to .0026219 & received cash	1.19	0.26	0.415			
.0026220 to .0233712 & received cash	1.12	0.24	0.615			
.0233713 to .3454737 & received cash	1.29	0.29	0.267			
Received cash & fraction of district raster cells w/increase in surface water (continuous)				4.90	8.23	0.344

¹Complete case analysis, n=6,513 across all models (97% of total sample).

²Likelihood Ratio test comparing linear and quadratic models was not significant, indicating the model with the quadratic term did not fit the data better than the linear model.

³Joint effect not included in models with increase in surface water as ordinal and continuous variables due to non-significant Likelihood Ratio test.

⁴These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining

Joint effect with any episodic flooding w/in district (indicator)	0.90	0.04	0.025						
Extent household moved during entire study period									
Did not move	Ref.			Ref.			Ref.		
Moved within the same village	1.09	0.13	0.490	1.08	0.13	0.512	1.09	0.13	0.503
Moved within the same district	1.22	0.20	0.217	1.23	0.20	0.202	1.24	0.20	0.178
Moved within the same regency	0.87	0.12	0.315	0.87	0.12	0.327	0.88	0.13	0.372
Moved within the same province	0.44	0.07	0.000	0.43	0.07	0.000	0.44	0.07	0.000
Moved to another province	0.86	0.13	0.324	0.86	0.13	0.327	0.86	0.14	0.345
Survey wave									
Wave 1 - 1993	Ref.			Ref.			Ref.		
Wave 2 - 1997	1.34	0.13	0.002	1.34	0.13	0.002	1.34	0.13	0.002
Wave 3 - 2000	1.94	0.19	0.000	1.94	0.19	0.000	1.94	0.19	0.000
Wave 4 - 2007	1.37	0.13	0.001	1.37	0.13	0.001	1.36	0.13	0.001
Wave 5 - 2014	1.05	0.10	0.605	1.05	0.10	0.614	1.05	0.10	0.630
Indicators of adaptive capacity (waves 3 to 5 only)⁴									
<i>Linear effects (model with no interactions)</i>									
Know of a place to borrow money	1.48	0.18	0.001	1.47	0.18	0.001	1.47	0.18	0.001
Secured a loan in past 12 months	1.45	0.12	0.000	1.45	0.12	0.000	1.45	0.11	0.000
Received any assistance from government or non-governmental organization in past 12 months ⁵	1.33	0.13	0.003	1.33	0.13	0.003	1.33	0.13	0.003
Received cash assistance	0.77	0.07	0.005	0.77	0.07	0.005	0.77	0.07	0.006
<i>Differential/joint effects with flooding</i>									
Know where to borrow & indicator of any episodic flooding w/in district	0.89	0.24	0.656						
Know where to borrow & fraction of district raster cells w/episodic flooding (ordinal)									
0 to .0003986 & knowing where to borrow money				Ref.					
.0003987 to .0040443 & knowing where to borrow money				2.23	0.84	0.034			
.0040444 to .0475874 & knowing where to borrow money				0.79	0.26	0.476			
.0475875 to .3115916 & knowing where to borrow money				0.60	0.17	0.074			
Know where to borrow & fraction of district raster cells w/episodic flooding (continuous)							0.01	0.02	0.010

Secured a loan & indicator of any episodic flooding w/in district	1.17	0.22	0.401				
Secured a loan & fraction of district raster cells w/episodic flooding (ordinal)							
0 to .0003986 & secured a loan				Ref.			
.0003987 to .0040443 & secured a loan				1.47	0.32	0.075	
.0040444 to .0475874 & secured a loan				1.44	0.29	0.076	
.0475875 to .3115916 & secured a loan				0.86	0.19	0.473	
Secured a loan & fraction of district raster cells w/episodic flooding (continuous)					0.15	0.20	0.147
Received any assistance & indicator of any episodic flooding w/in district	0.99	0.16	0.946				
Received any assistance & fraction of district raster cells w/episodic flooding (ordinal)							
0 to .0003986 & received any assistance				Ref.			
.0003987 to .0040443 & received any assistance				1.23	0.23	0.276	
.0040444 to .0475874 & received any assistance				1.23	0.22	0.245	
.0475875 to .3115916 & received any assistance				1.14	0.21	0.470	
Received any assistance & fraction of district raster cells w/episodic flooding (continuous)					4.16	4.48	0.184
Received cash & indicator of any episodic flooding w/in district	1.04	0.21	0.862				
Received cash & fraction of district raster cells w/episodic flooding (ordinal)							
0 to .0003986 & received cash				Ref.			
.0003987 to .0040443 & received cash				1.28	0.28	0.254	
.0040444 to .0475874 & received cash				1.18	0.26	0.458	
.0475875 to .3115916 & received cash				1.25	0.28	0.316	
Received cash & fraction of district raster cells w/episodic flooding (continuous)					3.10	4.14	0.398

¹Complete case analysis, n=6,513 across all models (97% of total sample).

²Likelihood Ratio test comparing linear and quadratic models was not significant, indicating the model with the quadratic term did not fit the data better than the linear model.

³Joint effect not included in models with episodic flooding as ordinal and continuous variables due to non-significant Likelihood Ratio test.

⁴These indicators were only available in IFLS waves 3 to 5. For each of the 3 flood variables, a series of five, multi-level linear regression models were fit adjusted for household characteristics and survey wave examining the linear effects of all four indicators of adaptive capacity and each of their joint effects with flooding. Each of the models included complete case analysis of n=4,579 observations (94% of total sample in waves 3 to 5).

⁵In the form of money, rice, vegetables, sugar, fruit, meat, snacks, other food, cooking oil, or kerosene.