
Leveraging Compact Satellite Embeddings and Graph Neural Networks for Large-Scale Poverty Mapping

Markus B. Pettersson

Chalmers University of Technology
markus.pettersson@chalmers.se

Adel Daoud

Chalmers University of Technology
Linköping University

Abstract

Accurate, fine-grained poverty maps remain scarce across much of the Global South. While Demographic and Health Surveys (DHS) provide high-quality socioeconomic data, their spatial coverage is limited and reported coordinates are randomly displaced for privacy, further reducing their quality. We propose a graph-based approach leveraging low-dimensional *AlphaEarth* satellite embeddings to predict cluster-level wealth indices across Sub-Saharan Africa. By modeling spatial relations between surveyed and unlabeled locations, and by introducing a probabilistic “fuzzy label” loss to account for coordinate displacement, we improve the generalization of wealth predictions beyond existing surveys. Our experiments on 37 DHS datasets (2017–2023) show that incorporating graph structure slightly improves accuracy compared to “image-only” baselines, demonstrating the potential of compact EO embeddings for large-scale socioeconomic mapping.

1 Introduction

Accurate local-scale poverty estimates are crucial for targeting interventions [Daoud and Dubhashi, 2023], yet such data remain scarce across much of the developing world. The most reliable socioeconomic information in many countries still comes from household surveys such as the Demographic and Health Surveys (DHS) [ICF, 2025]. Because these surveys cover only a limited set of locations, recent research has sought to predict local wealth or poverty indicators directly from satellite data [Jean et al., 2016, Yeh et al., 2020, Pettersson et al., 2023, Daoud et al., 2023].

A common approach trains regression models to predict a survey-based wealth index, such as the International Wealth Index (IWI) used here [Smits and Steendijk, 2015], from remote-sensing imagery such as Landsat or Sentinel-2 [Pettersson et al., 2023]. However, these methods often ignore spatial dependencies between nearby settlements. Neighboring communities tend to share infrastructure, markets, and environmental context, introducing autocorrelation that purely image-based models overlook. While including latitude and longitude as model inputs can partially capture this, such models fail to generalize to unseen countries and depend on absolute coordinates rather than spatial relationships.

A further complication arises from the way DHS anonymizes survey coordinates: each cluster’s location is randomly displaced by up to 2 km in urban and 10 km in rural areas [Burgert et al., 2013]. Consequently, the reported coordinates do not indicate the true village or neighborhood, introducing spatial uncertainty into all analyses [Kakooei and Daoud, 2024].

To address both issues jointly, we propose representing survey and nearby settlement locations as nodes in a graph, where edges capture geographic proximity. Using compact, information-rich *AlphaEarth* embeddings derived from satellite imagery [Brown et al., 2025], we train neural and graph-based models that incorporate these relationships. We further introduce a “fuzzy label” loss function that explicitly accounts for the DHS coordinate perturbations by weighting candidate settlement locations according to their displacement likelihood. Together, these ideas allow wealth

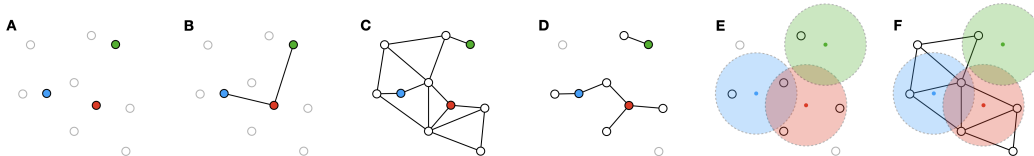


Figure 1: Structures of the six methods evaluated. In this figure, The colored nodes are different survey locations from the DHS while the white are settlement locations from GeoNames. Both of these have corresponding AlphaEarth embeddings x_i . The grayed out GeoNames locations in A, B, D and E signify that they are not included in the model training. For E and F, where the embeddings of the survey clusters are not included, the larger circle highlights the possible displacement area which assigns fuzzy labels to nearby GeoNames nodes.

prediction models to generalize spatially while remaining computationally lightweight enough for continental-scale training.

2 Data

We used 37 DHS surveys conducted between 2017 and 2023 across 25 Sub-Saharan African countries, comprising approximately 16,600 clusters, roughly corresponding to villages in rural areas or neighborhoods in urban areas Burgert et al. [2013]. For each cluster, we computed the mean International Wealth Index (IWI), a standardized measure of material wealth ranging from 0 to 100 Smits and Steendijk [2015].

To preserve respondent privacy, DHS displaces cluster coordinates randomly: urban clusters are perturbed by up to 2 km, and rural clusters by up to 10 km. This introduces spatial noise, making the reported coordinates approximate rather than exact; a key motivation for our “fuzzy label” approach. A more detailed discussion on this can be found in Appendix A.

As a reference set of potential settlement locations in the given countries, we extracted >167,000 points labeled as “Populated places” from the GeoNames database [Wick, 2005]. While this dataset is incomplete and is sure to contain selection bias (i.e., places with better connectivity reports more than those will worse), it provides a practical set of candidate locations for modeling both surveyed and surveyed villages.

For every location, whether from DHS or GeoNames, we obtain a 64-dimensional ALphaEarth embedding from Google Earth Engine, using a 500 m buffer around each coordinate and retrieving the mean pixel value [Brown et al., 2025, Gorelick et al., 2017]. These embeddings provide compact, informative representations of local land use and environmental context, which serve as the primary input for all our predictive models.

The smaller footprint allows us to work with continental scale data in memory, opening the door for utilizing powerful graph architectures. As an example, the corresponding dataset using Sentinel-2 data would require 51.5 GB, but with AlphaEarth, this becomes 45.6 MB, easily small enough to fit on any GPU [European Space Agency, 2015].

3 Method

We evaluate six modeling approaches that combine three design dimensions: (1) whether spatial relationships are modeled using a graph; (2) whether auxiliary unlabeled locations are included; and (3) whether fuzzy labels are used to account for coordinate displacement. Figure 1 summarizes these variants.

(A) Baseline: DHS Points. A simple multilayer perceptron predicts cluster-level IWI directly from the AlphaEarth embedding at the reported survey coordinate. This represents the standard “image-only” approach.

(B) DHS graph. We connect DHS clusters within each survey into a graph, drawing edges between locations less than 100 km apart, weighted inversely by distance. A graph convolutional network (GCN) then propagates information across nearby clusters [Kipf, 2016].

(C) Full graph (DHS + GeoNames). We expand the graph by adding unlabeled settlement nodes from the GeoNames database, linking each to its eight nearest neighbors (within 100 km). This allows the model to reason about relationships between surveyed and unsurveyed places.

(D) Ego graphs. For each DHS cluster, we construct a local subgraph connecting it to all nearby GeoNames settlements within its possible displacement radius (2-10 km). Edge weights represent the probability that each candidate is the true location given the DHS displacement mechanism.

To incorporate the uncertainty in cluster coordinates, we define a “fuzzy label loss function”:

$$\mathcal{L} = \frac{1}{N_l} \sum_{j=1}^{N_l} \sum_{i=1}^{N_s} P_{ji} (f_{\theta}(x_i) - y_j)^2$$

where (y_j) is the wealth label for survey cluster (j) , (x_i) is the embedding for candidate location (i) , and (P_{ji}) is the normalized probability that (i) is the true position of cluster (j) . This formulation effectively spreads each label across its plausible true locations.

(E) Points With Fuzzy labels. A non-graph version using the fuzzy loss, predicting wealth from embeddings of candidate locations only. The <3 % of reported survey coordinates with no candidate location within its possible displacement radius (due to missingness in GeoNames) gets included in the set as a “phantom location”.

(F) Graph With Fuzzy labels. A full GCN trained with the fuzzy-label objective, allowing message passing between both fuzzy-labeled, unlabeled, and phantom nodes.

3.1 Model and training

For all methods, we used small three-layer neural networks or equivalent GCNs with matched parameter counts for graphs. Methods A-D were trained with Mean-squared error loss (MSE), while E and F used the Fuzzy loss described above. Cross-validation was conducted across surveys to test generalization to unseen national contexts. More details on models and the cross validation can be found in Appendix B.

4 Results

We evaluate the performance of the methods on held-out cross-validation folds and the results can be found in Table 1.

Table 1: Results for the different method (mean \pm standard deviation across CV folds). Lower MAE and higher R^2 indicate better performance.

Method	GeoNames	Graph	Fuzzy labels	MAE \downarrow	R^2 \uparrow
(A) DHS Points				8.657 ± 0.343	0.546 ± 0.047
(B) DHS Graph		✓		8.682 ± 0.367	0.546 ± 0.040
(C) Full Graph	✓	✓		8.588 ± 0.472	0.546 ± 0.043
(D) Ego Graphs	✓	✓		8.354 ± 0.436	0.569 ± 0.047
(E) Points w. Fuzzy	✓		✓	9.217 ± 0.539	0.465 ± 0.075
(F) Graph w. Fuzzy	✓	✓	✓	9.665 ± 0.322	0.406 ± 0.084

Models incorporating spatial structure and unlabeled settlements (C–D) modestly outperform the image-only baseline (A), suggesting that nearby settlements contain complementary information not captured by single-site embeddings. The Ego Graphs model (D) performs best overall, indicating that constraining message passing to the plausible displacement radius helps leverage local spatial context without introducing noise from distant or unrelated nodes.

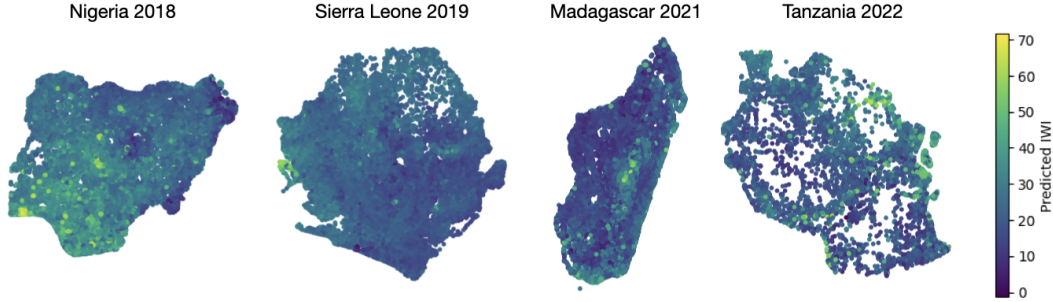


Figure 2: Predicted cluster-level IWI across GeoNames settlements in selected countries. Coverage and spatial density vary by country.

In contrast, models trained with fuzzy labels (E–F) perform notably worse. This suggests that while accounting for coordinate uncertainty is conceptually appealing, the reported displaced coordinates may still provide more reliable supervision than the set of nearby candidate locations. Possible reasons for this are discussed in the following section.

Figure 2 illustrates model predictions across all GeoNames settlements for a subset of countries. The resulting maps exhibit plausible spatial patterns, with smoother regional trends and higher predicted wealth near urban centers, broadly consistent with previous findings [Bank, 2009].

5 Discussion and Outlook

The poor performance of fuzzy-label variants highlights the difficulty of learning under spatial uncertainty. Although the DHS coordinate perturbation motivates such an approach, the GeoNames dataset is incomplete and likely biased toward larger or more accessible settlements. In addition, our current method does not account for the prior of the DHS displacement mechanism: in order to make a representative survey, clusters are more likely to be located in settlements with higher population. Our current probabilistic weighting treats any two settlements with the same distance to the reported coordinate as equally likely, underestimating this prior knowledge. Future work could address these issues by incorporating ancillary data such as population density when defining candidate locations and label weights [Kakoei and Daoud, 2024]. Otherwise, such biases could lead to e.g. misallocation of aid resources due to biased mapping.

Beyond methodological refinements, our experiments demonstrate the practicality of large scale graph learning with modern Earth observation embeddings. The compactness of representations such as AlphaEarth enables training on national or continental graphs with relatively minimal computational cost, making spatially aware poverty mapping feasible even on modest hardware. In the future, other embeddings such as ESA’s Terramind or Clay could be included for comparison [Jakubik et al., 2025, Clay Foundation, 2025].

Moving forward, several directions appear promising. First, refining the representation of spatial relationships could yield substantial gains. Rather than connecting nodes purely by Euclidean distance, edges could be weighted by estimated travel time derived from road networks or terrain models, better capturing real-world accessibility and socioeconomic interaction. Second, obtaining and incorporating the census enumeration areas or sampling frames actually used by DHS would provide a much more accurate set of candidate locations, replacing the incomplete GeoNames database. This, in combination with population weighting, would allow fuzzy-label models to operate on realistic spatial priors tied to how survey clusters are selected in practice. Finally, future work could explore attention-based or hierarchical graph neural networks and population- or time-weighted embeddings to capture multi-scale and temporal socioeconomic dynamics.

In summary, our study shows that high-quality, low-dimensional EO representations make it possible to train graph-based models at scale. We also show that encoding spatial relationships, especially within the known uncertainty bounds of survey locations, offers a tangible path toward more reliable, fine-grained poverty estimation in data-scarce regions.

References

- World Bank. World development report 2009: Reshaping economic geography. Technical report, World Bank, 2009. URL <http://hdl.handle.net/10986/5991>. License: CC BY 3.0 IGO.
- Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.
- Clara R. Burgert, Josh Colston, Thea Roy, and Blake Zachary. Geographic displacement procedure and georeferenced data release policy for the demographic and health surveys. Dhs spatial analysis reports no. 7, The DHS Program / ICF International, Calverton, Maryland, USA, September 2013. URL <https://www.dhsprogram.com/publications/publication-SAR7-Spatial-Analysis-Reports.cfm>. Publication ID: SAR7. Available from The DHS Program website.
- Clay Foundation. Clay foundation model, 2025. URL <https://clay-foundation.github.io/model/index.html>. Accessed: 2025-10-21.
- Adel Daoud and Devdatt Dubhashi. Statistical Modeling: The Three Cultures. 5(1), 2023. ISSN 2644-2353, 2688-8513. doi: 10.1162/99608f92.89f6fe66. URL <https://hdsr.mitpress.mit.edu/pub/uo4hjc6/release/1>.
- Adel Daoud, Felipe Jordán, Makkunda Sharma, Fredrik Johansson, Devdatt Dubhashi, Sourabh Paul, and Subhashis Banerjee. Using satellite images and deep learning to measure health and living standards in india. *Social Indicators Research*, 167(1):475–505, 2023.
- European Space Agency. *Sentinel-2 User Handbook*. European Space Agency, 2015. URL https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook. Accessed October 20, 2025.
- Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.
- Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 2017. doi: 10.1016/j.rse.2017.06.031. URL <https://doi.org/10.1016/j.rse.2017.06.031>.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.
- ICF. The DHS Program. <https://www.dhsprogram.com>, 2025. Funded by USAID. Accessed October 20, 2025.
- Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. Terramind: Large-scale generative multimodality for earth observation. *arXiv preprint arXiv:2504.11171*, 2025.
- Neal Jean, Marshall Burke, Michael Xie, W Matthew Alampay Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301): 790–794, 2016.
- Mohammad Kakooei and Adel Daoud. Increasing the confidence of predictive uncertainty: earth observations and deep learning for poverty estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024.
- TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

- Markus B Pettersson, Mohammad Kakooei, Julia Ortheden, Fredrik D Johansson, and Adel Daoud. Time series of satellite imagery improve deep learning estimates of neighborhood-level poverty in africa. In *IJCAI*, pages 6165–6173, 2023.
- Jeroen Smits and Roel Steendijk. The international wealth index (iwi). *Social indicators research*, 122(1):65–85, 2015.
- Marc Wick. Geonames geographical database, 2005. URL <https://www.geonames.org>. Accessed: 2025-10-11.
- Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):2583, 2020.

A Displacement mechanism

The DHS cluster displacement procedure, described in detail by Smits and Steendijk [2015], anonymizes household survey coordinates through a two-step randomization process. First, a random direction is drawn uniformly from $[0, 2\pi)$. Second, a displacement distance is sampled according to cluster type, which is reported for all clusters in the survey. For urban clusters, the distance is drawn uniformly from 0–2 km. For rural clusters, 99% are displaced by a random distance uniformly distributed between 0–5 km, and the remaining 1% between 0–10 km. The reported cluster coordinate is then obtained by displacing the true location by the sampled distance along the sampled direction. This randomized offset ensures respondent privacy while preserving approximate spatial structure for analytical use.

Let d denote the distance between the true and reported coordinates. The likelihood of observing the reported coordinate given the true location will thus be

$$L(d) = \begin{cases} \frac{1}{\pi 2^2} & d \leq 2 \\ 0 & d > 2 \end{cases}$$

for urban clusters and

$$L(d) = \begin{cases} \frac{0.99}{\pi 5^2} + \frac{0.01}{\pi 10^2} & d \leq 5 \\ \frac{0.01}{\pi 10^2} & 5 < d \leq 10 \\ 0 & d > 10 \end{cases}$$

for rural clusters.

B Model details

As was described in the text, we select one of two basic model architectures for all methods, depending on whether it uses a graph structure or not. The models were implemented in PyTorch, using the PyTorch Geometric package for GCN Convolution [Fey and Lenssen, 2019]. These two architectures are included below:

Listing 1: Simple Feedforward Neural Network used for methods (A) and (E).

```
class SimpleNN(torch.nn.Module):
    def __init__(self, input_dim, hidden_dim, output_dim):
        super().__init__()
        self.fc1 = torch.nn.Linear(input_dim, hidden_dim)
        self.fc2 = torch.nn.Linear(hidden_dim, hidden_dim)
        self.fc3 = torch.nn.Linear(hidden_dim, output_dim)

    def forward(self, x, edge_index, edge_attr=None, batch=None):
        # edge_index, edge_attr, and batch are unused in this model
        x = F.relu(self.fc1(x))
        x = F.relu(self.fc2(x))
        x = self.fc3(x)
        return x
```

Listing 2: Simple Graph Neural Network used for methods (B), (C), (D) and (F).

```
class SimpleGNN(torch.nn.Module):
    def __init__(self, input_dim, hidden_dim, output_dim):
        super().__init__()
        self.conv1 = GCNConv(input_dim, hidden_dim)
        self.conv2 = GCNConv(hidden_dim, hidden_dim)
        self.lin = torch.nn.Linear(hidden_dim, output_dim)

    def forward(self, x, edge_index, edge_attr=None, batch=None):
```

```

# edge_attr here is inverse distance
edge_weight = edge_attr.view(-1)
x = self.conv1(x, edge_index, edge_weight=edge_weight)
x = F.relu(x)
x = self.conv2(x, edge_index, edge_weight=edge_weight)
x = F.relu(x)

x = self.lin(x)
return x

```

For all methods, we used an input dimension of 64 (from the AlphaEarth embeddings), a hidden dimension of 128 and an output dimension of 1 (from the scalar IWI). The two models have the same number of parameters, to keep comparisons fair. For each method, we trained the corresponding model with Mean-Squared Error Loss (except for method E and F which use the Fuzzy loss described in the paper) and AdamW for 500 epochs, using checkpointing for the lowest validation loss [Loshchilov and Hutter, 2017]. We did a small grid search for learning rate on the cross-validation folds, but otherwise no hyperparameter tuning.

Although the smaller footprint of the AlphaEarth embeddings allows us to fit models on the full graphs in parallel, we found better performance when subsampling neighborhoods of the graph as described by Hamilton et al. [2017]. For each labeled node, a subgraph is created by randomly sampling four of it’s first degree neighbors and 16 of it’s second degree neighbors. This improvement is likely due to the added stochasticity of mini-batching.

All the models were trained on a single A100 GPU and they all converged within 15 minutes, with the majority of the computational time for the graph models being devoted to sampling subgraphs which is done on the CPU.

C Cross-Validation Folds

Table 2: Cross-validation splits for the different surveys.

Splits	Surveys
A	Kenya 2022 Standard DHS, Rwanda 2019-20 Standard DHS, Guinea 2018 Standard DHS, Ethiopia 2019 Interim DHS, Senegal 2018 Continuous DHS, Guinea 2021 MIS
B	Nigeria 2018 Standard DHS, Zambia 2018 Standard DHS, Cameroon 2018 Standard DHS, Liberia 2019-20 Standard DHS, Burkina Faso 2017-18 MIS, Mali 2021 MIS, Togo 2017 MIS
C	Mauritania 2019-21 Standard DHS, Sierra Leone 2019 Standard DHS, Burkina Faso 2021 Standard DHS, Mali 2018 Standard DHS, Uganda 2018-19 MIS, Senegal 2019 Continuous DHS, Niger 2021 MIS
D	Madagascar 2021 Standard DHS, Nigeria 2021 MIS, Benin 2017-18 Standard DHS, Cameroon 2022 MIS, Senegal 2017 Continuous DHS, Kenya 2020 MIS, Senegal 2020-21 MIS, Ghana 2019 MIS
E	Tanzania 2022 Standard DHS, Ghana 2022 Standard DHS, Cote d’Ivoire 2021 Standard DHS, Tanzania 2017 MIS, Gabon 2019-21 Standard DHS, Gambia 2019-20 Standard DHS, Mozambique 2018 MIS, Liberia 2022 MIS, Malawi 2017 MIS

Table 3: Cross-validation data splits for each fold

Train	Val	Test
CDE	B	A
ADE	C	B
ABE	D	C
ABC	E	D
BCD	A	E

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction clearly state the paper’s main contributions and limitations. Specifically, they claim (1) to propose a graph-based method using compact AlphaEarth embeddings for predicting cluster-level wealth across Sub-Saharan Africa, (2) to introduce a probabilistic “fuzzy label” loss for handling DHS coordinate displacement, and (3) to demonstrate modest improvements over image-only baselines on 37 DHS datasets. These claims are directly supported by the experiments and results reported in the paper. The abstract emphasizes that improvements are slight, and that the approach is computationally efficient rather than claiming major performance breakthroughs. Limitations such as the incomplete GeoNames coverage and the weak performance of fuzzy-label variants are also explicitly discussed, ensuring that the claims are proportionate to the presented evidence.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: The paper explicitly discusses several limitations in the Discussion and Outlook section, including the incomplete and biased coverage of GeoNames, the simplifying assumptions in the fuzzy-label weighting scheme, and the difficulty of learning under spatial uncertainty. It also notes the limited generalization of fuzzy-label models and highlights future directions to address these issues, such as population-weighted priors and better candidate location data.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.

- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not present formal theoretical results or proofs. It focuses on empirical modeling and evaluation rather than theoretical analysis.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides detailed information about datasets (37 DHS surveys and GeoNames settlements), preprocessing steps, model architectures, training procedures, and cross-validation splits. Hyperparameters, loss functions, and implementation details are listed in the appendix. We will also release the code on GitHub.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

- (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
- (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
- (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: We plan to release the full code and preprocessing scripts on GitHub. However, the DHS survey data are not fully open-access and currently require approved access requests through the DHS Program. Due to recent changes to the USAID department by the current American administration, it is unclear whether new requests are being processed. We will provide detailed instructions and dataset references so that authorized users can reproduce the experiments once access is granted.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specifies all key training and evaluation details, including optimizer (AdamW), number of epochs, early stopping, hidden dimensions, and loss functions. Full cross-validation splits, hyperparameters, and architecture definitions are provided in the appendix, ensuring the experimental setup is clear and reproducible.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report standard deviations for both MAE and R^2 across the cross-validation folds.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: A short discussion on computational needs can be found in the appendix. We also list the size of the AlphaEarth dataset in comparison to equivalent Sentinel-2 data in the Data section.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research uses publicly available, de-identified survey data in accordance with DHS data use agreements. The work aims to improve poverty estimation for socially beneficial purposes (targeting development interventions). We acknowledge data limitations and biases throughout the paper and do not make claims beyond what the evidence supports.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The work has potential positive impacts for targeting development interventions. Negative impacts discussed include risks from incomplete GeoNames data and spatial biases that could lead to misallocation of resources if predictions are inaccurate. Privacy is partially addressed through DHS coordinate displacement, though surveillance concerns from fine-grained wealth mapping should be acknowledged

Guidelines: The work has potential positive impacts for targeting development interventions. Negative impacts discussed include risks from incomplete GeoNames data and spatial biases that could lead to misallocation of resources if predictions are inaccurate. Privacy is partially addressed through DHS coordinate displacement, though surveillance concerns from fine-grained wealth mapping should be acknowledged

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not release new models or datasets with high risk for misuse. It uses existing public datasets (DHS surveys, GeoNames, AlphaEarth embeddings) and trains standard prediction models. The wealth prediction task does not pose dual-use or safety risks requiring safeguards.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [No]

Justification: While the paper cites the sources of all datasets (DHS, GeoNames, AlphaEarth), it does not explicitly mention the licenses or terms of use for these assets. The paper should add license information for DHS data usage terms, GeoNames Creative Commons license, and Google Earth Engine terms of service. This will be done for the camera-ready version.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The code as well as model weights will be published on GitHub. We will provide proper documentation.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.

- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper uses existing secondary survey data (DHS) collected by other organizations. No new crowdsourcing or human subjects research was conducted by the authors.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper uses publicly available, de-identified secondary data from DHS surveys. No new human subjects research was conducted requiring IRB approval. DHS data collection was approved by relevant IRBs in each survey country at the time of collection.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA] .

Justification:

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.