Prediction of demographic indicators from remote sensing images

PhD thesis

General information

- When to apply: until the 12th of May 2021
- Starting date: October 2021
- Funding: EDITE doctoral school **(subject to being approved by the EDITE committee.)**
- Institutes: Université de Paris, Laboratoire d'Informatique Paris Descartes (LIPADE), équipe Systèmes Intelligents de Perception and Institut national d'études démographiques, unité Démographie des populations du Sud (Demosud)
- Location: 45 rue des Saints-Pères, 75006 Paris (LIPADE) and 9, cours des Humanités, 93322 Aubervilliers (INED)
- Supervision: Sylvain Lobry, Camille Kurtz, Laurent Wendling - (first.lastname@u-paris.fr), Géraldine Duthé, Valérie Golaz (first.lastname@ined.fr)
- Keywords: Remote sensing, Demography, Computer vision, Deep learning, Africa

Proposed topic

Objective

In this PhD, which stems from and strenghtens an on-going collaboration between LIPADE and INED, the candidate will develop deep learning based methodologies using remote sensing data to predict indicators of the environment and environmental change, for demographic analysis. As such, the objective of this topic is twofold: to propose methodological contributions for the large-scale extraction of diachronic environmental indicators and to analyze their contribution to spatial population and health analyses. How do these indicators compare with the existing environmental data? What results do they yield in terms of the impact of environmental characteristics and environmental change on population structure and health in Sub-Saharan Africa? We expect prime results in the field of computer science (innovative methodologies) and demography (a better understanding of local inequalities in terms of population structure and health) as well as a contribution to the use of fine remote sensing data analysis for population studies.

Context

In a globalized context increasingly impacted by climate change, undergoing rapid population growth and urbanization, demographic studies would gain to better take environmental data into account and be carried out at the transnational level. However, this is not always possible in Sub-Saharan Africa, as matching harmonized demographic and environmental data are seldom available. One

major harmonized source of data on population and health in global South countries is the Demographic and Health Surveys (DHS) program. Since 2015, Demographic and Health Surveys were conducted once in about all the countries that participate in the program across Sub-Saharan Africa. To date, in spite of the delays in data collection and data set preparation due to the Covid-19 lockdowns, 16 data sets are already available for use, with matching geodata files. The large amount of spatial data regularly acquired since 2015 (in 2019 only, Sentinel satellites from the European Space Agency produced 7.54 PiB of open-access data[1]) are an opportunity to produce standardized and up-to-date indicators.

Background and state of the art

Several indicators have been developed to help understanding geographical realities in a consistent (i.e. not location dependent) manner. Among them, local climate zones (LCZ) have been proposed by WUDAPT (World Urban Database and Access Portal Tools) to systematically label urban areas [1]. Their goal is to provide a map of the world following this legend, in open-access, that can later be used by researchers for a wide range of studies. This data has been used to understand energy usage [2], climate [3] or geoscience modeling [4] or land consumption [5]. An important amount of work has been dedicated in the recent years to the automatic generation of such data, from sensors such as Landsat 8 or Sentinel 2. In a research competition organized by the IEEE IADF, several methods have been proposed to map LCZ from Landsat, Sentinel 2 and OpenStreetMap data [6]. Another recent study focused on the usage of Convolutional Neural Networks (CNNs) to tackle the task of automatically mapping LCZ using deep learning [7] and a large scale benchmark dataset was proposed in [8], with a baseline of an attention-based CNN. However, these works mostly focused on developed urban areas. For instance, the challenge of [6] targeted Berlin, Hong Kong, Paris, Rome, São Paulo, Amsterdam, Chicago, Madrid, and Xi'An. This is problematic, as developed cities are generally well mapped through governmental censuses, and that spatial generalization of machine learning based methods is a challenge [9]. It is therefore necessary to develop adapted methods for the global South [10].

In DHS surveys, a geospatial covariate dataset [11] corresponding to the approximate locations of the clusters interviewed can be matched to household, male and female datasets. The geospatial data stems from international programs aiming at providing estimates of environmental variables at the scale of planet Earth, such as population (Worldpop [12]), temperature and rainfall (CRU [13], Worldclim [14]), vegetation (VIP [15]), urbanisation (GHSL [16]), ... Most of these are based on large scale estimates derived from Landsat data and are defined over a 10km buffer zone in rural areas, 2 km in urban areas.

However, a smaller buffer around a precise location has been proven to bring out better results [17]. We therefore can expect that a high quality localised indicator such as LCZ in African urban metropolises would bring out better results when introduced in demographic analyses.

Desired background

We are looking for a student in Master 2 or engineering school in computer science data science or demography. The ideal candidate would have knowledge in image processing, computer vision, machine learning, Python programming, statistical data analysis and demographic research. The candidate should have an interest in large scale studies, remote sensing and demography.

[1] Sentinel Data Access 2019 annual report

How to apply?

You can apply by sending an email to stage-diip "at" listes.ined.fr before the 11th of May containing:

- A CV
- A cover letter
- Full set of transcripts
- Names, titles and emails of 2 persons who can write a reference letter. These persons should be expecting to be contacted during the application process.

Candidates will be selected and interviewed by the supervision team. The selected candidate will the be interviewed by the doctoral school (ED130) and a final answer will be provided at the beginning of July 2021.

Bibliography

[1] Benjamin Bechtel et al. "Mapping local climate zones for a worldwide database of the form and function of cities". In: *ISPRS International Journal of Geo-Information* 4.1 (2015), pp. 199–219.

[2] Paul John Alexander, Gerald Mills, and Rowan Fealy. "Using LCZ data to run an urban energy balance model". In: *Urban Climate* 13 (2015), pp. 14–37.

[3] Jan Geletič et al. "Spatial modelling of summer climate indices based on local climate zones: expected changes in the future climate of Brno, Czech Republic". In: *Climatic Change* 152.3-4 (2019), pp. 487–502.

[4] Hendrik Wouters et al. "The efficient urban canopy dependency parametrization (SURY) v1. 0 for atmospheric modelling: description and application with the COSMO-CLM model for a Belgian summer". In: *Geoscientific Model Development* 9.9 (2016), pp. 3027–3054.

[5] Jingliang Hu et al. "Land consumption in cities: A comparative study across the globe". In: *Cities* 113 (2021), p. 103163.

[6] Naoto Yokoya et al. "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.5 (2018), pp. 1363–1377.

[7] Chunping Qiu et al. "Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets". In: *Remote Sensing* 10.10 (2018), p. 1572.

[8] Xiao Xiang Zhu et al. "So2Sat LCZ42: A benchmark dataset for global local climate zones classification". In: *arXiv preprint arXiv:1912.12171* (2019).

[9] Emmanuel Maggiori et al. "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark". In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2017, pp. 3226–3229.

[10] John E Vargas-Muñoz et al. "Correcting rural building annotations in OpenStreetMap using convolutional neural networks". In: *ISPRS journal of photogrammetry and remote sensing* 147 (2019), pp. 283–293.

[11] Benjamin Mayala et al. "The DHS program geospatial covariate datasets manual". In: *Rockville: ICF* (2018).

[12] WorldPop. *Africa Continental Population Datasets (2000–2020)*. 2016. URL: https://www.worldpop.org/doi/10.5258/SOTON/WP00004.

[13] IPDJ Harris et al. "Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset". In: *International journal of climatology* 34.3 (2014), pp. 623–642.

[14] Stephen E Fick and Robert J Hijmans. "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas". In: *International journal of climatology* 37.12 (2017), pp. 4302–4315.

[15] K Didan and A Barreto. "NASA MEaSUREs vegetation index and phenology (VIP) vegetation indices monthly global 0.05 Deg CMG". In: *NASA EOSDIS Land Process. DAAC* 4 (2016).

[16] Martino Pesaresi et al. "GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014)". In: *European Commission, Joint Research Centre, JRC Data Catalogue* (2015).

[17] Kathryn Grace et al. "Integrating environmental context into DHS analysis while protecting participant confidentiality: A new remote sensing method". In: *Population and development review* 45.1 (2019), p. 197.