Earth Observation and Deep Learning for Sustainable Development

New Approaches To Facilitate Data-Informed Decision-Making

Fibæk, Casper Samsø

DOI (link to publication from Publisher):
10.54337/aau510571430

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

EARTH OBSERVATION AND DEEP LEARNING FOR SUSTAINABLE DEVELOPMENT

NEW APPROACHES TO FACILITATE
DATA-INFORMED DECISION-MAKING

BY
CASPER SAMSØ FIBÆK

DISSERTATION SUBMITTED 2022

AALBORG UNIVERSITY
DENMARK

EARTH OBSERVATION AND DEEP LEARNING FOR SUSTAINABLE DEVELOPMENT

NEW APPROACHES TO FACILITATE DATA-INFORMED DECISION-MAKING

by

Casper Samsø Fibæk

AALBORG UNIVERSITY

Submitted on the 21st of September 2022.

Dissertation submitted: September 21st, 2022

PhD supervisor: Professor Carsten Keßler
Department of Planning
Aalborg University Copenhagen, Denmark

PhD co-supervisor: Professor Jamal Jokar Arsanjani
Department of Planning
Aalborg University Copenhagen, Denmark

Industrial Supervisor: Head of Department Laurids Rolighed
Mapping and GIS
NIRAS A/S

PhD committee: Professor Ehsan Forootan (chairman)
Aalborg University, Denmark

Professor Sebastian van der Linden
University of Greifswald, Germany

Associate Professor Inge Sandholt
Technical University of Denmark, Denmark

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Planning

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-818-2

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2022

CV

Casper is an industrial PhD student at the engineering consultancy NIRAS and Aalborg University, working on developing tools and methods that utilise Earth Observation (EO) and Deep Learning to support the Sustainable Development Goals. The topic of the PhD came out of a personal interest in applying innovative technologies to concrete projects supporting global development. Casper is a part of the Mapping and GIS Department at NIRAS, where he has worked on diverse projects in numerous roles. He holds an AP. in Multimedia Design and Communication, a BSc. in Surveying, Planning and Land Management and an MSc. in Geoinformatics from Aalborg University in Copenhagen. His MSc. thesis was on the use of EO in monitoring compliance with the European Union's Common Agricultural Policy.

For most of the PhD, Casper has lived in Oxford, supporting the collaboration between NIRAS and Oxford Policy Management as the EO expert and industrial researcher on the "Savings at the Frontier" project. He currently serves as a quality expert on the EU Commission's Technical Assistance Facility for EO projects in Latin America as well as the EO and GIS expert on the $105 million Integrated Coastal Zone Management (ICZM) project for Egypt's Mediterranean coast. His previous work includes a feasibility study on new land management measures in Uganda - undertaken while working at the Ministry of Lands, Housing and Urban Development in Kampala, a study on using Sentinel imagery to detect invasive plant species in Denmark, as well as the creation of the WebGIS platform underpinning the coverage maps of the largest telecommunication service provider in Denmark.

Casper's research area is the application of EO and Deep Learning in service of the Sustainable Development Goals by facilitating data-informed geospatial decision-making. To that end, he has developed a software toolbox – Buteo and two distinct sets of Graphical User Interfaces to support the processes. He has published articles on estimating structural characteristics using satellite data, predicting population using data from the Copernicus Programme, forecasting urban growth using time series of satellite imagery and cellular automata, using geospatial analysis to support Financial Inclusion, and predicting the presence of protected dikes and walls using Deep Learning. He won the FIG Young Surveyor award, the CLGE Student Geoinformatics award, and the second-place award in the Integrated Geomatics track at the ESA/GSA Space week. He is a guest lecturer at multiple universities in Remote Sensing, has supervised master thesis projects and interns at NIRAS, and is the co-supervisor for an ongoing PhD project. Since July 2022, he has worked as a researcher at the European Space Agency's Φ-lab in Rome.

Intentionally left blank.

SUMMARY (ENGLISH)

Earth Observation (EO) is uniquely suited to support the Sustainable Development Goals (SDGs). It plays a central role in monitoring the Indicator Framework and directly contributes to several goals at both the Biosphere and Society levels. EO can support, track, and guide efforts towards reaching an estimated 1/4th of the targets and 1/8th of all the indicators. A key contribution of EO is in enabling global comparisons and coherent monitoring efforts. Central to the concept of Sustainable Development is recognising the finite resources of the globe, population growth, and unsustainable consumption patterns. This PhD project shows that these patterns can be monitored, measured, modelled, and predicted using EO data from sources such as the Copernicus Programme and Deep Learning (DL) methods.

The European Union's Copernicus programme alone generates approximately 16 terabytes of data per day, and analysing this large amount of data is a significant challenge. Artificial Intelligence and DL are uniquely suited to take on the challenge of analysing and making sense of these datasets. Recent advances in DL algorithms, the accessibility of cloud computing, and improved processing power mean that it is increasingly viable to undertake complex global monitoring tasks without expensive setups. In turn, it also means that small and medium-size enterprises can take advantage of these datasets to explore new and unique business cases. This PhD presents tools to assist Financial Service Providers, without subject matter experts, in setting up EO and DL analytical workflows using free and open software and data.

Despite the increased accessibility of EO data and data processing power, there are still major issues to tackle: There is a lack of easy-to-use monitoring interfaces and accessible endpoints for non-subject matter experts for information derived from satellite data. There is also a lack of labelled datasets and generic DL models tailored to ingest multi-sensor EO data. High quality, timely data with good geographical coverage is necessary to enable data-informed decision-making. Nevertheless, while countries in the Global South face major challenges, such as sustainable urban planning, less labelled data is available to inform decision-making. One way to alleviate this is by training DL models on labelled datasets from data-rich countries and carefully applying them in the Global South. Such efforts can lessen the data inequality, support the SDGs, and facilitate the creation of new business cases.

This PhD presents a series of methods and tools to create globally applicable DL models for predicting structural characteristics and population density while offering examples of how this data can be made accessible in easy-to-use decision support systems for non-EO experts. The models, tools, and datasets are available for anyone to use and further develop. They have been applied and are undergoing continuous development with partners in Denmark, Egypt, Ghana, and Tanzania to create business cases in the areas of Financial Inclusion and AgriTech.

RESUMÉ (DANSK)

Jordobservation (JO) er særligt velegnet til at understøtte verdensmålene for bæredygtig udvikling og bidrager direkte til flere af målene både på biosfære- og samfundsniveau. JO-data spiller ligeledes en central rolle i monitoreringen af resultatrammen for delmålenes indikatorer. Det anslås at JO kan understøtte, måle, og vejlede bestræbelserne på at opnå en 1/4 af målene og 1/8 af alle indikatorerne. JO datakilderne udgør et vigtigt bidrag i monitoreringsarbejdet, da de muliggør globale sammenligninger og en kontinuerlig, sammenhængende monitoreringsindsats. Et centralt element i begrebet bæredygtig udvikling er erkendelsen af klodens begrænsede ressourcer, befolkningstilvækst og uholdbare forbrugsmønstre. Dette ph.d.-projekt viser, at disse mønstre kan overvåges, måles, modelleres og forudsiges ved hjælp af JO-data og metoder som f.eks. Deep Learning (DL).

Copernicus-programmet genererer alene ca. 16 terabyte data om dagen, og det er en stor udfordring at analysere de store datamængder. Kunstig intelligens og DL teknikker er særdeles velegnede til at analysere og frembringe dybere indsigt fra disse datasæt. De seneste fremskridt inden for DL-algoritmer, tilgængeligheden af cloud-computing, og øget processeringskraft betyder, at det i stigende grad er muligt at udføre komplekse globale monitoreringsopgaver med lave omkostninger. Det betyder at små og mellemstore virksomheder kan drage fordel af disse datasæt til at undersøge nye forretningsmodeller. Denne ph.d. præsenterer værktøjer til at hjælpe finansielle tjenesteydere, som ikke har adgang til JO fageksperter, med at etablere JO- og DL-analytiske arbejdsgange ved hjælp af gratis og åben software og data.

På trods af den forøgede tilgængelighed af JO-data og databehandlingskapacitet er der stadig store problemer der skal løses: Der er mangel på brugervenlige grænseflader og tilgængelige produkter for ikke-fageksperter til at absorbere og anvende information afledt af satellitdata. Der er også mangel på træningsdatasæt og generiske DL-modeller, der er skræddersyet til multi-sensor JO-data. Aktuelle data af høj kvalitet med god geografisk dækning er nødvendige for at muliggøre datainformeret beslutningstagning. På trods af at landene i det globale syd står over for store udfordringer, med f.eks. bæredygtig byplanlægning, der kræver et solidt datagrundlag til at informere beslutningsprocesserne, er der ofte kun ganske få højkvalitets datasæt til rådighed. Adgangen til data i det globale syd kan styrkes ved at træne DL-modeller på træningsdatasæt fra datarige lande og forsigtigt anvende dem i det globale syd ved hjælp af Deep Transfer Learning. Udarbejdelsen og træning af sådanne DL-modeller kan mindske dataforskellene, understøtte verdensmålene for bæredygtig udvikling og lette udarbejdelsen af nye forretningsmodeller.

Denne ph.d. præsenterer en række metoder og værktøjer til at skabe globalt anvendelige DL-modeller for bygnings karakteristika og befolkningsskøn og giver samtidig eksempler på, hvordan disse data kan gøres tilgængelige i brugervenlige beslutningsstøttesystemer for ikke-JO-eksperter.

PREFACE AND ACKNOWLEDGEMENTS

The application for this PhD project was written during a deployment to Uganda for UN-Habitat and the International Federation of Surveyors (FIG). As part of a revision of the Ugandan land management policies, I investigated the use of Sentinel 2 data to help delineate property boundaries in rural communities. The experience guided the central topic of the PhD, and I have since explored ways of applying Earth Observation and Deep Learning to track and guide the Sustainable Development Goals (SDGs) – especially in the context of global development. The initial focus was on creating software and method modules that could be combined to create bespoke Spatial Decision Support Systems for SDG challenges. Over time, the focus shifted to concrete end-to-end tools and use-cases for actual business cases spurred by collaboration between NIRAS and partners. The shift meant increased efforts toward creating tools and datasets that were immediately useful for businesses and resulted in the project taking on an applied approach. The research has been inspired by the work and projects undertaken at NIRAS. I have strived to abide by open science principles: making datasets and code publicly available and using open-source software and datasets to ensure broad and global applicability of the research.

The Industrial PhD is funded by the Danish Innovation Fund, the NIRAS/Alectia Foundation and the Mapping and GIS department at NIRAS. I would like to sincerely thank all parties for funding the research and for the continuous encouragement and sincere interest in the work.

I want to thank my industrial supervisors at NIRAS, Laurids Rolighed and Søren Buch, for their guidance, experience, and support. Søren was the co-supervisor until his retirement in 2020. At Aalborg University, I would like to thank my academic supervisors, Carsten Keßler and Jamal Jokar Arsanjani, for sharing their knowledge and helping me guide the complex world of scientific publishing. I want to thank my collaborators: Hannah Laufer for introducing me to the world of FinTech, Marcia Trillo for her tireless contributions in testing the Buteo toolbox, Ana Fernandes for her support and contributions to the SatF tools, and Abednego Darko, Robert Madziva, Hany Ayad, and Foster Mensah for their collaboration and partnership.

Finally, I'd want to thank my partner Astrid for her loving support and for keeping me sane and Millie, our beloved old dog, who got me out of the house and were a great study companion.

Frascati, September 2022.

Casper Samsø Fibæk
MSc. Surveying, Planning, and Land Management

Intentionally left blank.

TABLE OF CONTENTS

1. Introduction ... 1
1.1 Earth Observation in Support of the Sustainable Development Goals ... 2
1.2 Deep Learning and Earth Observation ... 5
1.3 Research Questions ... 8

2. Summaries of Articles ... 9
2.1 Geodata-driven Approaches to Financial Inclusion ... 10
2.2 Estimating Structural Characteristics using Sentinel Data and Deep Learning ... 11
2.3 Creating Globally Applicable Population Estimates ... 12
2.4 Automatic Detection of Historic Stone Walls ... 13

3. Earth Observation and Deep Learning for Sustainable Development ... 15
3.1 Data and Decision-Making ... 16
3.2 Financial Inclusion ... 19
3.3 Integrated Coastal Zone Management ... 25
3.4 Cultural Heritage - Protected walls and dikes ... 27
3.5 Biodiversity, Planning, and Healthcare ... 29

4. Methods and Tools ... 33
4.1 Structural Characteristics ... 34
4.2 Estimating Population (SenPop) ... 38
4.3 Transfer Learning ... 41
4.4 The Buteo Toolbox ... 42
4.5 Easy Toolbox GUI ... 45
4.6 S2Super ... 46
4.7 The Savings at the Frontier Interfaces ... 47

5. Discussion ... 51

6. Conclusion ... 55

7. Bibliography ... 57

TABLE OF FIGURES

Figure 1. The contributions of different data types to the SDGs............................... 2
Figure 2. The contributions of EO derived data types to the SDGs 3
Figure 3. Examples of Deep Learning techniques 6
Figure 4. The contribution of the articles to the research questions........................... 8
Figure 5. Cancun in 1970... 16
Figure 6. Savings the Frontier workshops in Dar es Salaam and Accra 21
Figure 7. Formalising SUSU collection 22
Figure 8. Collaboratory development of the Spatial Decision Support System 24
Figure 9. Flooding in Alexandria, Egypt.. 25
Figure 10. Flooding scenario and impacted structures... 26
Figure 11. Descriptions from the official guide to protected stone walls and dikes. 27
Figure 12. Examples of labels and predictions of stone walls................................. 28
Figure 13. A WebGIS interface to manage invasive plant species. 29
Figure 14. A WebGIS interface to predict unmanaged urbanisation 30
Figure 15. Expected market penetration in Ghana 31
Figure 16. Processing Sentinel 1 data to proxy "permanent structures" 34
Figure 17. The effects of orbital direction on corner reflectors............................... 35
Figure 18. Classification of the structural volume in Silkeborg, Denmark............. 36
Figure 19. Predicted structures used for Integrated Coastal Zone Management...... 37
Figure 20. Comparison between WorldPop, MetaPop, and SenPop 38
Figure 21. Daytime and nighttime population in Tamale, Northern Ghana............. 39
Figure 22. SenPop time-dependent population predictions with default weights 40
Figure 23. Prepared training data for structures, Denmark. 41
Figure 24. Fractional ellipsoidal kernels are available in the Buteo Toolbox 42
Figure 25. A subset of the functions that are available in the Buteo Toolbox. 43
Figure 26. Using the mad collapse 44
Figure 27. GUI created using Easy Toolbox GUI from a simple setup file 45
Figure 28. Super-sampled Sentinel 2 bands over Copenhagen during winter.......... 46
Figure 29. SATF-Proximity tool - Application architecture diagram 47
Figure 30. The latest version of the Savings at the Frontier Excel interface............ 48

PHD PUBLICATIONS

Fibæk, Casper Samsø, Hanna Laufer, Carsten Keßler, and Jamal Jokar Arsanjani. "**Geodata-Driven Approaches to Financial Inclusion – Addressing the Challenge of Proximity.**" International Journal of Applied Earth Observation and Geoinformation 99 (2021): 102325. Regular Issue

Fibæk, Casper Samsø., Carsten Keßler, and Jamal Jokar Arsanjani. "**A Multi-Sensor Approach for Characterising Human-Made Structures by Estimating Area, Volume and Population Based on Sentinel Data and Deep Learning.**" International Journal of Applied Earth Observation and Geoinformation 105 (2021): 102628. Special Issue on Geospatial Artificial Intelligence.

Fibæk, Casper Samsø, Carsten Keßler, Jamal Jokar Arsanjani, and Marcia Luz Trillo. "**Creating Globally Applicable Population Estimates from Sentinel Data.**" Transactions in GIS (2022): https://doi.org/10.1111/tgis.12971 Special Issue on Methods, Models, and Resources for Geospatial Knowledge Graphs and GeoAI.

Trotter, Ezra, Ana Fernandes, Casper Samsø Fibæk, and Carsten Keßler. "**Machine Learning for Automatic Detection of Historic Stone Walls Using LiDAR Data.**" International Journal of Remote Sensing (2022): 101080. Regular Issue

PHD RELATED PUBLICATIONS

Mahboubi, Masoume, Astrid Grønbæk, Ana Fernandes, Casper Fibæk and Paolo Dabove. **"Geospatial Analysis of Safe Delivery App Events Based on the Geographically Weighted Regression Tool"**. Proceedings of the 2022 Agile Conference, 3(47), 1–6. https://doi.org/https://doi.org/10.5194/agile-giss-3-47-2022

Danish Agency for Data Supply and Efficiency. **Danish Uses of Copernicus – 50 User Stories Based on Earth Observation**. Casper Fibæk contributed three user stories: Supporting Financial Inclusion, Classification of Invasive Plant Species, and Satellite-Based Detection of Changes of Buildings (2021). ISBN 978-87-94056-03-8

Fibæk, Casper Samsø, and Simon Kamp Danielsen. "**Classification of Invasive Plant Species Using Sentinel Satellite Data and Other Free Datasets.**" Danish: "Klassifikation af Invasive Plantearter ved Hjælp af Sentinel Satellitdata og andre Frie Data." Ministry of the Environment. Project Nr.: 10401119 Publication Nr.: 1231391948 (2019)

Arsanjani, Jamal Jokar, Casper Samsø Fibæk, and Eric Vaz. "**Development of a Cellular Automata Model Using Open-Source Technologies for Monitoring Urbanisation in the Global South: The Case of Maputo, Mozambique.**" Habitat International 71 (2018): 38-48. Regular Issue.

DISSEMINATION

- Guest Lecturer on Earth Observation for Sustainable Development at the Centre for Remote Sensing and Geographic Information Services
  The University of Ghana – Legon, Ghana
- Guest Lecturer on the use of Remote Sensing for Sustainable Agriculture
  The Royal Agricultural School – Cirencester, United Kingdom
- Guest Lecturer on Earth Observation and the Sustainable Development Goals
  Aalborg University, Copenhagen
- Guest Lecturer on Deep Learning for Geospatial Analysis
  Aalborg University, Copenhagen
- Three-day workshop on the use of Geospatial Technology for Financial Inclusion
  Dar es Salaam, Tanzania
- Three-day workshop on the use of the Earth Observation and Deep Learning for Financial Inclusion and Agriculture in *Accra, Ghana*
- Two-day online workshop on Business-Cases of Earth Observation and Geospatial Analysis for Financial Service Providers
- Webinar: "Mapping it Out: Practical Tools for Reaching Remote Clients" for CGAP/World Bank Group
- Numerous presentations at NIRAS, NIRAS International and Oxford Policy Management on Earth Observation and Deep Learning in Support of the Sustainable Development Goals
- Three presentations at the annual mapping conference in Denmark: "Kortdage"
- Member of the organising committee of Kortdage for two years and organiser of the "Space"-tracks
- Participated in the ECMWF Conference on Artificial Intelligence and Space Data in *Reading, United Kingdom*
- Participated in UK Space Agency's "Space for Development" conference in *London*
- Supervisor for one PhD student, four master's students, one bachelor's student, and five interns at NIRAS.
- Invited speaker for the Danish Copernicus Committee on NIRAS use-cases for Copernicus data and the Aalborg University GIS day
- Presentation for the Brazilian Delegation to Denmark on "The Digital Transformation and Modernization in Brazil" on Remote Sensing for Forestry and Financial Inclusion. Organised by the Danish Ministry of Foreign Affairs and the Danish Technical University (DTU)

SOFTWARE AND DATA

SOFTWARE

The BUTEO Toolbox *Casper Fibæk and Marcia Trillo (Testing)*
https://github.com/casperfibaek/buteo
A toolbox to facilitate the use of EO data in Geospatial Decision Support Systems by significantly easing the processing of Earth Observation data for Deep Learning. Tools include: Downloading, preprocessing, postprocessing, tiling, and easy-to-use prediction workflows for satellite data. Includes an example catalogue. EO Specific tools are contained within the addon buteo_eo.

Sentinel Super-Resolution (S2Super) *Casper Fibæk*
https://github.com/casperfibaek/super_res_s2
A Deep Learning model to super-sample Sentinel 2 data, envisioned as part of a preprocessing pipeline to improve model accuracy. Trained on 3000 scenes across the entire globe, with samples from all climate zones and all major cities.

Savings at the Frontier (API) *Casper Fibæk and Ana Fernandes*
https://github.com/casperfibaek/satf-api
The API is an interface to translate requests from the Excel Interface Single Page Application (SPA) to spatial queries handled by a PostGIS spatial database underpinning the Savings at the Frontier project. It can be used and accessed without going through the front end. The database code and its proprietary functions are not available due to licensing.

Savings at the Frontier (SPA GUI) *Casper Fibæk and Ezra Trotter*
https://github.com/casperfibaek/satf-frontend
The Excel front-end and Single Page Application (SPA) is used to generate geospatial queries for a PostGIS database. It is designed to integrate into the day-to-day workflows of Financial Service Providers in Ghana and Tanzania. It contains many geospatial functions and a LeafletJS interface to georeference and visualise results. It interacts with Microsoft Excel through OfficeJS.

Easy Toolbox GUI Creator *Casper Fibæk*
https://github.com/casperfibaek/toolbox-gui
A small configuration file base tool to easily create bespoke Graphical User Interface toolboxes akin to the MAPLA interface of the Orfeo toolbox. It is designed to make it easy to create GUIs for the tools within the Buteo-toolbox, but the creator is target agnostic.

DATASETS

Fibæk, Casper Samsø (2021), "**Raster Dataset on the Area, Volume and Population of Structures in Denmark**.",
Mendeley Data, V1, DOI: 10.17632/mpkdmfdb8m.1

Fibaek, Casper Samsø (2022), "**Area and Population Estimates from Sentinel Data - Ghana and Mediterranean Egypt**.",
Mendeley Data, V1, DOI: 10.17632/gf8v525tm6.1

Fibaek, Casper Samsø (2022), "**Subset Containing 300 out of 3000 Prepared Sentinel 2 Scenes for Transfer Learning and Super-sampling.**"
Zenodo Open Data, V1, DOI: 10.5281/zenodo.7099984

All training and label datasets are available upon request.

1. INTRODUCTION

Earth Observation (EO) is more important than ever – it provides quantitative data on the state of a world undergoing major changes: from climate change, depleting resources, and urbanisation to disaster management and even *war*. It is a global, unbiased, and coherent data resource that is increasingly critical for addressing the world's biggest challenges while supporting individuals such as farmers, urban planners, and Financial Service Providers. It is truly a Big Data challenge that, when powered by modern breakthroughs in Deep Learning (DL), can provide global analysis, support sustainable development, and drive data-informed decision-making (Andries et al., 2019; Paganini et al., 2018).

The UN Reports on the Data Revolution and the Indicator Framework state that data is the enabler of the 2030 Agenda for Sustainable Development and "*[...] the lifeblood of decision-making and the raw material for accountability*" (United Nations, 2014; United Nations Statistical Commission, 2020). Free, global, high-frequency, and high-quality EO data are key to supporting data-informed decision-making and sustainable development. The data makes the monitoring framework for the Sustainable Development Goals (SDGs) technically feasible and financially viable (Anderson et al., 2017; Ferreira et al., 2020).

Open data and free and open-source software play a significant role in supporting sustainable development and are increasingly a part of public administration policy (French Presidency of the Council of the European Union, 2022). EO data sources have long been publicly available. However, the advent of DL, Cloud Computing, and the commercialisation of space has enabled global analysis of these data sources for consumers without requiring access to high-performance computing (Herweijer et al., 2018; Lord et al., 2021). Crowdsourced data, such as OpenStreetMap, has made high-quality vector data globally accessible, and when combined with open data from EO programmes such as Copernicus, it enables innovative new solutions to support sustainable development (Amadi et al., 2019; Schultz et al., 2017).

In a series of publications, this PhD dissertation presents novel approaches to analysing EO data using DL to support data-informed decision-making, especially in the Global South. The approaches significantly improve upon the state-of-the-art for estimating population (Fibæk et al., 2022) and structural characteristics (Fibæk, Keßler, et al., 2021) and provides interfaces to make the datasets accessible to decision-makers (Fibæk, Laufer, et al., 2021) without requiring EO or DL expert intermediaries. The work builds exclusively on open data and software. All derived datasets, code, tools, and articles are publicly available for free. They have been tested and implemented in collaboration with local partners.

The PhD project is an industrial "PhD by Publication". This document serves as the conceptual frame around the portfolio of academic work completed.

1.1 EARTH OBSERVATION IN SUPPORT OF THE SUSTAINABLE DEVELOPMENT GOALS

Earth Observation (EO) plays a key role in supporting sustainable development. It is used to track and guide the progress towards achieving the goals, aspirational targets and the indicators specified in the Global Indicator Framework (Anderson et al., 2017; Paganini et al., 2018). EO enables coherent and global monitoring while also contributing at the sub-national and business levels by efforts such as facilitating smart agriculture or informing urban planning. Figure 1 is derived from the report "Satellite Earth Observations in Support of the Sustainable Development Goals", wherein Paganini et al. argue that approximately 1/4th of the aspirational targets and 1/8th of indicators in the Indicator Framework can be contributed to or be directly measured using EO data (United Nations Statistical Commission, 2020).

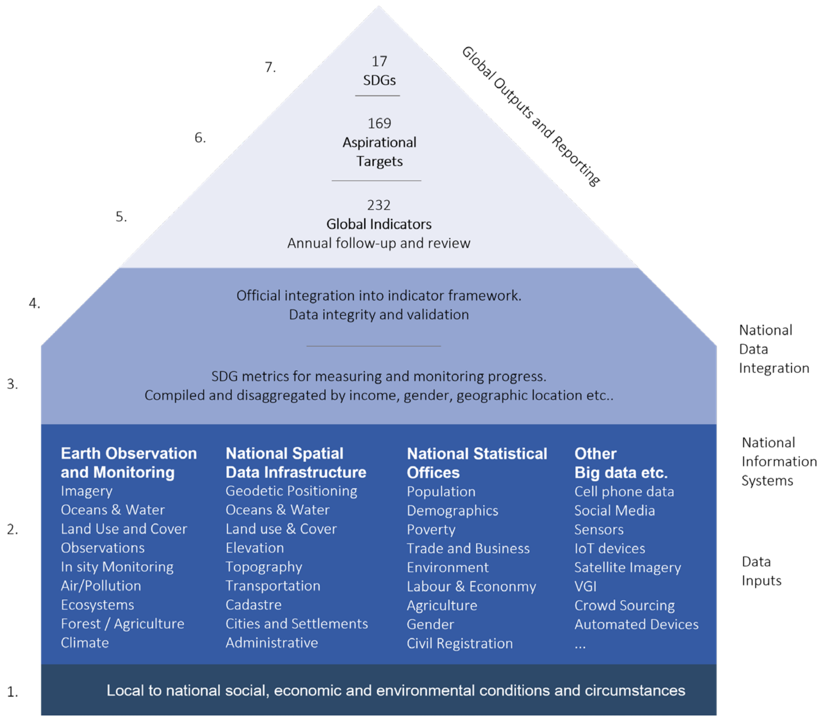

Figure 1. The contributions of different data types to the SDGs. The figure is derived from Paganini et al., 2018 (p. 10).

In GEO, 2017, the authors investigate EO use-cases that support the SDGs and produce a list of EO derived datasets that can support specific goals. In the report, the two EO datatypes that are highlighted as contributing to the broadest range of the Sustainable Development Goals (SDGs) are "Population Distribution" and "Cities and Infrastructure Mapping" (see Figure 2). Population distribution can be inferred from structural characteristics, which in turn is a central mapping target for cities and infrastructure mapping (Fibæk et al., 2022; Schug et al., 2021; Tiecke et al., 2017). Improving the state-of-the-art of mapping structures and their characteristics to derive population estimates is a theme for all but one of the core published articles that are part of this PhD.

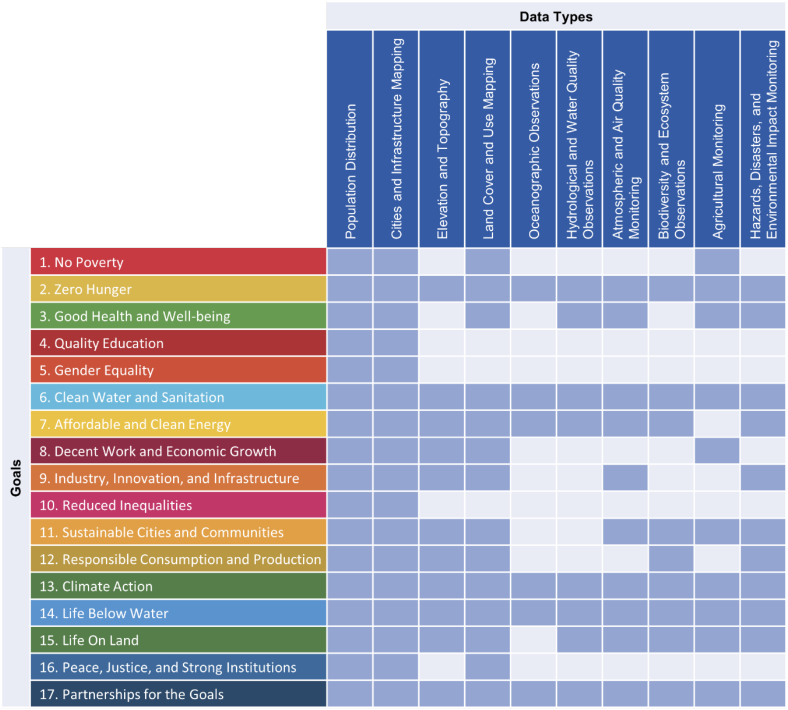

Figure 2. The contributions of EO derived data types to the SDGs. Based on the figure in GEO, 2017 *(p. 3).*

While the most straightforward applications of EO are for the biosphere orientated SDGs, Fibæk, Laufer, et al., 2021 show methods of applying EO to the society oriented goals *End Poverty* (SDG 1) and *Achieve Gender Equality* (SDG 5) through Financial Inclusion. Financial Inclusion concerns the equal access to fit-for-purpose financial products such as savings and transaction accounts, insurance, and loans. It is a theme for reaching several SDGs and is not mentioned in the Paganini report. Eight of the seventeen goals feature elements of Financial Inclusion (Fibæk, Keßler, et al., 2021). Goal one specifically targets equal rights to economic resources and access to financial services (Demirguc-Kunt et al., 2018).

Financial Service Providers (FSPs) use data to inform the location and services offered to customers while providing valuable insights into market saturation, supply-chain management, and customer segmentation. EO can support Financial Inclusion by providing up-to-date data on urban structures, infrastructure, population, wealth, and economic activity. Combining the temporal and spatial resolution of EO with high-quality local geospatial data makes it possible to assist FSPs in making data-informed decisions in previously insufficiently mapped areas (Fibæk, Keßler, et al., 2021). FSPs in Ghana, Zambia and Tanzania have been relying on official statistical data collected at a temporal and spatial resolution that is generally insufficient for day-to-day decision-making (Fibæk, Laufer, et al., 2021; J. E. Steele et al., 2017). The data insufficiency is partially caused by a reliance on housing and population census data, which is conventionally conducted decennially (Shi et al., 2018; United Nations, 2008). In some countries, national censuses are not conducted at all[1].

The European Union's Earth Observation Programme – Copernicus, is a central data source for many sustainable development efforts (Jutz & Milagro-Pérez, 2020). Several satellite constellations are part of the programme, including the radar constellation Sentinel 1, which can provide surface measurements in all-weather scenarios and Sentinel 2, which offers multispectral instruments capable of capturing a wide range of the electromagnetic spectrum. The wide range is helpful in tasks such as crop type classification, land-use and land-cover classifications, and human settlement classifications (Drusch et al., 2012; European Space Agency, 2015). The data products of the NASA Landsat satellites are similar to the Sentinel 2 satellites. Both datasets can be used in conjunction to increase the temporal resolution of the satellites (Wulder et al., 2019).

Paganini et al., 2018 list three key benefits of EO for sustainable development:

1. Makes the Global Indicator Framework viable.
2. Improves the temporal resolution of statistical outputs, such as census data.
3. Improves accuracy of reporting targets by being spatially explicit.

In addition to these three benefits, two additional benefits will be reasoned upon in this dissertation:

4. Enables data-informed decision-making in areas where it was previously not possible (Fibæk, Laufer, et al., 2021).
5. Deep Transfer Learning from data-rich to data-poor regions can lessen the impact of data inequality (Fibæk et al., 2022).

While the presented work can be used to support the Indicator Framework, the work focuses on supporting the goals themselves instead of developing new ways of measuring specific indicators.

[1] In Lebanon, to calm sectarian tension, no official census has been conducted since the 1932 census during the French Mandate, which played a significant role in the founding of the Lebanese post-colonial republic and the power sharing agreements (Mikdashi, 2020).

1.2 DEEP LEARNING AND EARTH OBSERVATION

Deep Learning (DL) methods are well suited for analysing Earth Observation (EO) data. The methods can outperform humans in image classification tasks and detect patterns and changes with previously unattainable accuracy. A key benefit to using DL for EO is how its accuracy generally scales with the amount of input data. As EO provides massive amounts of data from multiple sensors, switching from more traditional Machine Learning approaches to DL can significantly boost accuracy. As a methodology, it is highly flexible, able to simultaneously ingest images from varying sensors and generate outputs of arbitrary size. The flexibility makes it possible to create multi-sensor approaches and enables the creation of synthetic and up-scaled images. Working with DL approaches often mean combating overfitting. It is mainly an issue if little training data is available, and it is vital to take appropriate precautions – keeping in mind that DL models are biased by the intentions of their creators through the model design and the geography and semantics of the training data used (Chollet, 2022).

DL is a subfield of Machine Learning, which in turn is a sub-field of Artificial Intelligence. Artificial Intelligence encompasses the modelling of intelligent systems through rules, traditionally conditional statements. Machine Learning is explicitly about models that can learn to optimise and recognise specific patterns to perform tasks. Usually, Machine Learning entails using algorithms such as Linear Regression, Decision Trees, and Support Vector Machines. For DL, the basic design is an artificial neural network that seeks to mimic how the human brain works through networks of interconnected modelled neurons. The network is trained by fitting a loss function through backpropagation to the weights assigned to each "neuron" (Campesato, 2020).

DL is not new, but recent advances in Information and Communication Technologies have made it feasible to collect, process, and analyse massive amounts of data and distil the information (Buduma & Locascio, 2017). The key advances have come in the form of improved algorithms, cloud computing, and increased processing power. EO as a data source for up-to-date decision-making, business analysis and global monitoring has become increasingly viable due to these advances in AI (Herweijer et al., 2018). DL is key to unlocking the potential of all these new data streams, and the field and application areas are predicted to grow considerably in the coming years (European Space Agency, 2018).

There are many different methods for designing DL models to solve distinct EO tasks. The most common is image classification, where an EO image is divided into predefined semantic classes. Multiple techniques will often be used together to conduct an analysis, such as super-sampling the Sentinel 2 20m and 60m bands before performing dimension reduction and then doing a land cover classification. Another method could be using a DL model to reduce speckle noise in Sentinel 1 SAR images before performing change detection (Lillesand et al., 2016). Figure 3 shows some examples of techniques that can be used and studies that have used them.

Technique	Description	Examples	Litterature
Classification	Separating input data into predefined classes. Can be binary.	• Land cover & use • Crop types • Structure types	Fibæk, Laufer, et al., 2021 Arsanjani & Fibæk, 2018
Regression	Predicting a value for the input data. Can be constrained to a range.	• Land Value • Biomass • Structural Area	Fibæk et al., 2022 Fibæk, Keßler, et al., 2021
Clustering & Segmentation	Separating input data into areas with similar features.	• Twin neighbourhoods • Classification preprocess	Fibæk, Laufer, et al., 2021
Object-based Classification	Classifying segmented input data. Can be done using Zonal Statistics.	• Crop types • Forest Canopies	Fibæk, 2017
Gap-filling & Noise Reduction	Filling in missing parts of images or reducing noise in a dataset.	• SAR Noise Reduction • Cloud removal	Cresson, 2020
Dimension Reduction (Encoding)	Reducing the size of the input data while preserving the input variance.	• Preprocessing • Compress data	Vali et al., 2020
Super sampling	Increasing the resolution of input data.	• Sentinel 20m -> 10m	Lanaras et al., 2018
Synthetic Data Generation	From input data such as text, generate a new image.	• Multiseason training data	Wu et al., 2018

Figure 3. Examples of Deep Learning techniques and how they can be used to analyse Earth Observation data.

Many DL models created for EO analysis are inspired or derived from models trained on images taken by standard handheld cameras. Therefore, they are often trained solely on RGB inputs, even though many EO satellites, such as the Sentinel 2 and Landsat 8/9 satellites, capture a wider range of the electromagnetic spectrum (Cresson, 2020). Furthermore, due to the spatial aspect of EO, it is often possible to capture the exact location using multiple sensors within a short period of time, and these multi-sensor inputs are also seldom used in publicly available models. Some companies and NGOs are working on alleviating this through efforts such as Radiant Hub and the European Space Agency sponsored project: AI4EO[2].

Considerable efforts have been undertaken to map human settlements (Corbane et al., 2021) and their composition, structure, and population (Esch et al., 2017; Koppel et al., 2017) from DL and EO (Fibæk, Laufer, et al., 2021). However, given the unprecedented amount of EO data, the increased processing power available and advances in DL methods, there is a timely need to develop novel approaches for improving the mapping of these settlements at higher spatial and temporal resolution from multi-sensor approaches (Fibæk, Keßler, et al., 2021). Information on Population Distribution and Cities and Infrastructure Mapping is key to answering questions related to population growth, urbanisation, pollution, disaster management, spatial planning (see Figure 2), and even supporting Financial Service Providers in peri-urban

[2] https://ai4eo.eu and https://mlhub.earth

and rural areas by enabling data-informed decision-making about Financial Inclusion and market saturation (Fibæk, Laufer, et al., 2021; Haberl et al., 2021).

Due to the inherent geographic and semantic bias in curated training data, DL models usually require local training data to sufficiently generalise. Fortunately, a key benefit of using DL approaches is Transfer Learning, which can significantly reduce the local training data required. Using Transfer Learning, a model trained to solve one task is reused as the starting point of another model – usually by adding layers to the rear of the original model or refitting parts of the original model. The classic example is using a model trained on the tagged images in the ImageNet[3] database to classify pictures of dogs and cats (Chollet, 2022). These approaches also apply to EO data, where a model trained to predict the footprint of buildings can be used to classify structure types, or models trained in regions with ample training data can form the basis of models applied to data-poor regions. Urban planning and disaster management issues are felt dramatically in the developing world, yet there is a lack of data to facilitate data-informed decision-making (Fibæk, Laufer, et al., 2021). Training DL models in areas where significant ground truth data is available and adapting them to different contexts can reduce the information inequalities and improve decision-making. Sentinel 1 and 2 data is well suited for far-field applications as the data is global, open, and freely available (Fibæk et al., 2022).

Structural characteristics form the fundamental units of many geospatial analyses of human settlements. It can enable the classification of neighbourhood types, socio-economic status, and population estimates and can be used to assess topics such as energy needs, pollution, infrastructure requirements, urban and disaster planning, and health care coverage. Using open EO data to create Deep Learning models that can discern and map the characteristics of human-made structures can lower costs and barriers to make it feasible to continuously monitor structure and population patterns (Corbane et al., 2021). In cases where existing high-quality data on structures is already available, an EO-based model can increase the temporal resolution. It can also enable targeted topographic updates, thereby making large-scale, high-resolution mapping more affordable (Fibæk, Keßler, et al., 2021; Haberl et al., 2021).

The studies presented in this dissertation focus on designing multi-sensor, multi-resolution DL models for estimating structural characteristics such as area, volume, structure type, and population using open data and software. The focus ensures that the research can support a broad range of the Sustainable Development Goals, such as Financial Inclusion (*SDG 1: End poverty*) and Integrated Coastal Zone Management (*SDG 14: Life below water and SDG 15: Life on Land)*.

[3] ImageNet is a database of over 14 million hand-annotated images with bounding boxes and 20.000 classes. The database is used as the basis of many Deep Learning model and the focus of the ImageNet Large Scale Image Recognition Challenge.

1.3 RESEARCH QUESTIONS

The object of this PhD project is to investigate the use of Earth Observation (EO) data and Deep Learning (DL) methods within the context of sustainable development and how to equip decision-makers with the appropriate knowledge and tools to process and interpret EO data and DL methods to ultimately make data-informed decisions, especially in areas where there is currently little to no data available. The project presents a design for an EO-driven Graphical User Interface and datasets to enable data-informed decision-making in Ghana and Egypt and a series of tools based on open-source software and data to ease analysing EO data using DL. The interfaces support businesses without access to EO and DL experts in consuming and integrating EO data directly into their decision-making workflows.

How can Earth Observation and Deep Learning be used to improve data-informed decision-making on SDG related topics?

1. How can a toolbox be designed to ease Earth Observation data processing using Deep Learning?
2. How can Earth Observation and Deep Learning analytics be brought into decision-makers hands without access to subject-matter experts?
3. How can labelled data from data-rich countries such as Denmark be used to improve population and structure estimates in the Global South?
4. How can the state-of-the-art population and structure estimates, which form the basis of many SDG monitoring efforts, be improved?

Figure 4 shows the articles that contribute to the research questions. Section two contains summaries of the articles, while the conclusion section describes how the articles have answered the specific research questions.

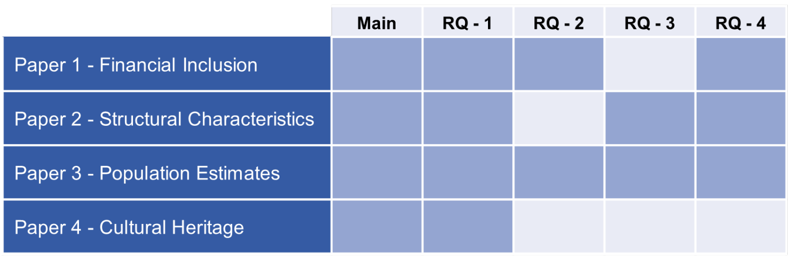

Figure 4. The contribution of the articles to the research questions.

The following section contains summaries of the core articles of the PhD. Afterwards, a section on “Earth Observation and Deep Learning for Sustainable Development” describes the broader context and topics where the research conducted throughout this PhD has benefitted sustainable development. Section four presents the tools and methods used and developed, while sections five and six are the discussion and conclusion.

2. SUMMARIES OF ARTICLES

This section consists of summaries of the four core articles of this PhD. The full articles are submitted for consideration together with the dissertation itself. The dissertation contains three articles where Casper is the primary author and one where he is a co-author. All articles have been peer-reviewed and accepted in journals.

2.1 GEODATA-DRIVEN APPROACHES TO FINANCIAL INCLUSION

Fibæk, Casper Samsø, Hanna Laufer, Carsten Keßler, and Jamal Jokar Arsanjani. "**Geodata-Driven Approaches to Financial Inclusion – Addressing the Challenge of Proximity.**" International Journal of Applied Earth Observation and Geoinformation Volume 99, July 2021, 102325. Open Access **https://doi.org/10.1016/j.jag.2021.102325**.

"*Financial Inclusion is, in many ways, a spatial planning issue: Where do financial institutions provide services, how far do customers travel to access mobile money, which services are available where and how is agent cash-flow handled?*" - Fibæk, Laufer, et al., 2021.

This article investigates the use of geodata, Earth Observation and Deep Learning to make Financial Inclusion (FI) an addressable challenge. FI is a central topic to achieving several Sustainable Development Goals (SDGs), although mainly goal one: *End Poverty*. It is about ensuring equal access to fit for purpose financial services, savings mechanisms, and financial products. Access to financial services is particularly an issue in under-served communities, such as low-income housing, self-organised, and rural areas (Varghese & Viswanathan, 2018). A key finding in Forster et al., 2013 and Peachey & Mutiso, 2019 is that physical and social proximity to financial services is central to improving FI and stresses that appropriate methods for measuring social and physical distances are important to reduce exclusion actively.

Earth Observation, Deep Learning, and a Decision Support System for Financial Service Providers are presented to address the challenge of measuring proximity in Ghana. It is possible to create an interactive map of Financial Inclusion by classifying different types of settlements, calculating population density, and georeferencing: mobile money agents, marketplaces, informal savings groups (SUSUs), and bank branches. However, mapping inclusion/exclusion is only the first step towards improved inclusion. It is necessary to enable the creation of business cases for private actors to reach the under-served communities (Demirguc-Kunt et al., 2018; Peachey & Mutiso, 2019).

The interactive Spatial Decision Support System presented support FI in two ways. (1) It allows Non-Governmental Organisations and National Statistical Offices to measure progress toward SDG 1, and (2) it enables the exploration of new business cases. The system allows the comparison of villages by accessibility, population, demography, and type to investigate the optimal placement of mobile money agents.

This article was written together with Hanna Laufer, a Financial Inclusion expert at Oxford Policy Management. The article was picked up and discussed in the blog post: https://blog.mondato.com/beyond-the-last-mile-reconsidering-rurality/

2.2 ESTIMATING STRUCTURAL CHARACTERISTICS USING SENTINEL DATA AND DEEP LEARNING

Fibæk, Casper Samsø, Carsten Keßler, and Jamal Jokar Arsanjani. "**A Multi-Sensor Approach for Characterising Human-Made Structures by Estimating Area, Volume and Population Based on Sentinel Data and Deep Learning.**" International Journal of Applied Earth Observation and Geoinformation. Volume 105, 25 December 2021, 102628. Special Issue on Geospatial Artificial Intelligence. Open Access. **https://doi.org/10.1016/j.jag.2021.102628**

A key takeaway from developing the datasets and Spatial Decision Support System for Financial Inclusion was that population density, neighbourhood types, and connectivity were the central datasets driving decision-making. While the previous article relied heavily on data from the WorldPop project (Tatem, 2017), it proved necessary to improve the resolution and ensure that data could be continuously updated. Haberl et al., 2021 and Koppel et al., 2017 show that it was possible to estimate building heights using the Sentinel 1 satellite. This article aims to improve the methodology of predicting structural characteristics by using a Deep Learning multi-sensor approach and comparing the importance of different input datasets.

Numerous studies (Li et al., 2019; Microsoft, 2021; Sirko et al., 2021) have shown that it is possible to extract vectorised building footprints from very high-resolution imagery, but there was a lack of research on measuring the physical characteristics of structures using open data policy satellites. Corbane et al., 2020 and others have made binary classifications of human settlements using Sentinel or Landsat data. However, more information on structures is needed to create neighbourhood-level classification schemes and a high-quality conversion layer from structural characteristics to population estimates.

The case study area of the article is Denmark, which, while largely a geographically homogeneous area, was chosen due to the accessibility of data that enabled the creation of labels for the area, volume, and population of structures at the resolution of the Sentinel satellites, thus serving as an "ideal" testbed. Different combinations of data products from the Sentinel 1 and 2 data were investigated: What is the effect of only using RGB, and does combining Sentinel 1 and 2 improve the results? What is the influence on the accuracy of the results from using one or both orbital directions?

The best combination of data included all the spectral bands from the Sentinel 2 satellite, both orbital directions from the Sentinel 1 Satellites and utilised the complex data products from Sentinel 1 to produce interferometric coherence estimates. The improvements in accuracy from using the complex products were minor, while using backscatter from both orbital directions significantly improved results. Finally, the study led to the design of a Deep Learning model that provides good estimates for the characteristics, decreasing in accuracy as the complexity of the target metric increase.

2.3 CREATING GLOBALLY APPLICABLE POPULATION ESTIMATES

Fibæk, Casper Samsø, Carsten Keßler, Jamal Jokar Arsanjani and Marcia Trillo. "**A Deep Learning Method for Creating Globally Applicable Population Estimates from Sentinel Data**" Transactions in GIS – Special Issue on Methods, Models, and Resources for Geospatial Knowledge Graphs and GeoAI. Open Access. https://doi.org/10.1111/tgis.12971

The article summarised in Section 2.2 (Fibæk, Keßler, et al., 2021) presented a methodology for estimating various structural characteristics using the Sentinel 1 and Sentinel 2 data. This article focuses on improving upon the model design, testing the far-field applications and presents a novel method for converting the structural characteristics into population estimates, comparing them with state of the art – namely WorldPop and Meta's (formerly Facebook) globally population estimate layers (Meta, 2022; Tatem, 2017).

Having accurate, high-resolution population estimates is essential for a wide range of use-cases: Urban planning, pollution estimates, and energy needs, but also for generating good business cases for Financial Service Providers to explore new service areas and improving Financial Inclusion (Fibæk, Laufer, et al., 2021).

The Deep Learning model was updated to include adaptive pooling layers, output clipping, a significant increase in the number of parameters, spatial dropout, and batch normalisation after each concatenation. These efforts improved the prediction results and the model's ability to generalise. The presented methods focus on predicting the structural area, which means that the methodology should be globally applicable, as most countries have access to some level of vector data on the structural area through public participatory GIS services such as OpenStreetMap. The study areas were the entire Mediterranean coast of Egypt and the whole country of Ghana. Volume and population were not predicted directly, as no labelled datasets were available. Area predictions were used as the basis for the population estimates.

The population estimates were made in three steps: First, the structural area of the study areas was predicted using ground truth data collected in collaboration with the University of Alexandria in Egypt and the Centre for Remote Sensing and Geographical Services in Ghana (see Figure 8). The area prediction models performed as well or better than models using very high-resolution input data, although at a coarser minimum mapping unit. The models trained to predict structural area were adapted, through Deep Transfer Learning, to classify the types of the area as either: residential, non-residential, or self-organised. These classes, along with the predicted area and the regional demography, were used to generate high-resolution daytime and nighttime population estimates. These estimates were compared to the state-of-the-art and were promising, performing as well or better in most comparisons.

2.4 AUTOMATIC DETECTION OF HISTORIC STONE WALLS

Trotter, Ezra., Ana Fernandes, Casper Samsø Fibæk and Carsten Keßler. "**Machine Learning for Automatic Detection of Historic Stone Walls Using LiDAR Data.**" International Journal of Remote Sensing. Volume 43, 2022 – Regular Issue 6. Open Access. **https://doi.org/10.1080/01431161.2022.2057206**

This article is significantly different from the other three and came about as Casper was supervising Ezra and Ana at NIRAS for their master thesis – which was a continuation of previous work done by Casper at NIRAS. The project used the Buteo Toolbox for preprocessing, tiling, and predicting the Deep Learning model and shows the broad application area of the tools available in the toolbox.

Protected stone walls are an essential part of the cultural and historical Danish landscape. They are important for supporting biodiversity and cultural heritage and are protected by Danish law. However, many of the protected walls and dikes have never been registered or are registered in the wrong location. As the walls have distinct geographical features, the article presents a methodology for mapping these using the national terrain models (GeoDenmark, 2021) and Sobel filtering (Sharifrazi et al., 2021) as the input data, Deep Learning as the analysis methodology, and the Buteo toolbox as the analytical framework.

The Island of Ærø was chosen as the study area due to its high amount of protected stone walls in relation to its size. Two methods are investigated, one for verifying the existence of registered stone walls by drawing orthogonal cross-sections along the length of the registered walls, extracting the terrain information, and conducting a regression analysis to determine the presence of a wall. The other is a convolution neural network trained on rasterized, anti-aliased and verified walls to predict the presence of walls over the entire study area. The last approach proved the most promising and has since been adopted by the Danish Ministry of Culture for national application. The design of the presented Deep Learning model is now being updated to use the Deep Learning model design used for extracting structural characteristics in Ghana and Egypt.

While the methods presented performed well, there were also drawbacks. Areas along roads and drainage canals along agricultural fields are often classified as walls, and the model has issues in densely forested areas. The produced rasters also show some border noise due to the model not using adaptive pooling layers and an average prediction merging methodology. The border noise issue has since been fixed in the Buteo Toolbox, which now uses Median Absolute Deviation (MAD) adjusted prediction merging (see Figure 26) as a standard for model prediction (Fibæk, 2022b).

Intentionally left blank.

3. EARTH OBSERVATION AND DEEP LEARNING FOR SUSTAINABLE DEVELOPMENT

This section presents specific contributions of this PhD project to sustainable development through the use of Earth Observation (EO) data and Deep Learning (DL). It focuses on the application of the contributions, while the following section presents the tools and methods involved. The presented applications are diverse to show that the same data sources, tools, and methods can be used to solve diverse and complex issues and help drive data-informed decision-making. The covered themes are Financial Inclusion, Integrated Coastal Zone Management, cultural heritage and biodiversity and health care coverage. The themes use the same common toolbox to support sustainable development in novel ways. Before presenting the applications, an introduction to using EO data to inform decision-making is made.

A large amount of research is focused on applying EO at the national and international level to support the work of National Statistics Offices and their reporting and monitoring metrics for the Sustainable Development Goals (SDGs) (Andries et al., 2019; Kavvada et al., 2020). This project instead focuses on the downstream applications of EO that can support the SDGs themselves: Deep Transfer Learning of population density and urban infrastructure estimates from the data-rich to data-poor countries, improving Financial Inclusion by enabling Financial Service Providers to explore business cases in under-served communities, create flood risk management tools, management plans for invasive plant species and the identification of unmapped cultural heritage walls and dikes.

3.1 DATA AND DECISION-MAKING

In 1968, when the Mexican government decided to build a new resort town to attract the growing number of American tourists, they used a computer and a data-informed approach to find the optimal location. The government looked at the proximity to Disney Land, availability of local labour needing jobs, access to water, good weather, presence of sharks, and even the patterns of migratory wood-cutter ants. The choice was narrowed down to 25 undeveloped locations and each physically visited before finally deciding on Cancun (Figure 5). The formula was so successful that it has since been used to develop multiple resorts (Renee Pelas, 2011; L. Steele, 2019).

Figure 5. Cancun in 1970.
Image taken before the development of the Resorts began. © Yucatan magazine.

Planned tourism is an excellent example of data-informed decision-making and shows the benefits of Decision Support Systems. A Decision Support System is an interactive system to assist people in decision-making by facilitating easy access to information, documentation, people, and analysis. Its simplest form can be described as a computer-based system to assist decision-making and generally focuses on helping decision-makers solve *ill-defined* problems (Sprague, 1980).

Research into Decision Support Systems was formalised under the name "Management Support Systems" by Gorry & Morton, 1971 and generalised from *management* support to *decision* support by Sprague, 1980. The research was driven by a desire to support business managers in making informed decisions that relied less on intuition. During the 1990's spatial data was increasingly incorporated into DSSs, creating the *Spatial* DSSs. Comprehensive tests of the addition of spatial data to support decision-making found:

> "*... unequivocal evidence that addition of GIS technology to the decision environment for a spatial reference decision task reduced the decision time and increased the accuracy of individual decision-makers*." - Crossland et al., 1995

With the increasing amount of available data came the concept of "*Data-Driven Decision-Making*", an umbrella term for multiple types of data-use in the decision-making process. Data-*driven* decision-making relies on hard data to make a final decision, while for data-*informed* decision-making, data supports and informs the decision-making process. In literature, data-driven can refer to both types of decision-making. In this project, the following definitions have been used:

Data-*Driven* Decision-Making
An automated system that ultimately decides on a specific action.
Example: Automatically fining farmers for non-compliance with agricultural policies using Earth Observation and Deep Learning.

Data-*Informed* or *Human-in-the-loop* Decision-Making
A system where the decision is ultimately in the hands of a human caseworker, but they are supported and potentially recommended a specific action based on data.
Example: Finding the optimal placement of a new bank branch or holiday resort by visiting recommended sites from a SDSS.

The Spatial Decision Support System presented in this project for Financial Inclusion supports data-*informed* decision-making. As part of a whole, the produced data and the user interfaces are referred to as a Spatial Decision Support System.

There are many examples of using Earth Observation to support the Sustainable Development Goals (SDGs) through Spatial Decision Support Systems, mainly through global annual statistics of land cover/land use and population, as well as monitoring water resources, early warning systems for crop failure, average household pollution estimates and the GEOGLAM project for global agricultural monitoring (Whitcraft et al., 2015). A good example of a recently open-sourced decision support system is the Google Global Drinking Water project, which uses geospatial data to map global potentials for extracting water from solar energy (Lord et al., 2021). The study uses WorldPop (Tatem, 2017) and geospatial data to map potential and find areas for possible interventions.

Research into the development of these SDSSs for the SDGs benefits from the data revolution and the strides in Artificial Intelligence (AI) and Modern UI/UX Design research. AI research is headed mainly by American universities such as Stanford and MIT and industrial labs. ESA and the French Space Agency CNES implement these cutting-edge tools in the EO sphere through the SNAP and Orfeo Toolboxes (European Space Agency, 2021; Grizonnet et al., 2017). Google has made much of their AI research available through the TensorFlow library and Meta (Facebook) through the PyTorch library. These two libraries are a basis on which a significant amount of research is being conducted (Abadi et al., 2016; Paszke et al., 2019).

When incorporating AI into the decision-making process, it is crucial to be mindful of introducing potential biases (Osonde & Welser, 2017). In the case of Financial Inclusion, it is important to be aware of the inherent biases in such systems. They are biased both by the intentions of their creators and the training data used. In application areas such as remote credit scoring and access to loans and other financial services, there is considerable potential for significant unintended human consequences caused by ill-designed models (Wang, 2021).

Osonde & Welser, 2017 suggests in the book “An Intelligence in Our Image” to use a “human-in-the-loop” approach or “data-*informed* decision-making” approach to not introduce additional unknown biases into the final systems. A famous example of bias in AI research is in health care, where Deep Learning and computer vision can support medical diagnosis (Esteva et al., 2019). However, sometimes there is a significant data disparity in how these models have been trained. Suppose the training data is collected solely from people in high-HDI countries or single ethnic groups. The resulting model’s predictions could be wrong or biased for a large segment of the global population, especially in the Global South (Gao & Cui, 2020). Another example is biased training data used in predictive policing algorithms, leading to discriminatory outcomes (Oswald et al., 2018) or AI used to detect childcare fraud with incomplete training data and little oversight, leading to the Dutch Childcare tax scandal, which resulted in the government resigning (Ranchordás, 2022).

Like the data inequality in healthcare, there is a significant global imbalance in the availability geospatial data. Despite being home to almost half the global population, low and medium HDI countries account for only 28% of the total amount of structures mapped in OpenStreetMap (Herfort et al., 2021). Gao & Cui, 2020 and Fibaek et al., 2022 suggest using Deep Transfer Learning to alleviate the impact of data inequality and be mindful of the data availability gap when training Deep Learning models.

3.2 FINANCIAL INCLUSION

Financial Inclusion concerns the equal access and availability of useful financial services. Some examples of these services are access to loans, insurance, savings accounts, or general banking. Efforts to increase Financial Inclusion focus on under-served communities and map areas and communities systematically excluded and under-served. These communities are often vulnerable or marginalized groups that face discrimination (Varghese & Viswanathan, 2018). It is part of several Sustainable Development Goals, but mainly number one: *No Poverty* (Fibæk, Laufer, et al., 2021). In the absence of meaningful access to formal services, traditional saving mechanisms such as Susus are used in many countries. A Susu is an informal savings club or organisation where people, predominantly women, contribute a fixed amount to a fund at fixed terms. The collected amount is paid to a club member at each term based on a rotating queue. The system is based on trust, and interests are not paid. Some FinTechs in Africa are working on connecting Susus to the formal financial sector to offer interest, loans, and physical protection of capital (Peachey & Mutiso, 2019).

The service landscape of finance is changing rapidly, and access to financial services in the Global South has been improved significantly by the increasing access to mobile money services from FinTech services such as M-Pesa and MTN Mobile Money (Porteous, 2006). Traditional banks in Africa are increasingly using agency banking to reach last-mile customers to increase their catchment areas. However, there is a clear tendency for these agents and services to be clustered in the urban cores and service levels drops quickly with distance to urban centres in what is called the "proximity cliff-edge" (Forster et al., 2013).

There is a great divide in service levels between urban and rural communities. Rural areas are often defined using a residual classification. First, urban areas are classified, and accessibility, service levels, and structural density that does not reach certain thresholds are considered "rural". A danger of this approach is over-estimating the homogeneity of rural areas (Woods, 2010). By using such "residual definitions", rural areas will, almost per definition, be underserved communities. Some people argue that the urban/rural divide is better described using local perception and culture rather than physical realities (Wunderlich, 2016). In Fibæk, Laufer et al., 2021, labels for rural areas in Ghana were created with the assistance of an individual with local knowledge. Later, in Fibæk et al., 2022, the approach was changed from an urban/rural dichotomy to measurements of the density of structures to facilitate global cohesion in the measurements and circumvent the use of local definitions (Dijkstra et al., 2021).

Financial Inclusion can be considered a spatial planning issue, and a key issue for measuring it is due to the spatial aspects and the reliance on large scale representative household surveys and census data that take place infrequently and at a scale that makes targeted interventions difficult (Demirguc-Kunt et al., 2018). The increased accessibility of Earth Observation data, improved processing power, and Deep

Learning techniques have fundamentally altered the level of detail available for Financial Inclusion studies (Fibæk, Laufer, et al., 2021; Mondato, 2021).

USING EARTH OBSERVATION FOR FINANCIAL INCLUSION

In Fibæk, Laufer et al., 2021, the authors show that geodata can play a central role in mapping social exclusion and assist FinTechs in exploring novel approaches to expand catchment areas to include under-served communities. The authors identify two important data issues facing national and corporate efforts to improve Financial Inclusion:

1. **National and international level**: What is the current level of Financial Inclusion, how does it compare to other regions, and how is the data trending. Where should interventions be supported, and efforts localised.

2. **Corporate level**: Where are the under-served communities, and what are the appropriate services to target them: formalised Susus, agency banking, brick and mortar bank branches, remote credit scoring, and auxiliary services such as crop insurance or leasing out farm equipment and fertilisers.

In part, measurements of Financial Inclusion derived from Earth Observation data can benefit both issues. The benefits come from the increased spatial and temporal resolution of the data, which can reduce the reliance on large scale household surveys and census data that is insufficient for monitoring missions and exploring business cases. The spatial explicitness of the Earth Observation derived data on population, and georeferenced information on services can provide a more nuanced image of inclusion. Earth Observation can contribute data that enable coherent global proxy measurements of inclusion, such as:

- **Population** - Using structures, structure types, or direct estimates
- **Wealth** - Nightlights, road surface, roof material, physical layout
- **Accessibility** - Road networks, waterways
- **Type of areas** - Commercial, residential, self-organised, rural

Combining this information with data from FinTechs on the placement of bank branches, mobile money agent, and Susus enable a targeted approach to improving inclusion. Supplementing the information with national data or voluntarily contributed geospatial data, such as OpenStreetMap data, allows further granularity by providing access to isochrone travel time calculation, placement of marketplaces, bus and train stations. Additionally, data such as cell phone coverage and historical weather data can be added to inform decision-making further. Combining these datasets allows for highly detailed mapping of Financial Inclusion, customer segments, market saturation, and the identification of potential business areas.

THE SAVINGS AT THE FRONTIER PROJECT

The Savings at the Frontier project is a seven-year partnership between the Mastercard Foundation and Oxford Policy Management, with NIRAS assisting with data analysis and geospatial research. The goal is to further Financial Inclusion by improving the linkage between informal savings groups and the formal financial sector by partnering with FinTechs to reach rural and under-served communities through workshops, tools, and research into last-mile delivery of services, proximity, and novel business cases (Oxford Policy Management, 2019).

Figure 6. Savings the Frontier workshops in Dar es Salaam and Accra.
Top: CEOs and representatives from banks in Tanzania. Casper Fibæk is by the banner.
Bottom: A Savings at the Frontier Technical Session in Accra.

As part of the Savings at the Frontier Project (see Figures 6 and 7), research began to look into the issue of proximity to answer such questions as what does the proximity to financial services mean for inclusion, how far are people willing to travel, and what are the exceptions to the patterns. Forster et al., 2013 and Peachey & Mutiso, 2019 show that the use of financial services is determined by the social and physical distances to Financial Service Providers. These distances are especially pronounced in rural and self-organised areas (Mahendra, 2006), necessitating a formal mapping mechanism of population and neighbourhood types.

The Peachey & Mutiso 2019 report "*Moving Proximity From Critical Issue to Addressable Challenge: Possible Approaches and Tools*" showed that for efforts to increase Financial Inclusion to be successful – Financial Service Providers must be able to find sustainable business models for expanding catchment areas and targeted customer segments. The World Bank Group specifies that one way to create these business models is to invest in new technologies and data-driven business models. An example of this is applying geospatial data to route to market models (Demirguc-Kunt et al., 2018).

Figure 7. Formalising SUSU collection.
Left: Man applying for a small business loan in a self-organised area in Accra.
Right: Hannah Laufer is waiting at the head office of the local mobile money agent.

In collaboration with the FinTech partners of the Savings at the Frontier project, a list of relevant datasets was created that could help them make data-informed decisions about the creation of new business cases and better inform current ones (NIRAS, 2020).

- The population within a certain distance in time or physical distance
- Spatially explicit demography
- Travel time between any two locations
- Wealth Estimates
- Urban traders (Agents, Susus, businesses, banks)
- Nearest: Market place, water source, transport hub, …
- Estimate agricultural output and harvest times

Initially, the aim was to create a national map of neighbourhood types in Ghana, Tanzania, and Zambia to see how services reached different neighbourhoods and map the physical and social distance to services. The initial layer made a traditional classification of urban/rural areas based on perceived local definitions. The approach was since changed to population density-based classes. Classifications are complicated and risk making the conductor a "prisoner of the terminology" (Wunderlich, 2016). In close collaboration with the FinTechs, it became apparent that there was a need to disseminate the information from the Earth Observation classification that was not a static map but a quarriable database the FinTechs and NGOs could incorporate into their software and workflow to track proximity and reach.

A DECISION SUPPORT SYSTEM FOR FINANCIAL INCLUSION

In Fibæk, Laufer et al., 2021, we presented a Spatial Decision Support System, with a frontend embedded within a spreadsheet interface. The system enabled measuring the accessibility and planning of financial services through geospatial analysis and Earth Observation derived products. The Decision support system has been under development throughout the PhD project and currently has 39 functions accessible in the interface and more available through the API and the database. The basis for many of the functions was initially data from the WorldPop project (Tatem, 2017); this was later changed to SenPop (Fibæk et al., 2022) as it became apparent that there was a need for better quality and resolution population estimates, updatable on demand (NIRAS, 2020). The Decision Support System relies on data from the Copernicus programme and Public Participatory GIS through OpenStreetMap. The underlying data has a high temporal resolution which allows a decoupling from the census cycle and ensures the potential of global applicability. The Decision Support System allows querying the Earth Observation data through the Excel interface, and a map in the side pane allows visualisation of layers and manual geocoding of locations. The current version of the system allows, amongst other functions, the query of information about:

- The population within "x" (Isochrones calculated using the OSM network)
- Demographics (Using population estimates and the census demographics)
- Type of area (Classification of area type)
- Nightlights over time (VIIRS)
- Nearest bank, water source, marketplace, point of interest (OSM)
- Cell phone coverage (GSMA)
- Vegetation indices, droughts, weather forecasts.

Casper developed the Spatial Decision Support System, preliminarily called the "SatF Tool", and the continued development and maintenance are conducted with Ana Fernandes and Ezra Trotter at NIRAS. Section 4.7 explains the use of the system. The full GIT history is available on GitHub (Fibæk & Fernandes, 2022; Fibæk & Trotter, 2022).

Figure 8. Collaboratory development of the Spatial Decision Support System.
Top: Interactive learning and feedback session in Dar es Salaam.
Bottom: Collaborators from the Centre for Remote Sensing and GIS, University of Accra.

3.3 INTEGRATED COASTAL ZONE MANAGEMENT

Planning and mitigation efforts are needed to combat sea-level rise and the increase in occurrences of extreme weather events caused by climate change. Coastal zones are home to 40 % of the global population, and 10 % of the global population live in coastal zones at less than 10 meters above sea level (United Nations, 2017). The coastal zones face many threats, such as flooding events, coastal erosion, sea inundation and saltwater intrusion (Intergovernmental Panel on Climate Change, 2022). The vulnerable coastal zones can be managed through Integrated Coastal Zone Management.

Integrated Coastal Zone Management (ICZM) is a process to facilitate sustainable management of coastal areas. It brings stakeholders together to gather data to enable data-informed decision-making, resource management, and monitoring to achieve societal goals, mitigate risks and enable better planning. The methodology considers multiple facets of spatial planning, such as the environment, economic impact, and cultural heritage. It is necessary to collect data and provide continuous monitoring to ensure data-informed decision-making is possible (Fabbri, 1998). Earth Observation data provides the best datasets available for large-scale monitoring of coastal environments in many countries, especially in the Global South, where less data is available (Elnabwy et al., 2020).

Figure 9. Flooding in Alexandria, Egypt. Oct. 2015. AP Photo - Heba Khamis

The Mediterranean coast of Egypt and the Nile delta is susceptible to sea inundation, flooding, and salt-water intrusion of the agricultural lands located, and Egypt has one of the highest numbers of people living in the Low Elevation Coastal Zone in the world (Abou-Mahmoud, 2021; Neumann et al., 2015; NIRAS, 2021).

A Deep Learning model for predicting structural area, type and population density was created based on experiences from the population estimates made for Denmark and Ghana (Fibæk, 2022b; Fibæk et al., 2022). It was produced using the Buteo Toolbox and used to facilitate ICZM in Egypt (Fibæk, 2022b). Local training data was collected with the assistance of the University of Alexandria. A belt of 100 km along the entire Mediterranean coast and the Nile Delta up to Cairo was mapped, and the following data was collected:

- Land Cover / Land Use Classification (European Space Agency, 2020)
- Area and Type of Structures (Fibæk et al., 2022)
- SenPop - Population (Fibæk et al., 2022)
- Rivers and Oceans (OpenStreetMap)
- Terrain Models (Initially Copernicus DEM, then Hawker et al., 2022)

These datasets form the basis of a vulnerability ranking map, which is continuously updated with additional layers and higher resolution data sources (Abou-Mahmoud, 2021) as they become available. Using the input data to create flooding scenarios allows the creation of impact assessment and valuation maps. The *Forest and Buildings Removed DEM* (FABDEM), which is based on the Copernicus DEM, is currently used for testing (Hawker et al., 2022). However, for increased accuracy – especially regarding rivers and minor bodies of water, it is necessary to gather higher resolution data, such as from LIDAR Scans. An example is shown in Figure 10.

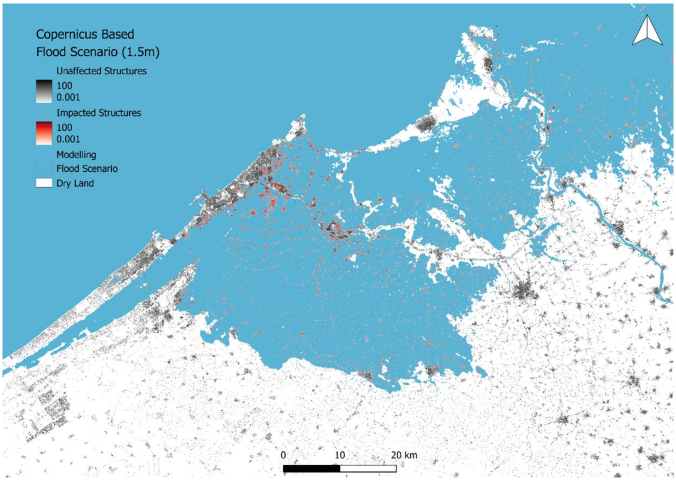

Figure 10. Flooding scenario and impacted structures. Calculated using the Copernicus DEM and the SenPop Structures. Based on an original figure created by Marcia Trillo.

The Buteo Toolbox played a central role in downloading, preprocessing, and analysing the Earth Observation data. The Deep Learning Deep Learning Multi-Sensor model design presented in Fibæk, Keßler, et al., 2021 and Fibæk et al., 2022 was used to map structures, areas, and types.

3.4 CULTURAL HERITAGE - PROTECTED WALLS AND DIKES

Historic stone walls and dikes are protected in Denmark due to their cultural and historical value and role in supporting biodiversity. Historic maps noted the location of walls, but many have disappeared or become dilapidated because of maleficence or neglect, depending on the awareness of their protected status (see Figure 11). The registry keeping track of the location and status of these walls and dikes needs to be updated to remove instances where they are no longer present. It is also necessary to locate unregistered walls missing from the official registry (Vejledning Om Beskyttede Sten-Og Jorddiger, 2009)

> *"Man-made, linear elevations of stone, earth, turf, seaweed or similar materials which function or have functioned as fences and have or have had the purpose of marking administrative property or use boundaries in the landscape"* – Description of protected walls or dikes. Guidance on Protected Rock and Earth Dikes, Ministry of Culture, 2009.

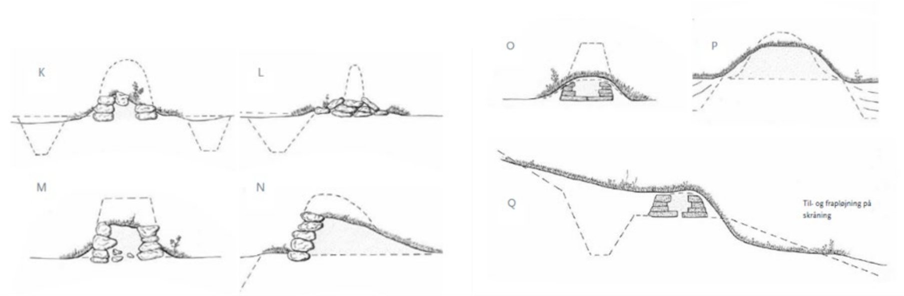

Figure 11. Descriptions from the official guide to protected stone walls and dikes demonstrate the original physical appearance of walls and how they might appear today (Vejledning Om Beskyttede Sten-Og Jorddiger, 2009).

In Trotter et al., 2022, the authors suggest implementing a Deep Learning approach based on the tooling in the Buteo Toolbox (Fibæk, 2022b) to locate the walls using the terrain and surface models derived from the Danish National Terrain Model, which has a 40 by 40-centimetre resolution (GeoDenmark, 2021). A Sobel filter and the Height Above Terrain were applied to the terrain model, which improved the model's accuracy (Sharifrazi et al., 2021; Vincent & Folorunso, 2009). The suggested methodology achieved 93 % accuracy using a pixel-wise comparison between the labels and predictions. An example of the predicted walls is shown in Figure 12.

The model presented in the article was based on a U-Net (Ronneberger et al., 2015) approach and not the Inception ResNet (Szegedy et al., 2017) used in the other three core articles in this PhD Project. The Ministry of Culture has since decided to roll out the model for all of Denmark. As part of that roll-out, the model is being updated to the Inception ResNet style used in the other core papers, as well as adding the following additional layers to the computation:

- **Slope** of terrain model in degrees. Normalised to 0.0 – 1.0
- **Aspect** (COS) of terrain model in radians (r). Normalised as: $\frac{\cos(r)+1}{2}$
- **Aspect** (SIN) of terrain model in radians (r). Normalised as: $\frac{\sin(r)+1}{2}$

In initial testing, adding the above layers derived from the FABDEM (Hawker et al., 2022) to the multi-sensor Sentinel approach described in Fibæk, Keßler, et al., 2021 showed promising results.

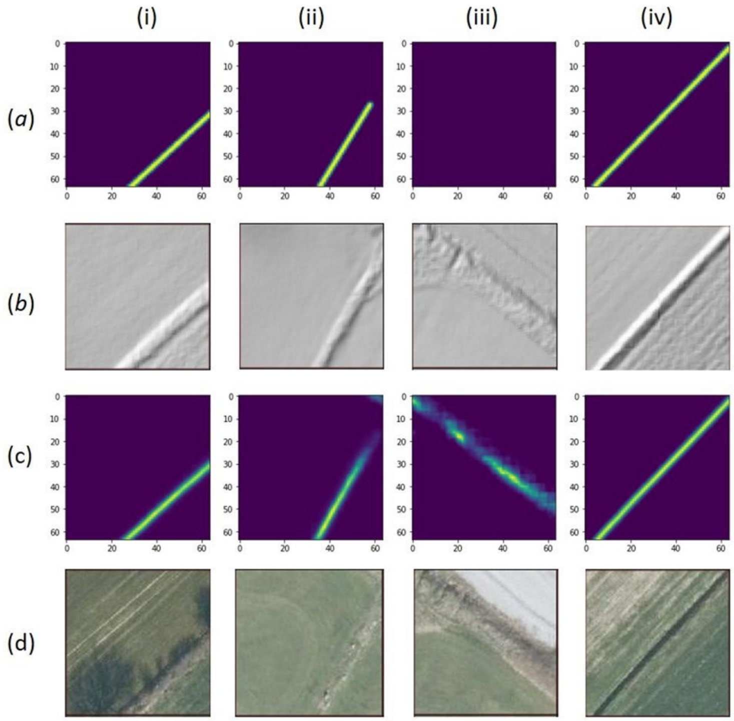

Figure 12. Examples of labels and predictions of stone walls.
Row A: Label, Row B: Terrain Model, Row C: Prediction, Row D: Aerial Imagery.
From Trotter et al., 2022.

The model generalises well but has issues in dense urban areas and forested areas. Dense urban areas are not a target area for applying the model; however, the difficulty in mapping walls in densely forested areas is troublesome. It is likely due to the processing method used to generate the terrain model from the LiDAR point clouds. An additional improvement to the methodology could be to use the raw Point Cloud as input data and convert the neural network to 3D Convolutions (Guo et al., 2021).

The methodology deployed was inspired by the work of (Fibæk, Keßler, et al., 2021) and the data was preprocessed and predicted using the Buteo Toolbox.

3.5 BIODIVERSITY, PLANNING, AND HEALTHCARE

Besides the projects included in the core articles, the methods and tools were applied to a range of distinct and diverse projects to show the diversity of the application field of the tools and methods and Deep Learning and Earth Observation in general.

INVASIVE PLANT SPECIES

Tackling invasive plant species is specifically mentioned as one of the areas in which Earth Observation can support the Sustainable Development Goals (Paganini et al., 2018). The EU has published the EU Biodiversity Strategy to support biodiversity in the European Union, which sets targets for combating invasive species (European Comission, 2020), and the biodiversity law of 2014 compels the member states to manage these (The Prevention and Management of the Introduction and Spread of Invasive Alien Species, 2014). The Danish Ministry of the Environment started a project to investigate the use of Sentinel data to monitor and ease the creation of management plans for five specific invasive plant species.

In Fibæk & Danielsen, 2019, the authors present a system to monitor these five types of invasive plant species in Denmark using Sentinel data combined with other open data sources. When designing decision support systems that use Earth Observation, the systems can often benefit greatly from being combined with other data sources to increase the resolution and improve the models' ability to generalise.

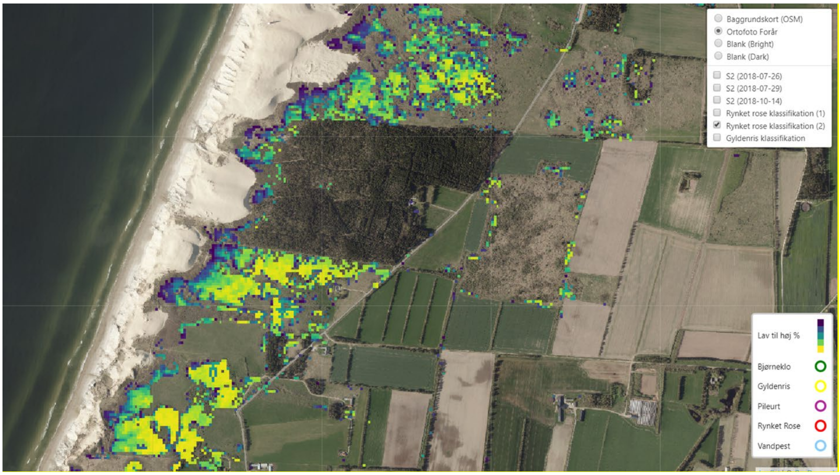

Figure 13. A WebGIS interface to manage invasive plant species.

In collaboration with biologists, plant profiles were created from historical sightings and relevant geospatial data: terrain aspect and slope, soil type, water table, proximity to roads, water, agriculture, and forests. The biologists collected up to date sightings of invasive species on cloud-free days covered by the Sentinel 2 satellites. A Machine Learning model was created using the plant profiles, up-to-date sightings, and Sentinel

2 data to recognise the likelihood that clusters of the invasive plants are currently present at any location in Denmark. A WebGIS platform shown in Figure 13 was created to facilitate plant maintenance.

URBAN PLANNING

Earth Observation is often used to create a snapshot of current circumstances or provide a historical overview of developments. Earth Observation can also be used to estimate future scenarios. By creating land cover classifications of historical Landsat images and settings rules for cellular automata models to map the developments as closely as possible, it is possible to mimic the historical development of cities and extrapolate the results into the future (Arsanjani & Fibæk, 2018).

The methodology is based on 2D Cellular Automata initially described in John Conway's famous article "The Game of Life", in which he describes the building blocks, grids and states of a cellular automaton (Conway, 1970). The state and transition rules of the automaton presented in Arsanjani & Fibæk, 2018, can be interactively adjusted. The changes to future land cover can be inspected, along with graphs showing the development of the built-up areas (see Figure 14). The methods were conceived to predict future urbanisation scenarios in regions with a significant amount of self-organised housing in the periphery of cities.

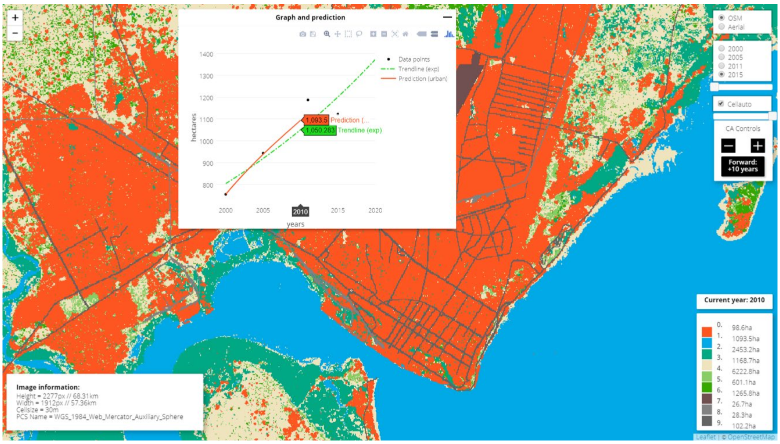

Figure 14. A WebGIS interface to predict unmanaged urbanisation.
The purpose of the model is to predict future urban expansion in Maputo, Mozambique.

The tools show that mapping future urbanisation scenarios using Earth Observation and artificial intelligence makes it possible to design a system to try out different scenarios for the future and plan interactively. The built-up classes were based on random forest classifications, which could be upgraded to the approach described in (Fibæk et al., 2022) to improve modelling accuracy. The current implementation does not support zonal laws besides no-built zones, which is an avenue for future research.

ACCESS TO HEALTHCARE

The Safe Delivery App (SDA) is a Mobile Learning Application for midwives and birth attendants developed by the Maternity Foundation, The University of Copenhagen, and the University of Southern Denmark. The application provides birth attendants with up-to-date guidelines and training for emergency obstetric and neonatal care (Thomsen et al., 2019). The app was introduced to Ghanaian midwives in 2014, and a Ghanaian version of the app was launched in late 2017. If agreed to by the users of the application, the location of app interactions is recorded and anonymously submitted to a central database (Mahboubi et al., 2022).

Earth Observation derived population estimates and urban classifications from Fibæk et al., 2022 were used to calculate the potential catchment areas of the app and determine the market saturation rate. The population estimates were combined with the 2021 Ghana census demographics at the regional level to estimate the number of women of childbearing age and infants (Mahboubi et al., 2022). Earth Observation and Deep Learning combined with public participatory GIS layers allow detailed mapping of customer segments in areas where little other data is available. A subset of the results is shown in Figure 15.

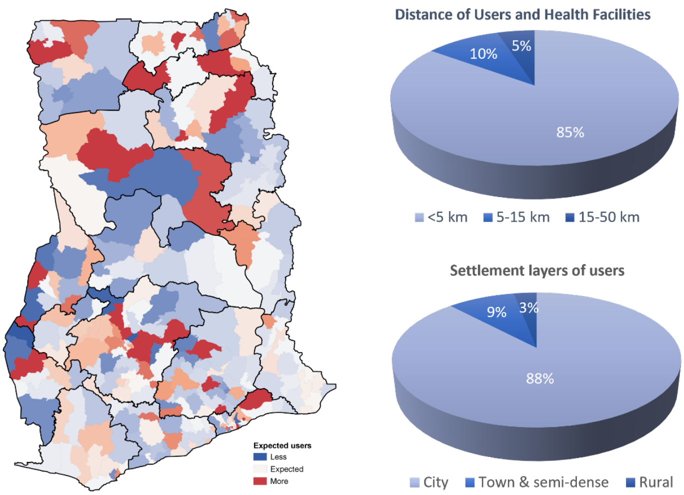

Figure 15. Expected market penetration in Ghana.
More than expected (red), expected (white), and less than expected (blue).
Figure based on Mahboubi et al., 2022.

1.7 million interactions within the application were joined with data on the location of health facilities, mobile network coverage and SenPop data to locate areas with high and low market penetration in Ghana. The results have guided the further roll-out of the SDA and enabled training seminars to be targeted at areas without access to sufficient healthcare services, and helped highlight the regions and districts that need further support to roll out the SDA (Mahboubi et al., 2022).

Intentionally left blank.

4. METHODS AND TOOLS

The initial focus of this project was a theoretical approach focusing on the creation of modular tools and methods for Earth Observation (EO) and Deep Learning (DL) that could be combined in various ways to create bespoke decision support systems to benefit sustainable development. The idea of a modular approach was kept intact in the design of the Buteo Toolbox, which was used as the central analytical framework for the presented projects in the previous section. As the work progressed, the project took on a more applied approach, and the focus became on developing complete systems for furthering sustainable development using EO data and DL. These systems are primarily exemplified through the Savings at the Frontier and Integrated Coastal Zone Management projects.

Supporting open science is important to the European Union (French Presidency of the Council of the European Union, 2022). It enables scalability, increases inclusion, and reduces the barrier of entry – especially in the global south. Open Science policies have been described as an “Accelerator of the Sustainable Development Goals” (Unesco, 2022). All methods have been developed on top of open-source software and are designed to ingest data from open data policy EO programmes and build upon Open Science ideas. The developed tools, data, and methodologies have been made available on GitHub and in the data repositories linked at the beginning of the dissertation.

This section expands upon the approaches presented in the previous section and the four core articles. After presenting the main approaches, the developed tools are described.

4.1 STRUCTURAL CHARACTERISTICS

A central research topic of the PhD has been the production of accurate estimates of structural characteristics from data from the Copernicus Programme. Initially, the idea was to use the double bounce effect exhibited from corner reflectors in SAR data (Koppel et al., 2017) and an object-orientated classification approach to classify different neighbourhood types using Sentinel 2 to do an initial unsupervised segmentation (Fibæk, Laufer, et al., 2021). The first iteration of the methodology used random forest machine learning and was inspired by an approach previously used for crop type classification (Fibæk, 2017). The initial steps were as follows:

1. Unsupervised Large Scale Mean Shift segmentation on Sentinel 2 bands
2. Texture analysis using the Grey-Level Co-Occurrence Matrix
3. Zonal statistics on Sentinel 1 (GRD and SLC), Sentinel 2 bands (10m and 20m) and textures (GLCM)
4. Dimensionality reduction (PCA)
5. Random Forest classification on tabular data in CSV format

The Random Forest object-based approach provided a good initial estimate of what was possible with the data sources for urban mapping. However, the approach was severely limited in its resolution due to the reliance on the unsupervised segmentation approach. Updating the method to using a Convolutional Neural Network, or potentially a Vision Transformer network (ViT), would significantly improve the results, enable transfer learning, and increase the flexibility and spatial resolution.

Sentinel 1 SAR data was the starting point for the research into structural characteristics, and the aim was to investigate how much information could be extracted using that data source alone. Initial testing showed that combining the backscatter with interferometric coherence, as shown in Figure 16, provided good results on the density of urban structures.

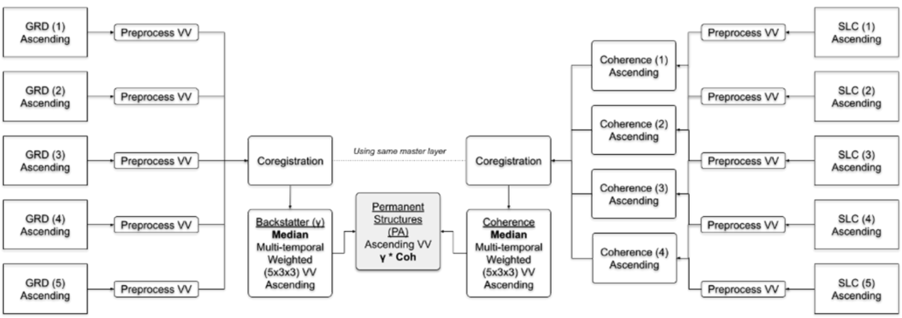

Figure 16. Processing Sentinel 1 data to proxy "permanent structures". The above figure shows only the Ascending data, while the actual tests involved combining data from both the Ascending and Descending orbital directions and both VV and VH polarisations.

The approach was investigated by Fibæk, Keßler et al., 2021 and Fibæk, Laufer, et al., 2021 to determine the impact of different data sources on the accuracy of the estimates. Combining data from both orbital directions and using the GRD (Ground Range Detected) and SLC (Single Look Complex) Sentinel 1 data has several drawbacks: processing coherence is a time-consuming process requiring significant processing power, and both orbital directions are not available over much of Africa. The tests in Fibæk, Keßler et al., 2021 showed that adding coherence and dual orbital direction Sentinel 1 data to the analysis improved estimates slightly but significantly increased processing time. For the Deep Learning models deployed in Fibæk et al., 2022, it was decided to rely only on GRD Sentinel 1 data from a single orbital direction. Figure 17 shows the effects of different orbital directions on the radar backscatter observed over Aarhus, Denmark.

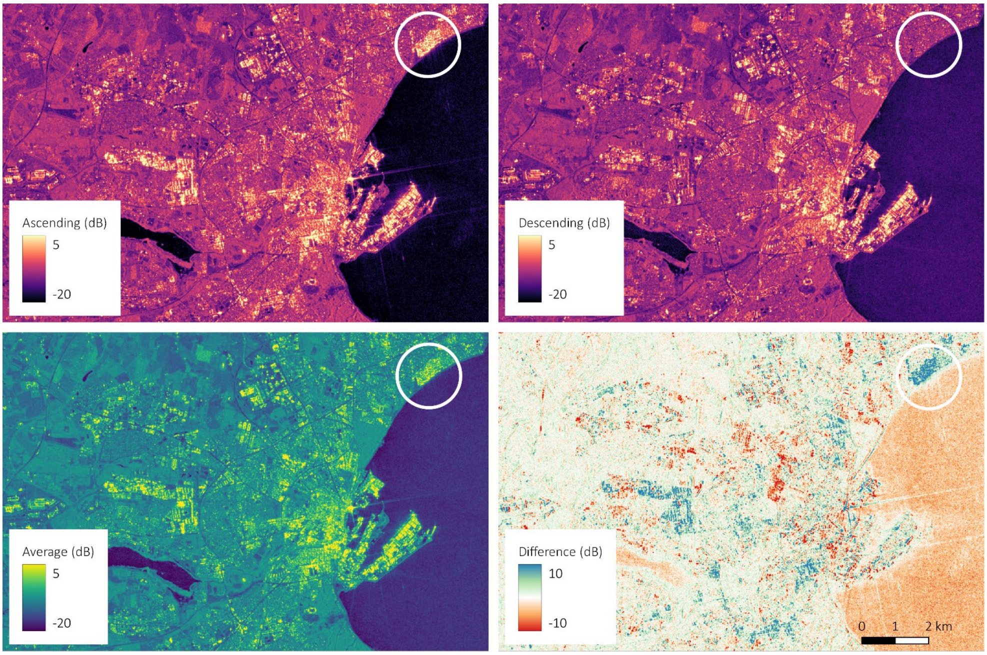

Figure 17. The effects of orbital direction on corner reflectors. Aarhus, Denmark.

As a proof-of-concept, a per tile U-Net (Ronneberger et al., 2015) style Deep Learning model was created to model area, volume, and population in the Central Region of Denmark. The model used 16x16 pixel tiles and made predictions at the tile level using zonal statistics like the initial random forest based approach (see Figure 18). It showed that it was feasible to reproduce the overall structural volume patterns in three different Danish municipalities using Deep Learning (DL) and Earth Observations (EO). The training and testing data used in the initial study and a description of how it was made are available in Fibæk, 2021.

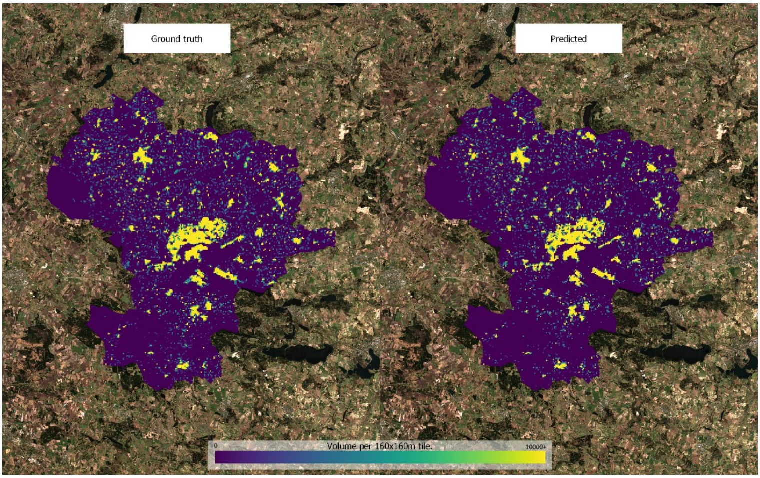

Figure 18. Classification of the structural volume in Silkeborg, Denmark. The method used a summary statistics approach on 160 by 160m grid cells.

After the successful proof-of-concept, an enhanced approach was deployed using a convolutional neural network approach inspired by Inception ResNet and the complexity of the models was increased significantly (Fibæk, Keßler, et al., 2021; Szegedy et al., 2017). The approach works by using a clipped regression output with MSE activation for the networks instead of binary or multi-class classification output activation functions such as SoftMax or sigmoid. A key insight is the need to use the Mean Squared Loss or other loss functions that optimise towards the mean, as they appear to be the only loss functions able to learn due to the small number of structures on most tiles meaning the median value most often is zero (Hyndman & Athanasopoulos, 2018). The final network architecture is described in Fibæk et al., 2022 and the core design is available at artificial_intelligence/model_base.py in the Buteo Toolbox on GitHub.

Some global maps of urban areas from EO derived datasets are based on binary classifications of built-up areas (Corbane et al., 2021; Esch et al., 2017), while the focus of this PhD research has been on estimating the physical characteristics of these built-up areas using regression. A regression-based approach using Machine Learning (SVM) has been featured in research from Germany (Frantz et al., 2021; Haberl et al., 2021; Schug et al., 2021). The main difference between the results in the German studies and the studies presented in this dissertation is the use of Deep Learning and Transfer Learning to enable broad applicability of the models. The models presented here are deployed in data-poor regions, while the German studies focus on Germany. The results from the presented studies were not compared to the German studies as there was no geographical overlap in the study areas. The input tile size of the network had little impact on the model's overall accuracy. Smaller tile sizes increased the flexibility of training parameters by reducing the memory footprint, while large footprints slightly improved the classification of large homogeneous structures, such as warehouses. As a result, 32x32 pixel tiles were chosen as a compromise. The final model was used to predict the area of structures in Egypt and Ghana, using mainly training data from Denmark, with minor amounts of local training data. A subset of the results is visible in Figure 19. The complete results are available and described in detail in Fibæk, 2022a.

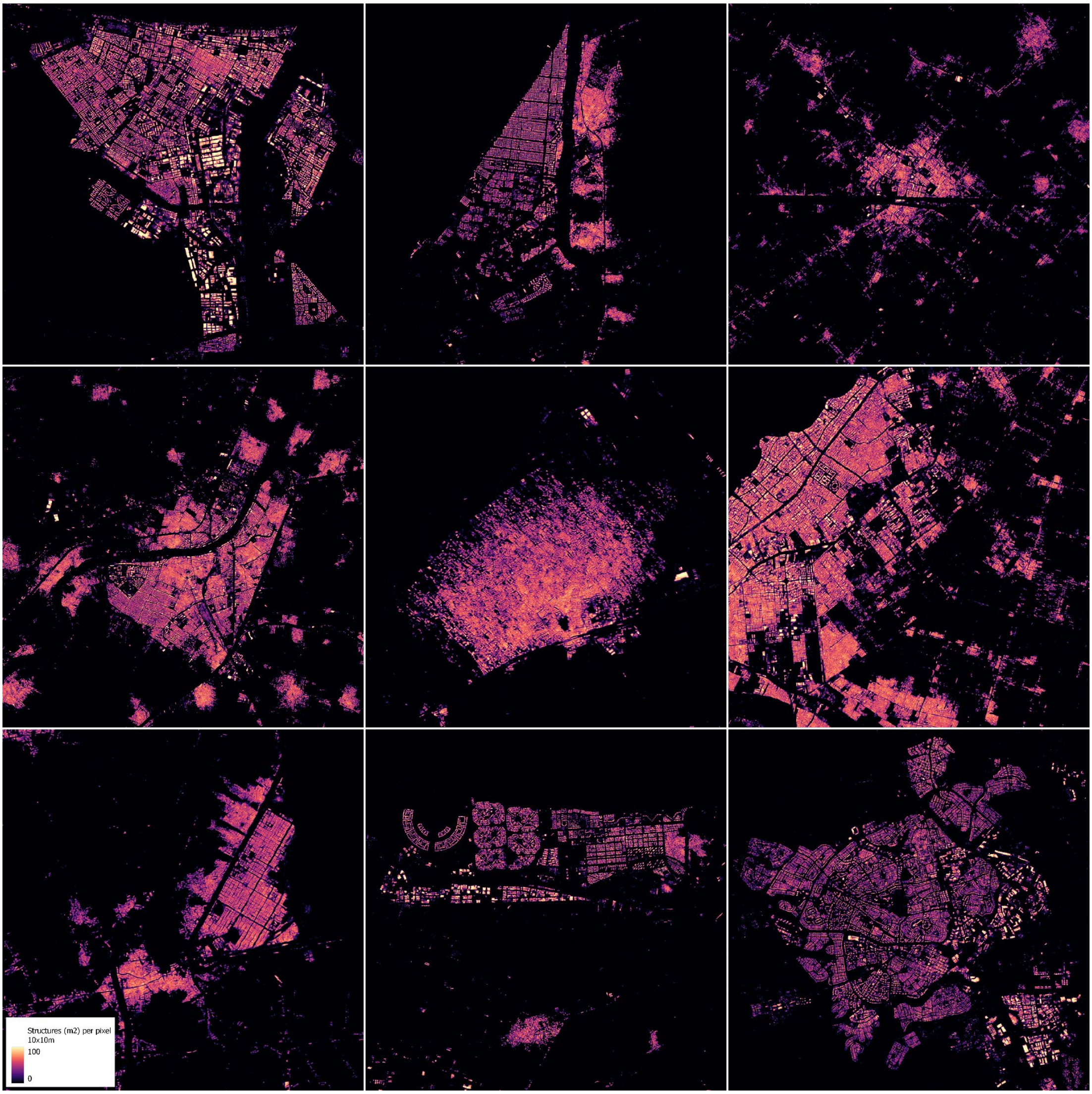

Figure 19. Predicted structures used for Integrated Coastal Zone Management. The figure shows cities and villages along the Mediterranean Coast of Egypt.

4.2 ESTIMATING POPULATION (SENPOP)

Throughout the Savings at the Frontier project, it was apparent that, while mapping and classifying structures was helpful, the most pressing data concern for the Financial Service Providers was access to high resolution and up to data population estimates. WorldPop and Meta have recently shown a method of converting structure count into population estimates at 100m and 30m resolution (Bondarenko et al., 2020; Meta, 2022). Their approaches use data from very-high resolution satellites to do a semantic segmentation of buildings and vectorising the building footprints. These footprints are then aggregated to a grid, and the population census data is distributed according to the proportion of structures within the whole statistical aggregation unit. In Fibæk et al., 2022, the authors introduced a new approach called "SenPop" to derive population estimates from Sentinel data and Deep Learning. Resampled results can be seen in Figure 20. The census aggregation shapes and the boundaries of the underlying very-high-resolution satellite imagery used can clearly be seen in the WorldPop, and Meta approaches, and the counting approach provides a less granular than SenPop.

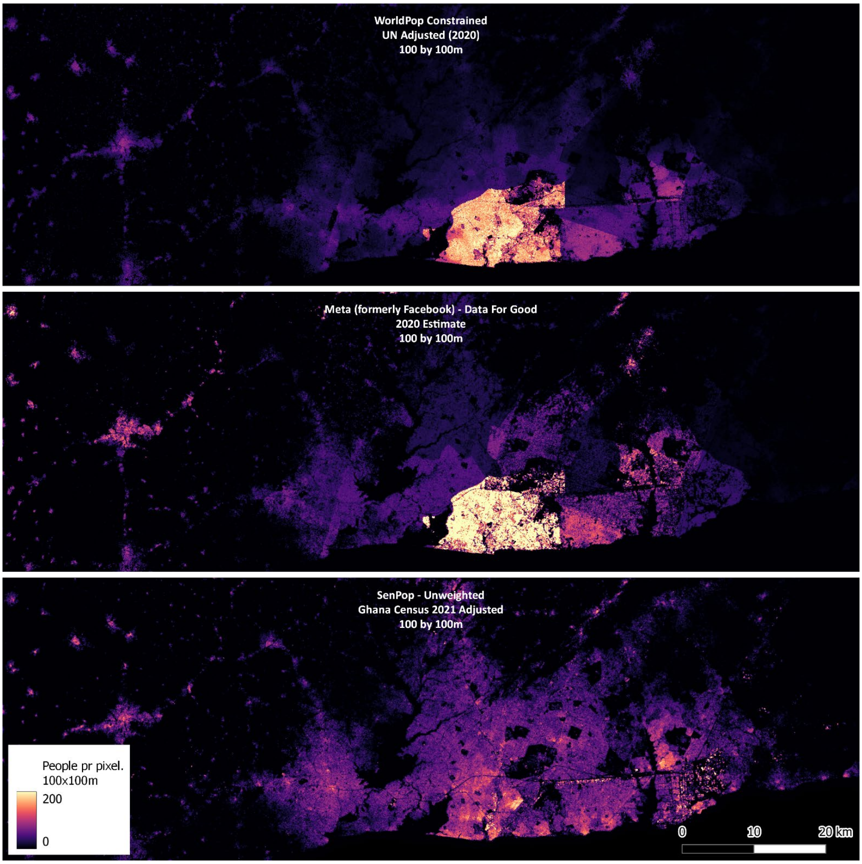

Figure 20. Comparison between WorldPop, MetaPop, and SenPop

The SenPop approach provides estimates at a 10-meter spatial resolution. The increased spatial resolution that comes from the very-high-resolution satellites means the minimum mapping unit of such approaches is significantly smaller than what is offered by the SenPop approach – but population mapping at a scale of 10 meters is already close to excessive. The SenPop article proves that it is possible to create maps of structures at the 10m resolution that provide equal or better accuracy than the very-high-resolution satellite imagery approaches owing to the increased spectral and temporal resolution and the multi-sensor approach used (Fibæk et al., 2022).

The model used for predicting structural area was repurposed using Deep Transfer Learning to classify the type of neighbourhood as either residential, non-residential, or self-organised. The classification of neighbourhood types can be used to further inform the population estimates by providing numbers for daytime and nighttime population (Foley, 1954).

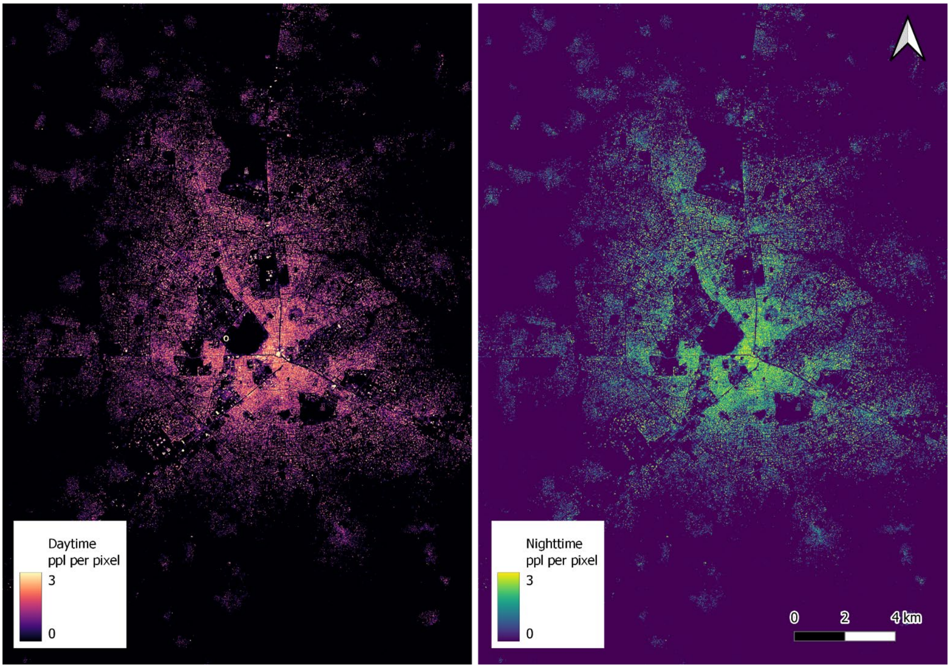

Figure 21. Daytime and nighttime population in Tamale, Northern Ghana.

Daytime and nighttime population estimates can be very beneficial for Financial Service Providers to determine the expected population throughout the day. The estimates rely on a coarse conversion scale, subject to regional differences. However, since re-calculation is a matter of changing the weights assigned to people per m^2 of residential or non-residential structure, calibration can be done quickly on the CPU to investigate differences. Examples of SenPop daytime and nighttime population estimates can be seen in Figures 21 and 22.

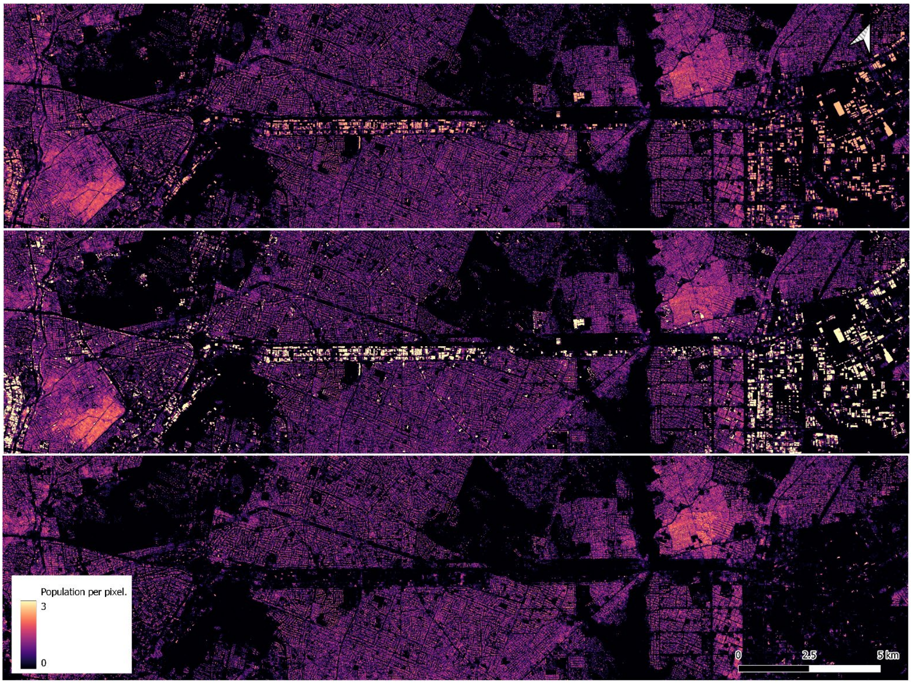

Figure 22. SenPop time-dependent population predictions with default weights. Top: Unweighted population. Middle: Daytime population. Bottom: Nighttime population. Accra and Tema, Ghana.

Both Fibæk, Keßler et al., 2021 and Frantz et al., 2021 showed that estimating building heights or volumes using Sentinel 1 data is possible. In the initial draft of the study to predict the population in Egypt and Ghana, volume estimates were also created but were since removed from the article as the results were not good. Coarse ground truth was collected through Ghana by visually estimating the height of 43 flat-roofed buildings in Google StreetView using the approximate height of known objects. The result was not promising, with a mean average error of 1.6 meters and a mean absolute percentage error of 36.4% in urban areas. The model generally under-predicted with a total percentage error of -19.0%. Fitting the data with minor amounts of local training data on building heights and volume would likely boost the accuracy significantly.

All the population estimates presented suffer from working in two dimensions, as houses are three dimensional. While the Danish models applied directly to Ghana yielded poor results, a solution could be to do a tiered classification into single, multi, and high-rise buildings instead. These three classes could then be assigned a more reasonable proportion of the floor area in target areas. In suburban and rural areas, the effects of not working in three dimensions are likely negligible, but more research should be conducted into globally applicable population estimates.

4.3 TRANSFER LEARNING

Earth Observation and Deep Transfer Learning can help alleviate the impact of the global inequality regarding access to high-quality training data (Fibæk et al., 2022; Jasper et al., 2021). At its core, Deep Transfer Learning is training a neural network for one task and reusing the learned weights for different tasks. An example of such is using a Convolutional Neural Network trained to differentiate between cats and dogs as the basis for a network to classify hats by style or era (Trask, 2019).

Housing and settlement patterns, materials, and textures look different throughout the world. However, a network trained on classifying structures in one area – can be repurposed or refitted with minor amounts of additional data to predict structures in different geographies. Furthermore, a model trained to estimate the area of structures on the ground can be augmented through Deep Transfer Learning to classify structure types or predict volume and building heights (Fibæk et al., 2022). This flexibility offered by Deep Transfer Learning shows promise in lessening the data disparity.

For many geospatial Deep Learning tasks, Denmark and other data-rich countries can serve as a rich-data case-study area to train models. Geospatial labels on roads, agriculture, structural area, volume and population, and aerial, lidar and oblique imagery are freely available and cover the entire country. Figure 23 shows examples of available training data in Denmark that served as the basis for predicting the population in Ghana and Egypt.

Figure 23. Prepared training data for structures, Denmark.

By using Deep Transfer Learning to repurpose the area prediction model to classify structure types in Ghana and Egypt, it was possible to produce population estimates split into daytime and nighttime populations and provide a basis for mapping self-organised communities (Fibæk et al., 2022).

4.4 THE BUTEO TOOLBOX

The idea of a modular approach to Earth Observation Spatial based Decision Support Systems was the starting point of the research for this PhD. While the focus shifted to creating more end-to-end systems, the Buteo[4] Toolbox has continuously been expanded by incorporating the tools and methods used to complete the research. It is divided into sub-folders focussing predominantly on the processing of raster data, readying data for Deep Learning analysis, and the interoperability of the toolbox with functionality from the SNAP and Orfeo Toolboxes (European Space Agency, 2021; Grizonnet et al., 2017) by providing custom python bindings to these libraries.

There are many novel concepts included within the toolbox that has not been covered substantially by the academic work submitted alongside this dissertation. Concepts such as ellipsoidal (see Figure 24), distance weighted, and fractional kernels (buteo/filters/kernel_generator), unique approaches to creating mosaics from multispectral images (buteo_eo/sentinel2_mosaic) and easy prediction and tiling of Earth Observation imagery.

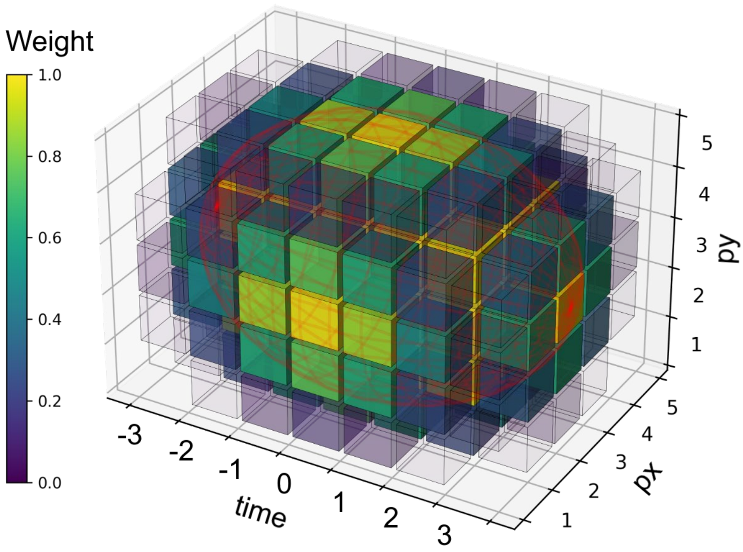

Figure 24. Fractional ellipsoidal kernels are available in the Buteo Toolbox (before normalisation).

[4] The name Buteo is a play on the name of the common buzzard (Buteo Buteo), which in Danish is the same word as "awake" or "alert". The word contains the abbreviation of Earth Observation (EO), and the keen eyesight of the predatory birds is reminiscent of the sensors on the multispectral satellites. As an additional benefit, the name is close to one of its inspirations, the Orfeo toolbox.

The novel mosaic approach works by assigning a quality score, based on the scene classification, feathered circular kernels, and pixel values, to each pixel and picking the highest overall quality image. Cloud-free data from other images are added to this image until a certain quality threshold or a maximum number of images is reached. The pixels are merged using a feathered approach and optional colour balancing. The mosaic approach has been used in all presented studies for the Sentinel 2 imagery. Sentinel 1 imagery is mosaiced by preprocessing the imagery using the Buteo SNAP Bindings by applying a weighted median merging technique using fractional ellipsoidal kernels of multiple timestamps to reduce the speckle noise. Figure 25 shows some of the functions available in the toolbox.

Earth Observation
- Download Sentinel Imagery
- Mosaic Sentinel 1 (incl. preprocessing)
- Mosaic Sentinel 2 (incl. preprocessing)

...

Artificial Intelligence
- Data Augmentation (Scale, band, pixel, drop, clamp, ...)
- Patch Extraction (Tile size, overlaps, borders, align, ...)
- Complete predictions (From raw imagery to predictions)
- Tensorflow callbacks (Overfitting protection, timing, ...)
- Example models (Multi-sensor Inception ResNet, U-net, Transformer, ...)

...

Vector
- Read and Write (index, shape measurement, filter, convert, ...)
- Clip (Vector, raster, targets)
- Reproject, rasterize, merge, intersect, dissolve
- Attributes (Filters, Pandas)

...

Examples
- Preprocessing Sentinel data
- Custom bindings for the Graph Processing Command Line tool for the ESA SNAP Toolbox

Filters
- Novel Kernels (Ellipsodial, distance weighted, ...)
- Convolution Filters (Mad, Sigma Lee, Morphology, DEGURBA)
- Pixel-wise Filters (truncate, normalise, invert, ...)
- Pansharpen

...

Raster
- Read and Write (Stack, meta, NumPy interop, nodata, ...)
- Align (To master or targets)
- Clip (Vector, raster, targets)
- Resample, reproject, shift, warp

...

Bindings
- Custom Python Bindings for the Orfeo Toolbox
- Custom bindings for the Graph Processing Command Line tool for the ESA SNAP Toolbox

Figure 25. A subset of the functions that are available in the Buteo Toolbox.

The Toolbox contains many functions that are not novel but significantly speed up the EO data to DL prediction workflow. Functions such as "align" allow the reading of any number of raster files in a glob pattern to be read, aligned, and clipped to an area of interest provided as a vector or raster format. The most used tools within the toolbox are preprocessing, prediction and postprocessing of Earth Observation data tools located within the artificial_intelligence folder. Imagery can quickly be readied into compressed NumPy arrays with functions provided to calculate overlaps, normalisation, and data augmentation. The same approach can be used to predict raw Sentinel 1 and 2 imagery using a single function providing using overlapped predictions and choosing the MAD collapsed value as described in Figure 26 and Fibæk et al., 2022.

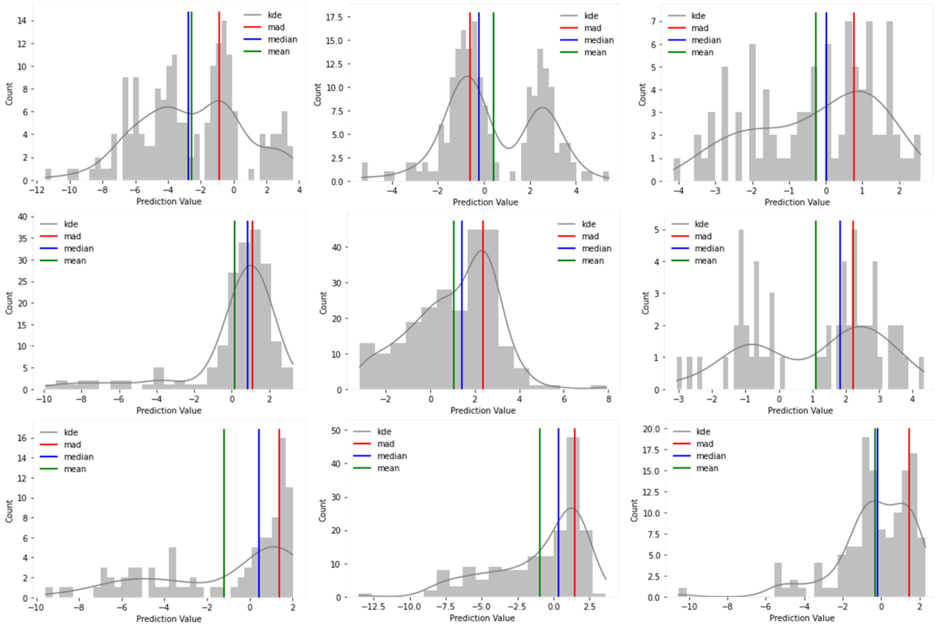

Figure 26. Using the mad collapse method to determine the predicted value out of several predictions.

Providing bindings to the SNAP and Orfeo Toolboxes means that the user does not have to leave the single python environment but regrettably also significantly complicates the maintenance of the library. The other dependencies are kept to a minimum to reduce these maintenance requirements, and the integration with SNAP and Orfeo is kept optional. The core dependencies are NumPy, Pandas, and the Geospatial Abstraction Library (GDAL) (Warmerdam, 2008). Tensorflow (Abadi et al., 2016) is required to make Deep Learning predictions but is not required for other functions. Predictions with PyTorch (Paszke et al., 2019) are currently not supported but are scheduled for a future release.

Besides the functionality, the toolbox provides an example folder with examples for preprocessing, analysing, and predicting Earth Observation data using Deep Learning. Furthermore, the template for the multi-sensor model presented in Fibæk et al., 2022 is provided in the artificial_intelligence folder, along with an example of how to train the model. The entire library is available from Fibæk, 2022b, which also contains example notebooks for using the toolbox.

The toolbox is available on PyPI and Conda:

```
!pip install buteo --upgrade
import buteo as beo
```

4.5 EASY TOOLBOX GUI

Many Earth Observation and geospatial data practitioners conduct their analysis without writing code but instead use GUIs such as QGIS or the Orfeo Toolbox. The Easy Toolbox GUI Python framework was created to ease the creation of a bespoke toolbox tailored to users' needs within NIRAS and other organisations. It is based upon PySimpleGUI and the QT Framework (Driscoll, 2022; Qt Company & Qt Project, 2022) and enables the creation of toolboxes in the style of the Orfeo Toolbox (Grizonnet et al., 2017) for any Python framework. The main interface contains two panels: One for viewing function descriptions and selecting the function to run and another for running the python function while setting parameters interactively. A basic example of an interface for creating .CSV files is shown in Figure 27.

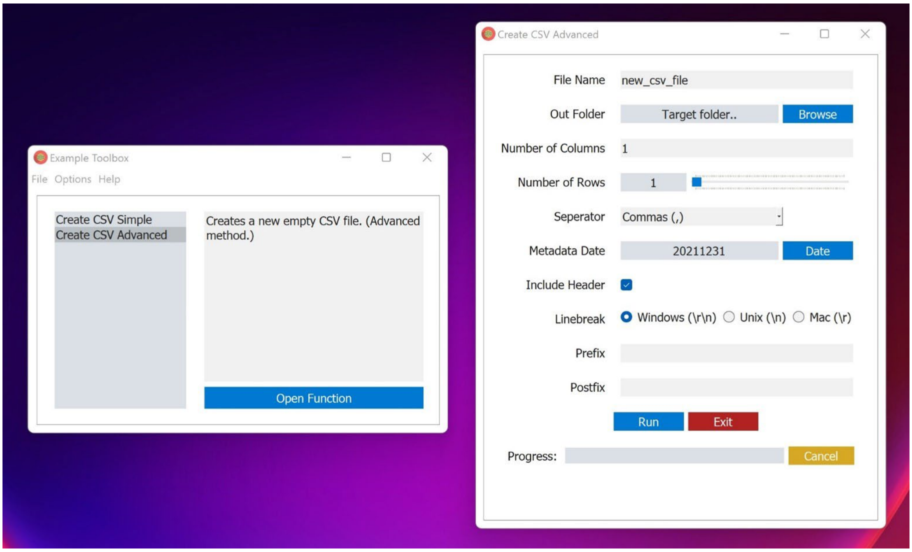

Figure 27. GUI created using Easy Toolbox GUI from a simple setup file.

The framework makes it easy to create cross-platform toolboxes from python functions contained within a single executable. A GUI has been made to download, preprocess, and predict Earth Observation data interactively and has been used to facilitate Earth Observation analysis without using code within NIRAS. It was initially part of the Buteo Toolbox but was since moved to be able to benefit projects that were not Earth Observation or Deep Learning orientated.

The framework exposes a single function called “create_gui”, which takes a JSON ordered Python dictionary as a setup file. The dictionary can take any number of tools that are then added to the first window. The python function and its dependencies are directly included in the setup file and must be imported. The resulting executable is a bundled Python environment, and it is possible to expose the running python console directly to view feedback from the running programmes (Fibæk, 2022c).

4.6 S2SUPER

Using pansharpening and Deep Learning to increase the resolution of satellite imagery is not a new concept (Lanaras et al., 2018; Nikolakopoulos, 2008). However, the methodology provided by the S2Super repository merges pansharpening and DL techniques by constraining a deep learning model by a local mean-matching approach. The model is trained by resampling and predicting the NIR band using the RGB bands. The result is a conservative super-resolution approach envisioned as a part of the preprocessing steps.

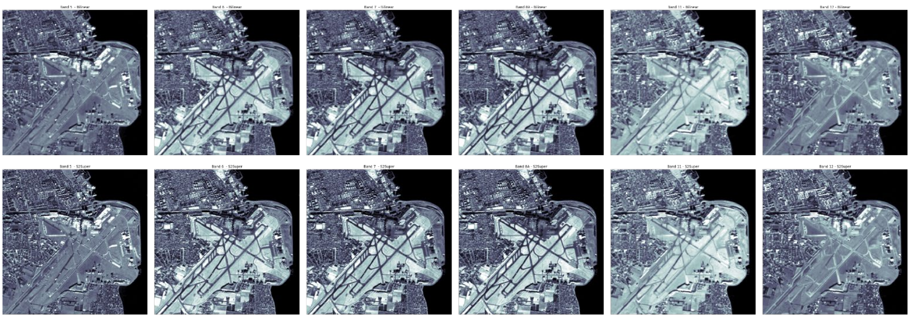

Figure 28. Super-sampled Sentinel 2 bands over Copenhagen during winter.

The core DL model was trained using mosaics from 3000 scenes worldwide to ensure that the model has global applicability. The scenes were chosen to represent each Köppen-Geiger climate zone and every city in the world with over one million inhabitants (Beck et al., 2018). The data ingestion steps in the repository extensively use the buteo toolbox.

The methodology and model are available on PyPI:

```
!pip install s2super --upgrade
from s2super import super_sample, get_s2super_model
```

4.7 THE SAVINGS AT THE FRONTIER INTERFACES

The design of the Spatial Decision Support System for the Savings at the Frontier project came out of the questions asked by the Financial Service Providers at the initial workshops (NIRAS, 2020). It was developed further based on feedback iterations at subsequent workshops and submitted use-cases of the individual Financial Service Providers (Oxford Policy Management & NIRAS, 2021). The providers asked questions such as:

- How many people live within “x” of “y”?
- What type of area is my business in?
- Can the tooling recommend a new area for expansion?

The base of the system is Microsoft Excel. While spreadsheet-based systems are not considered very innovative solutions today[5], they are a central piece of software for most Financial Service Providers. All the participants were familiar with calling excel functions. At the same time, none of the participants had experience with geospatial software. The solution was to bring advanced geospatial functionality to spreadsheets. The idea was for the system to allow integration into the providers' workflows at any level comfortable for their organisation. Some wanted a database, some wanted a Web API for integration into mobile applications, and some wanted a simple interactive GUI (see Figure 28). The solution was to open the backend of the software to direct integration and to use web-based front-end technologies (Microsoft, 2022).

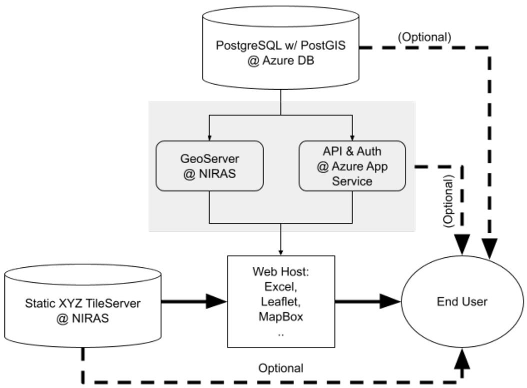

Figure 29. SATF-Proximity tool - Application architecture diagram.
The figure shows how the end-user can integrate the system at different levels.

[5] Spreadsheet software is widely accepted to have begun with VisiCalc (Visible Calculation), released on the Apple II computer in 1979, developed by Dan Bricklin and Bog Frankston. Spreadsheets, as a user application tool, is so popular that it has framed how many think of calculations in what has been called “A Spreadsheet Way of Knowledge” (Levy, 1984).

The system's core design is described in detail in Fibæk, Laufer et al., 2021, but it has since been expanded with additional functionality. Examples of the functions and the interface are shown in Figure 29.

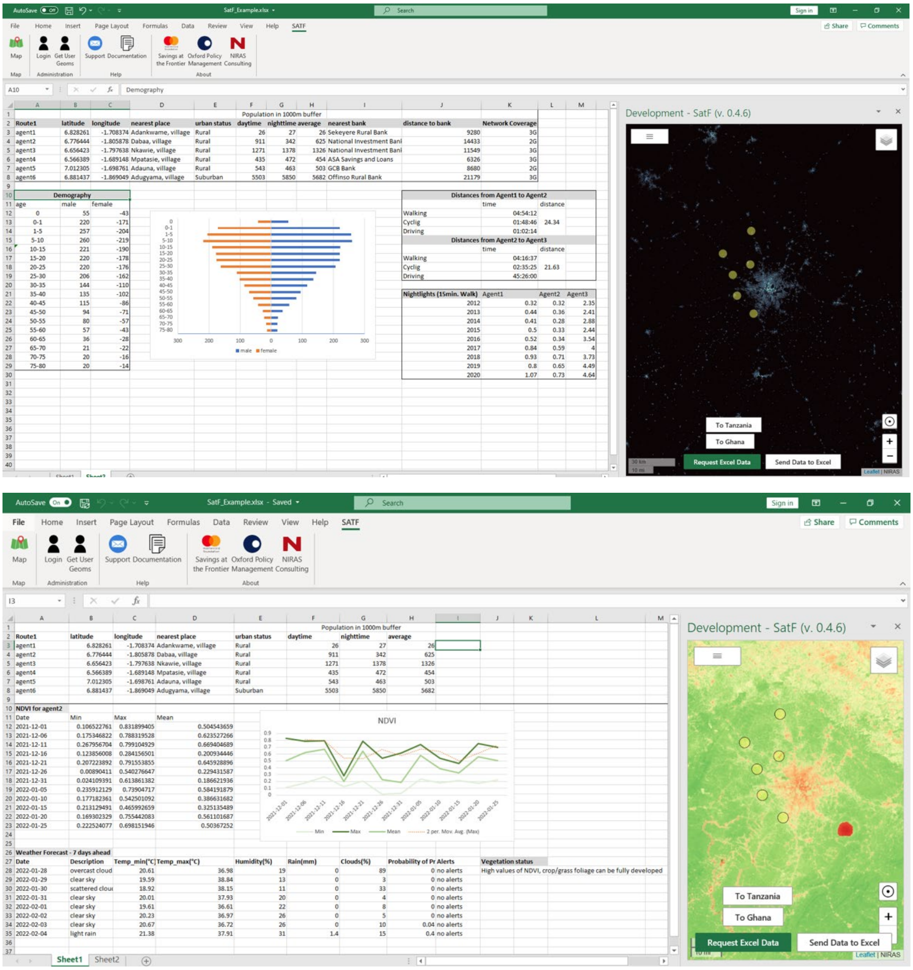

Figure 30. The latest version of the Savings at the Frontier Excel interface (April 2022).

The new functions added to the interface include weather forecasts and NDVI calculations to support AgriTech focused FinTech providers, who supply loans and lease equipment to farmers and Agri-dealers. Moreover, the population estimates have been upgraded by moving from WorldPop to SenPop estimates (Fibæk et al., 2022; Tatem, 2017). Additional basemaps on population and nightlights have been added, and the ability for users to store, share, and retrieve point layers in the geospatial database has been added.

The interface currently has 39 functions drawing on the following data sources:

- Sentinel 1 (Population, neighbourhood classification, structure type)
- Sentinel 2 (Population, neighbourhood classification, structure type, NDVI)
- Official Census data (Demographics, population)
- VIIRS Nightlights (History and proxy of development)
- Open Street Map (Roads, marketplaces, banks, rivers, POIs)
- Weather Forecasts (OpenWeather)
- Address Conversion (Lat/Lng, What3Words, Google PlusCodes, and Ghana Postal GPS)
- Cell Phone Coverage (Both: OpenCellID and GSMA)

The data allows queries such as:

POPDENS_BIKE_ISO(*time_minutes, lat[location]*[6]*, lng)**: How many people live within "x" minutes of travel time on a bike from a given location. Locations can be supplied in one of four valid address systems. The SenPop population counts are used.

DEMOGRAPHY_BUFFER(*distance_meters, lat[location], lng)**: What is the demographic breakdown of the people living within "x" meters of a given location. The function uses the proportions of the newest available demographics from the census for the location and applies them to the SenPop population counts.

URBAN_STATUS(*lat[location], lng)**: What type of area is the given location. Is it rural? Follows the classes defined in Fibæk, Laufer, et al., 2021. DEGURBA based classes are also possible.

The Decision Support System is integrated into the workflow of two Financial Service Providers, one in Ghana (DSS Systems Accra) and one in Tanzania (Digital Mobile Africa), with work ongoing to integrate the system into the workflow of additional providers. An introduction to the interface and the Financial Service Providers is available from the webinar "Mapping it Out: Practical Tools for Reaching Rural Clients"[7], held together with CGAP and Oxford Policy Management. The use of the Decision Support System and data analytics to support business development in developing countries is described in Laufer, 2022.

The system does not provide recommendations, as the Financial Service Providers were reluctant to share data on the location of their services and agents. Instead, the functions and the interface allow the providers to design a recommendation engine tailored to their needs by using the provided functionality as part of an End-User Developed Application (Raković, 2019).

[6] […] denotes that the parameter can be either lat/lng or a location/address. * = Optional.

[7] www.findevgateway.org/findev-webinar/mapping-it-out-practical-tools-reaching-remote-clients

Intentionally left blank.

5. DISCUSSION

Access to high quality, global datasets on population density derived from free data and created using open-source software significantly benefits the efforts towards attaining the Sustainable Development Goals. The research presented in this PhD shows that it is possible to derive high-quality population estimates in the Global South from structural characteristics that outperform the state-of-the-art. The research has shown how these Earth Observation (EO) and Deep Learning (DL) derived datasets can be made available to decision-makers through familiar interfaces while also allowing them to perform complex spatial queries.

Meaning and Implications

The research highlights the broad application areas of EO data and DL methods and their potential for large scale geospatial analysis. As described in European Space Agency, 2018, we are at the beginning of an artificial intelligence revolution. Earth Observation will likely see its role as a central data source for supporting sustainable development grow in the future (Herweijer et al., 2018). As shown in Fibæk et al., 2022, it is possible to estimate structural characteristics, neighbourhood types, and population density for any country, at a high resolution, from a standard laptop, at home, using only free data and open source tools. As the field matures, the scope of predictable targets at high resolution will likely expand to encompass socio-economic variables such as wealth – as has already been explored from very high-resolution imagery by the World Bank Group and WorldPop (Engstrom et al., 2017). As the size of the back catalogue of high-quality data from open data policy satellites increases, it becomes increasingly viable to use EO and DL to do predictive urban planning and simulation, as initially explored in Arsanjani & Fibæk, 2018.

Very high-resolution satellite imagery from tasking satellites is becoming increasingly accessible due to the commercialisation of space. The imagery provides a higher spatial resolution than the Sentinel and Landsat satellites. However, the imagery does not necessarily outperform the open data satellites at higher minimum mapping units (Fibæk, Keßler, et al., 2021). The high performance of the open data policy satellites is due to the satellites' high spectral, radiometric, and temporal resolution. It is crucial to consider the correct data source for a project. For global population studies, the minimum mapping unit of the Sentinel satellites is sufficient for high-quality population estimates at no acquisition cost and lower processing time (Fibæk et al., 2022).

Deep Learning and Deep Transfer Learning methods boost the flexibility of EO analytics by providing flexibility in the input and output data sources and support the creation of generic models that can support diverse analyses. Using DL methods and models that rely on open data as inputs significantly lowers the barriers to developing, sharing, and co-creating models. By having a standard way of sharing and preprocessing open data, it will be possible to do Transfer Learning of complex models from data-rich areas to data-poor areas to lessen the impact of the global data inequality. Denmark, The Netherlands, the United Kingdom and other countries with expansive spatial data infrastructures and open data policies can form the basis for training DL models that could be applied globally. Generic models on structures, roads, and agriculture could form the basis for many future studies (Fibæk et al., 2022). For countries with less-developed or no spatial data infrastructures, Humanitarian Open Street Map and other public participatory GIS services can provide the limited local data needed to train models to make predictions.

EO and DL opens up the possibility of improving Financial Inclusion by supporting Financial Service Providers in the Global South in exploring new, financially viable areas for market expansion. The urban/rural distinction is used to map the "proximity cliff-edge", but relying on per country classifications based on local semantics lowers the value of the datasets as a cohesive global measurement (Peachey & Mutiso, 2019). Changing the basis of the classification to work on the DEGURBA standard derived from the SenPop predictions could enable a global measure of proximity. For the data to be useful, it is important that it is accessible to decision-makers and helps make issues addressable. Through the Savings at the Frontier project, such an interface was created and proved successful in making proximity an addressable challenge (Laufer, 2022).

Limitations

The presented method of population estimation relies on a top-down approach, where the census data is fitted to and distributed using the predicted structural characteristics. As the methodology matures, bottom-up approaches will be increasingly viable, and it will be increasingly possible to decouple census taking and population estimates. The issue with a top-down approach is the increased potential for errors as both the census and the structural maps can be wrong. The proposed methodology attempts to take into account the temporal nature of the population, which is not considered in most censuses and available population datasets – the lack of temporality results in errors along regional boundaries as people commute to work. In many countries, people will spend a significant amount of time a day outside of structures or in transport. These transit periods are not considered, but could be very helpful for creating business cases for Financial Service Providers. Classifying structure types into residential and non-residential areas depends on localised knowledge and is very difficult to determine with a high level of accuracy. Buildings will often be a mix of residential and non-residential, which is challenging to discern from non-oblique imagery. The approaches presented focus on structures that distinctively belong to either class, such as industrial buildings or suburban housing.

The system to disseminate the EO and DL derived datasets to decision-makers do not contain a recommendation engine, as it requires information on the target task to solve. Without access to information on the location of Mobile Money Agents, Bank Branches, Susus, and other services, it is not possible to assess the market saturation and locate under-served communities, similar to what was done in Mahboubi et al., 2022 regarding access to health care services. At the FinTech workshops in Ghana and Tanzania, the Financial Service Providers specifically requested the addition of a recommendation engine for placing new financial services but were unwilling to share data on the location of their offered services (Laufer, 2022; NIRAS, 2020). As such, the implementation of recommendations is left to the providers themselves, using spreadsheets to create end-user applications to produce recommendation engines tailored to their use case (Raković, 2019).

DL models risk of becoming a “Black Box” is a well-described issue where it is difficult to determine the specific reasoning behind a given classification. The challenge has led to the research area of “Explainable AI” and various tools to investigate the reasoning within the models (Vinuesa & Sirmacek, 2021). The research presented in this dissertation does not go into depth with explainable AI and does not deploy tools to investigate the filters and reasoning behind the specific classifications. Some models, such as the s2super model (see section 4.6), constrain the deep learning models by using known physical limitations, which increases the explainability of the models. Suppose DL models, based on the approaches described, are to be used in a legal capacity. In that case, it is crucial to be able to explain the reasoning and biases inherent in the models – as the person has been taken out of the decision-making process. For some cases, it might be more prudent to use more straightforward decision tree approaches, as they offer higher interpretability although with lower accuracy when sufficient training data is available.

Recommendations

The presented approaches for population estimation in the Global South rely on the structural area instead of the volume (Fibæk et al., 2022). Further research should be conducted, and additional training data collected in diverse geographies to improve the population estimates by considering the volume of structures. In the absence of additional local training data on volume, it might be possible to classify ranges of the floor heights of structures using data from data-rich countries alone. The methodology used for preprocessing and merging Sentinel 1 data in the Buteo Toolbox does not consider the “look” and “incidence” angles of the captured data. For some target applications, it might be beneficial to apply weights to the angles. Data collected at steep angles could be given a higher weight if the goal is to map structural volume, as the side of the structure is important to determine its height. A lower weight could be applied if the target is to measure soil water content. Suppose time series of single Sentinel 1 layers are used in a Deep Learning model. In that case, rasters with the individual images' look and incidence angles could be added as an additional layer to improve the results. Training auto-encoder or diffusion models on Sentinel 1 and 2 data through geography-aware, self-supervised learning to predict the location an image was taken, could increase the generalisability of models, when used for transfer learning of the structural volume prediction models (Ayush et al., 2022).

The models designed to predict structural characteristics and type could be repurposed to estimate local wealth and socioeconomic variables. This can be estimated by a model ingesting Sentinel derived variables on the type of structure, neighbourhood type, road material and condition, structure material, and access to greenery. similar estimates to Engstrom et al., 2017 and J. E. Steele et al., 2017, could potentially be made, using open high-resolution data instead of commercial very high-resolution data.

There are several land areas on the Earth where both orbital directions of the Sentinel 1 satellite constellation are unavailable. The approaches presented in this dissertation only use one direction. Suppose the pre-processing approach was changed to generate synthetic data for the one missing direction in areas where it is not available. In that case, both directions could be used in areas where it is possible. Using both directions has been shown to significantly increase accuracy (Fibæk, Keßler, et al., 2021).

Terrain data have been used in preprocessing the Sentinel 1 data, but incorporating the terrain data directly into the model architecture might improve the accuracy further by informing the model of the height, slope, and aspect of the local terrain. However, this runs the risk of causing temporal misalignment, where the prediction on new satellite data is temporally out of sync with the used terrain models – this might especially cause issues in densely urban areas. Adding textures to the input data, such as GLCM, or various indices such as NDVI, might improve results but would significantly increase the training time required, if the model design is kept the same as in Fibæk, Keßler, et al., 2021. Furthermore, due to the nature of large-scale Convolution Neural Networks, the benefit of preprocessed textures will likely be limited. The S2Super model has yet to be deploying as part of the data-processing step of the SenPop model, but the super-sampled sentinel 2 20m bands would be expected to improve results.

6. CONCLUSION

Earth Observation (EO) and Deep Learning (DL) can play a prominent role in supporting the Sustainable Development Goals both in tracking and guiding the progress as well as assisting the achieving the goals themselves (Kavvada et al., 2020). DL models based on data from the Sentinel satellites can outperform models based on very high-resolution data if the minimum mapping unit required is closer to the spatial resolution of the Sentinel satellites. The presented studies show that it is possible to map population with a state-of-the-art accuracy in developing countries while relying predominantly on data from data-rich regions and open data policy satellites and open software. Structural characteristics are a good starting point for population density studies, and the methods can be applied globally at a low cost. The mapping can serve as the baseline for other studies on energy needs, emissions, and various service levels.

The Buteo Toolbox

The presented projects have demonstrated the broad applicability of the Buteo Toolbox in processing and analysing EO data through DL. The same methodologies that work well for classifying structural characteristics can be applied to predict the location of protected stone walls in Denmark. The toolbox provides an excellent analytical base for many research topics involving EO and DL. The toolbox was continuously expanded throughout the work of the four core articles. The methodologies applied and datasets used when using EO to solve problems often converge on a core set of analytical approaches that do not vary significantly across seemingly diverse projects. For example, radar-based change detection can be used to designate areas to update topographic maps of structures. At the same time, it can also be used to monitor reservoir water levels or illegal logging continuously. The approaches could all use the same core methodology of harvesting, cleaning, and algorithmically detecting changes. This recurrence of methodologies has been used to design the toolboxes to support a diverse range of analyses.

Data-Informed Decision-Making and Financial Inclusion

EO data and DL approaches can be presented inside spreadsheets to ease the adoption of spatial decision support systems to make data-informed decisions. It is important to transfer and disseminate EO data and DL approaches into domains that are not foreign to decision-makers. The research presented in this dissertation clearly shows that Financial Inclusion efforts can benefit from EO derived datasets. Two Financial Service Providers have already implemented the tools, with more working on integration to make proximity and improving Financial Inclusion an “addressable challenge”. The proposed decision support system helps decision-makers by a) allowing them to measure the reach of financial services – a significant benefit to NGOs and NSOs tracking progress towards the SDGs. b) enable Financial Service Providers to plan and optimise the distribution of their offered financial services and handle cash flows.

Deep Transfer Learning
There is an immense potential for Deep Transfer Learning from generic EO data-based DL models to benefit many future studies. Using transfer learning techniques, or fitting minor amounts of local data, can drastically widen the application areas of models. Through Deep Transfer Learning, models created to predict the area of structures can be repurposed to predict structural types, volume, and even population, as shown in Fibæk et al., 2022.

Structural Characteristics
Multi-sensor Inception ResNet Style DL networks can predict structural characteristics well. For Denmark, the presented findings show that the model design can predict area with 2.6% total percentage error, volume with 7.6% and estimate population directly at 17.0% accuracy. Improving the design and applying the models to Ghana resulted in a 3.7m^2 Mean Average Error (MAE) at 10m and 1.0m^2 at 100m resolution (when resampled by average). For Egypt, the results were 3.5m^2 MAE at 10m and 1.3m^2 at 100m. For Egypt, both the 100m results and the 10m results outperformed the Google Open Buildings dataset, while for Ghana, the 100m results were similar and the 10m slightly worse. When designing the models, combining Sentinel 1 and 2 data provided the best results while adding interferometric coherence resulted in minor improvements. Using both orbital directions improves accuracy considerably, as found in Frantz et al., 2021 and replicated in Fibæk, Keßler et al., 2021. However, they are not available in much of Africa. The accuracy of predictions declines when the target labels increase in complexity. The models perform best for predicting area, slightly worse for volume, and only moderately well for predicting population directly.

SenPop
The models created to predict structural characteristics were repurposed to predict the structural type. The predictions were combined with census data, and UN forecasts to create top-down population density predictions. The results provided finer granularity than the state-of-the-art, and visual inspection indicates that the results are as good or better than the state-of-the-art. Using the structure types to predict population dynamics makes it possible to estimate the daytime and nighttime population density.

Two additional benefits of Earth Observation to support the Sustainable Development Goals were proposed in the introduction and elaborated upon in this dissertation: (1) it enables data-informed decision-making in areas previously not possible, and (2) the impact of data inequality can be lessened by Deep Transfer Learning of EO models from data-rich to data-poor regions. There is tremendous research interest in Deep Learning and a vast amount of EO data captured daily - we must not let this resource go untapped and actively use it to improve livelihoods and the state of our shared world.

7. BIBLIOGRAPHY

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016*, *10*(July), 265–283. https://doi.org/10.5555/3026877.3026899

Abou-Mahmoud, M. M. E. (2021). Assessing coastal susceptibility to sea-level rise in Alexandria, Egypt. *Egyptian Journal of Aquatic Research*, *47*(2), 133–141. https://doi.org/10.1016/j.ejar.2021.04.002

Amadi, S., Beutler, J., Clark, C., Maron, M., Patel, D., Radform, T., Soden, R., & Uithol, P. (2019). *Open Mapping for SDGs*. https://sdgs.hotosm.org/

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-Spatial Information Science*, *20*(2), 77–96. https://doi.org/10.1080/10095020.2017.1333230

Andries, A., Morse, S., Murphy, R., Lynch, J., Woolliams, E., & Fonweban, J. (2019). Translation of Earth observation data into sustainable development indicators: An analytical framework. *Sustainable Development*, *27*(3), 366–376. https://doi.org/10.1002/sd.1908

Arsanjani, J., & Fibæk, C. S. (2018). Development of a cellular automata model using open source technologies for monitoring urbanisation in the global south: The case of Maputo, Mozambique. *Habitat International*, *71*, 38–48. https://doi.org/10.1016/j.habitatint.2017.11.003

Ayush, K., Uzkent, B., Meng, C., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2022). *Geography-Aware Self-Supervised Learning*. 10161–10170. https://doi.org/10.1109/iccv48922.2021.01002

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2018.214

Bondarenko, M., Kerr, D., Sorichetta, A., & Tatem, A. (2020). Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 51 countries across sub-Saharan Africa using building footprints. *University of South Hampton (WorldPop)*, *10*(2). https://doi.org/10.1371/JOURNAL.PONE.0107042

Buduma, N., & Locascio, N. (2017). *Fundamentals of Deep Learning* (1st ed.). O'Reilly.

Campesato, O. (2020). *Artificial Intelligence, Machine Learning, and Deep Learning*. Mercury Learning and Information.

Chollet, F. (2022). Deep Learning with Python. In *Deep Learning with Python* (2nd ed.). Manning.

Conway, J. (1970). The Game of Life. *Scientific America*, *223*(4).

Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2021). Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications*, *33*(12), 6697–6720. https://doi.org/10.1007/s00521-020-05449-7

Cresson, R. (2020). *Deep Learning for Remote Sensing Images with Open Source Software* (1st ed.). CRC Press. https://doi.org/10.1201/9781003020851

Crossland, M. D., Wynne, B. E., & Perkins, W. C. (1995). Spatial decision support systems: An overview of technology and a test of efficacy. *Decision Support Systems*, *14*(3), 219–235. https://doi.org/10.1016/0167-9236(94)00018-N

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., & Hess, J. (2018). The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. In *World Bank Publications*. https://doi.org/10.1596/978-1-4648-1259-0

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., & Schiavina, M. (2021). Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics*, *125*, 103312. https://doi.org/10.1016/j.jue.2020.103312

Driscoll, M. (2022). *PySimpleGUI*. https://github.com/PySimpleGUI

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Elnabwy, M. T., Elbeltagi, E., El Banna, M. M., Elshikh, M. M. Y., Motawa, I., & Kaloop, M. R. (2020). An approach based on landsat images for shoreline monitoring to support integrated coastal management - A case study, ezbet elborg, nile delta, Egypt. *ISPRS International Journal of Geo-Information*, *9*(4), 199. https://doi.org/10.3390/ijgi9040199

Engstrom, R., Hersh, J., & Newhouse, D. (2017). *Poverty from Space Using High-Resolution Satellite Imagery for Estimating Economic Well-Being*. http://econ.worldbank.org.

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., & Strano, E. (2017). Breaking new ground in mapping human settlements from space – The Global Urban Footprint. *ISPRS Journal of Photogrammetry and Remote Sensing*, *134*, 30–42. https://doi.org/10.1016/j.isprsjprs.2017.10.012

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

European Comission. (2020). *Biodiversity Strategy for 2030 Bringing nature back into our lives*. https://data.europa.eu/doi/10.2779/048

The prevention and management of the introduction and spread of invasive alien species, Pub. L. No. 1143/oj (2014). http://data.europa.eu/eli/reg/2014/1143/oj

European Space Agency. (2015). *SENTINEL-1 Observation Scenario*. https://sentinel.esa.int/web/Sentinel/missions/Sentinel-1/observation-scenario

European Space Agency. (2018). *Towards a European AI for Earth Observation Research & Innovation Agenda*.

European Space Agency. (2020). *WorldCover 2020*. https://worldcover2020.esa.int/

European Space Agency. (2021). *Science Toolbox Exploitation Platform*. http://step.esa.int/main/

Fabbri, K. P. (1998). A methodology for supporting decision making in integrated coastal zone management. *Ocean & Coastal Management*, *39*(1–2), 51–62. https://doi.org/10.1016/S0964-5691(98)00013-1

Ferreira, B., Iten, M., & Silva, R. G. (2020). Monitoring sustainable development by means of earth observation data and machine learning: a review. *Environmental Sciences Europe*, *32*(1), 1–17. https://doi.org/10.1186/s12302-020-00397-4

Fibæk, C. S. (2017). *Monitoring Compliance with the Common Agricultural Policy* [Aalborg University]. https://projekter.aau.dk/projekter/files/259740642/casperFibaek_thesis.pdf

Fibæk, C. S. (2021). Raster dataset on the area, volume and population of structures in Denmark at ten by ten-meter resolution January 2021. *Mendeley Data*, *1*.

https://doi.org/10.17632/MPKDMFDB8M.1

Fibæk, C. S. (2022a). Area and Population Estimates from Sentinel Data - Ghana and Mediterranean Egypt. *Mendeley Data*, *1*. https://doi.org/10.17632/GF8V525TM6.1

Fibæk, C. S. (2022b). *Buteo Toolbox*. Github. https://github.com/casperfibaek/buteo

Fibæk, C. S. (2022c). *Easy Toolbox GUI*. Github. https://github.com/casperfibaek/toolbox-gui

Fibæk, C. S., & Danielsen, S. K. (2019). *Pilot Study: Classification of Invasive Plant Species Using Sentinel Satellitedata and Other Free Datasets.*

Fibæk, C. S., & Fernandes, A. (2022). *SATF Backend*. GitHub. https://github.com/casperfibaek/satf-api

Fibæk, C. S., Keßler, C., & Arsanjani, J. J. (2021). A multi-sensor approach for characterising human-made structures by estimating area, volume and population based on sentinel data and deep learning. *International Journal of Applied Earth Observation and Geoinformation*, *105*, 102628. https://doi.org/10.1016/j.jag.2021.102628

Fibæk, C. S., Keßler, C., Jokar Arsanjani, J., & Trillo, M. (2022). Globally Applicable Models for Creating Population Estimates from Sentinel Data. *Transactions in GIS*. https://doi.org/10.1111/tgis.12971

Fibæk, C. S., Laufer, H., Keßler, C., & Jokar Arsanjani, J. (2021). Geodata-driven approaches to financial inclusion – Addressing the challenge of proximity. *International Journal of Applied Earth Observation and Geoinformation*, *99*, 102325. https://doi.org/10.1016/J.JAG.2021.102325

Fibæk, C. S., & Trotter, E. (2022). *SATF Frontend*. Github. https://github.com/casperfibaek/satf-frontend

Foley, D. L. (1954). Urban daytime population: A field for demographic-ecological analysis. *Social Forces*, *32*(4), 323–330. https://doi.org/10.2307/2574113

Forster, D., Peachey, S., & Stahl, L. (2013). Mapping Proximity Bringing Products and Services Close Enough to the Poor to Be Meaningfully Usable and Still Keep Them Sustainable for WSBI Partner Banks - Working Paper. *WSBI.*

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., & Hostert, P. (2021). National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, *252*, 112128. https://doi.org/10.1016/j.rse.2020.112128

French Presidency of the Council of the European Union. (2022). *Strasbourg Declaration on the Common values and challenges of European Public Administrations*. https://www.transformation.gouv.fr/files/presse/Declaration_common_values_challenges_european public_administrations_1.pdf

Gao, Y., & Cui, Y. (2020). Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nature Communications*, *11*(1), 1–8. https://doi.org/10.1038/s41467-020-18918-3

GEO. (2017). *Earth Observations in Support of the 2030 Agenda for Sustainable Development*. https://www.earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf

GeoDenmark. (2021). *GeoDanmark Specification 6.0.1*. http://www.geodanmark.nu/Spec6/HTML5/DK/601/StartHer.htm

Gorry, G. A., & Morton, M. S. S. (1971). A framework for management information systems. *Sloan Management Review*, *13*, 55–70. http://mis.njit.edu/ullman/cis465/Articles/gorry.pdf

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., & Cresson, R. (2017). Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, *2*(1), 1–8. https://doi.org/10.1186/s40965-017-0031-6

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2021). Deep Learning for 3D Point Clouds: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(12), 4338–4364. https://doi.org/10.1109/TPAMI.2020.3005434

Haberl, H., Wiedenhofer, D., Schug, F., Frantz, D., Virag, D., Plutzar, C., Gruhler, K., Lederer, J., Schiller, G., Fishman, T., Lanau, M., Gattringer, A., Kemper, T., Liu, G., Tanikawa, H., Van Der Linden, S., & Hostert, P. (2021). High-Resolution Maps of Material Stocks in Buildings and Infrastructures in Austria and Germany. *Environmental Science and Technology*, *55*(5), 3368–3379. https://doi.org/10.1021/acs.est.0c05642

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, *17*(2), 024016. https://doi.org/10.1088/1748-9326/ac4d4f

Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., & Zipf, A. (2021). The evolution of humanitarian mapping within the OpenStreetMap community.

Scientific Reports, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-82404-z

Herweijer, C., Combes, B., Ramchandani, P., & Sidhu, J. (2018). Fourth industrial revolution for the earth: harnessing artificial intelligence for the earth. In *PwC*. Harnessing_Artificial_Intelligence_for_the_Earth_report_2018

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (3rd ed.). OTexts.

Intergovernmental Panel on Climate Change. (2022). Mitigation of Climate Change Summary for Policymakers Climate Change 2022. Working Group III contribution to the Sixth Assessment Report by the IPCC. In *UNEP / WMO*. https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf.

Jasper, P., Luna Hernandez, A., Scott, M., & Tzavidis, N. (2021). How can new technology support better measurement of extreme poverty? Annex A - Methods for producing high-resolution and high-frequency poverty estimates. *DEEP*.

Jutz, S., & Milagro-Pérez, M. P. (2020). Copernicus: The European earth observation programme. *Revista de Teledeteccion*, *2020*(56), v–xi. https://doi.org/10.4995/raet.2020.14346

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., & Chuvieco, E. (2020). Towards delivering on the sustainable development goals using earth observations. *Remote Sensing of Environment*, *247*, 111930. https://doi.org/10.1016/j.rse.2020.111930

Koppel, K., Zalite, K., Voormansik, K., & Jagdhuber, T. (2017). Sensitivity of Sentinel-1 backscatter to characteristics of buildings. *International Journal of Remote Sensing*, *38*(22), 6298–6318. https://doi.org/10.1080/01431161.2017.1353160

Vejledning om beskyttede sten-og jorddiger, (2009). www.kulturarv.dk

Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., & Schindler, K. (2018). Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018

Laufer, H. (2022). The role of data analytics in business development in low-income markets. In *Oxford Policy Management*. https://www.opml.co.uk/projects/savings-frontier

Levy, S. (1984). A Spreadsheet Way of Knowledge. *Harper's*, 1–17.

https://www.wired.com/2014/10/a-spreadsheet-way-of-knowledge/

Li, W., He, C., Fang, J., Zheng, J., Fu, H., & Yu, L. (2019). Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data. *Remote Sensing 2019, Vol. 11, Page 403*, *11*(4), 403. https://doi.org/10.3390/RS11040403

Lillesand, T., Kiefer, R. W., & Jonathan, C. (2016). *Remote Sensing and Image Interpretation* (7th ed.). Wiley.

Lord, J., Thomas, A., Treat, N., Forkin, M., Bain, R., Dulac, P., Behroozi, C. H., Mamutov, T., Fongheiser, J., Kobilansky, N., Washburn, S., Truesdell, C., Lee, C., & Schmaelzle, P. H. (2021). Global potential for harvesting drinking water from air using solar energy. *Nature*, *598*(7882), 611–617. https://doi.org/10.1038/s41586-021-03900-w

Mahboubi, M., Grønbæk, A., Fernandes, A., Fibæk, C. S., & Dabove, P. (2022). Geospatial Analysis of Safe Delivery App Events Based on the Geographically Weighted Regression Tool. *Proceedings of the 2022 Agile Conference*, *3*(47), 1–6. https://doi.org/https://doi.org/10.5194/agile-giss-3-47-2022

Mahendra, D. S. (2006). Financial Inclusion: Issues and Challenges. *Economic and Political Weekly*, *41*(41), 4310–4313. https://www.jstor.org/stable/4418799

Meta. (2022). *Facebook Data For Good High Resolution Population Density Maps*. Data for Good. https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps

Microsoft. (2021). *Open dataset of machine extracted buildings in Uganda and Tanzania*. https://github.com/microsoft/Uganda-Tanzania-Building-Footprints

Microsoft. (2022). *OfficeDev - OfficeJS*. GitHub. https://github.com/OfficeDev/office-js

Mikdashi, M. (2020). Sextarianism: Notes on Studying the Lebanese State. In *The Oxford Handbook of Contemporary Middle Eastern and North African History* (pp. 443–473). https://doi.org/10.1093/oxfordhb/9780199672530.013.24

Mondato. (2021). *Beyond the Last Mile: Reconsidering "Rurality"*. Mondato Blog. https://blog.mondato.com/beyond-the-last-mile-reconsidering-rurality/

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0118571

Nikolakopoulos, K. G. (2008). Comparison of nine fusion techniques for very high resolution data. *Photogrammetric Engineering and Remote Sensing*, *74*(5), 647–659. https://doi.org/10.14358/PERS.74.5.647

NIRAS. (2020). *Savings at the Frontier - Proximity Workshop. Ghana, Accra*.

NIRAS. (2021, December 15). *Increasing the resilience of the North coast and Nile River Delta in Egypt*. https://www.niras.com/projects/increasing-the-resilience-of-the-north-coast-and-nile-river-delta-in-egypt/

Osonde, O., & Welser, W. (2017). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. In *Rand Corporation*. https://doi.org/https://doi.org/10.7249/RR1744

Oswald, M., Grace, J., Urwin, S., & Barnes, G. (2018). Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and Experimentall Proportionality. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3029345

Oxford Policy Management. (2019). *The Savings at the Frontier Project*. https://www.opml.co.uk/projects/savings-frontier

Oxford Policy Management, & NIRAS. (2021). *Dissemination and Use-Case Meetings*.

Paganini, M., Petiteville, I., Ward, S., Dyke, G., Steventon, M., & Harry, J. (2018). *Satellite Earth Observations in Support of the Sustainable Development Goals*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

Peachey, S., & Mutiso, M. (2019). Moving Proximity From Critical Issue to Addressable Challenge: Possible Approaches and Tools. Main Report. *Oxford Policy Management*. http://www.opml.co.uk/projects/savings-frontier

Porteous, D. (2006). The Enabling Environment for Mobile Banking in Africa. In *Bankable Frontiers Associates Boston USA* (Issue April). www.bankablefrontier.com

Qt Company, & Qt Project. (2022). *Qt*. https://www.qt.io/

Raković, L. (2019). A framework for managing spreadsheet-based end user applications. *International Journal of Management and Decision Making*, *18*(1), 76–92. https://doi.org/10.1504/ijmdm.2019.096695

Ranchordás, S. (2022). Empathy in the Digital Administrative State. *Duke Law Journal*, *71*(6), 1341–1389. https://doi.org/10.2139/ssrn.3946487

Renee Pelas, H. (2011). *Tourism Development in Cancun, Mexico: An Analysis of State-Directed Tourism Initiatives in a Developing Nation*. Georgetown University.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Schug, F., Frantz, D., van der Linden, S., & Hostert, P. (2021). Gridded population mapping for Germany based on building density, height and type from Earth Observation data using census disaggregation and bottom-up estimates. *PLoS ONE*, *16*(3 March), e0249044. https://doi.org/10.1371/journal.pone.0249044

Schultz, M., Voss, J., Auer, M., Carter, S., & Zipf, A. (2017). Open land cover from OpenStreetMap and remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *63*, 206–213. https://doi.org/10.1016/j.jag.2017.07.014

Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J. H., Shoeibi, A., Jafari, M., Hussain, S., Sani, Z. A., Hasanzadeh, F., Khozeimeh, F., Khosravi, A., Nahavandi, S., Panahiazar, M., Zare, A., Islam, S. M. S., & Acharya, U. R. (2021). Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomedical Signal Processing and Control*, *68*, 102622. https://doi.org/10.1016/j.bspc.2021.102622

Shi, G., Jiang, N., & Yao, L. (2018). Land use and cover change during the rapid economic growth period from 1990 to 2010: A case study of Shanghai. *Sustainability (Switzerland)*, *10*(2), 426. https://doi.org/10.3390/su10020426

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y. S. E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., & Quinn, J. (2021). Continental-Scale Building Detection from High Resolution Satellite Imagery. *ArXiv*. https://arxiv.org/abs/2107.12283v2

Sprague, R. H. (1980). A framework for the development of decision support systems. *MIS Quarterly: Management Information Systems*, *4*(4), 1–26.

https://doi.org/10.2307/248957

Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y. A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*, *14*(127). https://doi.org/10.1098/rsif.2016.0690

Steele, L. (2019, January 3). How Mexico built Cancun from scratch. *Yucatan Magazine*. https://yucatanmagazine.com/how-mexico-built-cancun-from-scratch/

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 4278–4284. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*(1), 1–4. https://doi.org/10.1038/sdata.2017.4

Thomsen, C. F., Barrie, A. M. F., Boas, I. M., Lund, S., Sørensen, B. L., Oljira, F. G., & Tersbøl, B. P. (2019). Health workers' experiences with the Safe Delivery App in West Wollega Zone, Ethiopia: A qualitative study. *Reproductive Health*, *16*(1), 1–11. https://doi.org/10.1186/s12978-019-0725-6

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B., & Dang, H.-A. H. (2017). Mapping the World Population One Building at a Time. In *Mapping the World Population One Building at a Time*. World Bank, Washington, DC. https://doi.org/10.1596/33700

Trask, A. W. (2019). *Grokking Deep Learning* (1st ed.). Manning.

Trotter, E. F. L., Fernandes, A. C. M., Fibæk, C. S., & Keßler, C. (2022). Machine learning for automatic detection of historic stone walls using LiDAR data. *International Journal of Remote Sensing*, *43*(6), 2185–2211. https://doi.org/10.1080/01431161.2022.2057206

United Nations. (2008). Principles and recommendations for population and housing censuses - Revision 2. *Statistical Papers*, *3*(March), 420. http://unstats.un.org/unsd

United Nations. (2014). *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. https://www.undatarevolution.org/

United Nations. (2017). Factsheet: people and oceans. In *The Ocean Conference*

(Issues 5-9 June 2017). United Nations. oceanconference.un.org

United Nations Statistical Commission. (2020). Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development Goals and targets (from the 2030 Agenda for Sustainable Development) Indicators. *IAEG-SDGs*, 21. https://lirias.kuleuven.be/retrieve/572545

Varghese, G., & Viswanathan, L. (2018). Financial Inclusion: Opportunities, Issues and Challenges. *Theoretical Economics Letters*, *08*(11), 1935–1942. https://doi.org/10.4236/tel.2018.811126

Vincent, O., & Folorunso, O. (2009). A Descriptive Algorithm for Sobel Image Edge Detection. *Proceedings of the 2009 InSITE Conference*. https://doi.org/10.28945/3351

Vinuesa, R., & Sirmacek, B. (2021). Interpretable deep-learning models to help achieve the Sustainable Development Goals. *Nature Machine Intelligence*, *3*(11), 926. https://doi.org/10.1038/s42256-021-00414-y

Wang, Y. (2021). When artificial intelligence meets educational leaders' data-informed decision-making: A cautionary tale. *Studies in Educational Evaluation*, *69*, 100872. https://doi.org/10.1016/j.stueduc.2020.100872

Warmerdam, F. (2008). The Geospatial Data Abstraction Library. *Open Source Approaches in Spatial Data Handling*, 87–104. https://doi.org/10.1007/978-3-540-74831-1_5

Whitcraft, A. K., Becker-Reshef, I., & Justice, C. O. (2015). A framework for defining spatially explicit earth observation requirements for a global agricultural monitoring initiative (GEOGLAM). *Remote Sensing*, *7*(2), 1461–1481. https://doi.org/10.3390/rs70201461

Woods, M. (2010). Rural. In *Rural* (1st ed.). Taylor and Francis. https://doi.org/10.4324/9780203844304

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., … Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, *225*, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

Wunderlich, G. S. (2016). *Rationalizing Rural Area Classifications for the Economic Research Service*. National Academies Press. https://doi.org/10.17226/21843

ISSN (online): 2446-1628
ISBN (online): 978-87-7573-818-2

AALBORG UNIVERSITY PRESS