MSc thesis topics in Geoinformatics – spatial big data

Meng Lu

meng.lu@uni-bayreuth.de

Near real-time change monitoring from remote sensing image time series

Near real-time change monitoring from time series continuously monitor change when new data enters the data stream. As remote sensing time series are becoming dense, near real-time change monitoring from remote sensing image time series has opened immense opportunities in environmental monitoring. For example, an automatic deforestation monitoring system significantly contributes to forest conservation from relevant environmental and political agencies. Time series structural change methods based on cusum (cumulative sum) statistics have been developed to automatically detect forest change from optical satellite image time series. However, the method has difficulties in separating between man-made and natural change (e.g., drought) and requires the modelling of seasonality, which is complex in many forest systems. Also, it is difficult to integrate multidimensional information from space, spectral bands, as well as multiple sensor data into a 1D time series analysis. Deep learning methods have dominated remote sensing due to their great predictive power, and the fact that Earth observations have been growing tremendously in diversity and volumes. Recurrent neural networks and their relatives (e.g. Long short-term memory (LSTM) [1]) are state-of-the-art in time series forecasting and change monitoring [2] but they have been rarely applied in environmental modelling. This study focuses on evaluating deep learning-based change monitoring methods (e.g. recurrent neural network and its relatives) and compare them with statistics methods. *Novel application cases and own datasets are welcomed* but is not a prerequisite. Sentinel 2 and other open archives will be used if the study case is in deforestation monitoring.

Keywords: deep learning, time series analysis, recurrent neural networks, change detection, remote sensing

Reference:

[1] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.

[2] Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K. and Funaya, K., 2016, October. Robust online time series prediction with recurrent neural networks. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 816-825). Ieee.

Deep learning for intertidal-flats mapping from remote sensing imagery

Spatiotemporal geomorphological mapping of intertidal areas is essential for understanding system dynamics and provides information for ecological conservation and management. Ecological quality of intertidal areas is important because of the European Water Framework Directive and as they are designated as Natura 2000 areas, which is the implementation of the European birds directive and the European habitat directive. Mapping the geomorphology of intertidal areas is a considerable challenge mainly because spectral differences are oftentimes relatively small while transitions between geomorphological units are oftentimes gradual. Also, the intertidal areas are highly dynamic. Surface water, saltmarsh, and tidal flats are relatively simple to distinguish but considerable challenges remain for distinguishing between different types of tidal flats, specifically, low and high dynamic shoal flats, sandy and silty low dynamic flats, and mega-ripple areas. The challenge is reflected in recent studies using rule-based OBIA (Object-based Image Analysis) methods to classify aerial imagery at 0.25 m resolution with spectral bands of red, blue, and NIR. In addition, monitoring and modelling the coastal tidal-flat dynamics over time is essential for coastal tidal-flat conservation and understanding, but has not been addressed.

In recent decades, deep convolutional neural networks, as a powerful representation learning method, have brought a breakthrough in information extraction from imagery. Predicting each pixel of images to a category is called segmentation. This study aims to apply and develop semantic segmentation neural networks to address challenges inter-tidal area classification. The results will also be compared with rule-based and tree-based OBIA methods.

Keywords: deep learning, classification, natural environment, remote sensing, coastal geomorphology

Image-based spatiotemporal air pollution with deep learning methods

Poor ambient air quality represents one of the largest environmental risks to public health. Long-term exposure to air pollution associates with chronic respiratory and cardiovascular diseases. Spatial-temporal air quality prediction is of vital importance in understanding the health effects of air pollution and for making scientific recommendations. Ground sensor networks are usually sparse and incidental, especially in developing countries, and can have a low resolution in space and/or time. But air pollutants such as NO2 are highly dynamic spatial-temporally. Statistical mapping methods of air pollution commonly use relationships between GIS predictors (e.g., population) and covariates (regression) but it is difficult to find all the relevant predictors. This project invests in novel deep learning-based computer vision methods to automatically extract features from remote sensing imagery and thematic maps for air pollution mapping.

Keywords: deep learning, air pollution, remote sensing

OpenStreetMap and deep learning for road information completion and extraction

In many middle and low-income countries, air pollution monitor networks are deficient or non-existing, but in these countries, people are the most vulnerable to air pollution, with young children suffer the most. The idea is to borrow information from countries where relatively dense ground monitors are available, and integrate information from satellite measurements and geospatial predictors, to give an estimation of global air quality. Global air quality has a strong and complex relationship with transport networks, capturing this relationship is a key in spatial prediction of traffic-related pollutants (e.g. NO2).

So far, OpenStreetMaps (OSM) provides the most open-source transport network, as in different types of tracks (e.g. roads, rails). A key challenge is that the transport network provided by OSM is incomplete in many countries, such as in China and African countries. High-Resolution satellite imagery (e.g. worldview2) and machine learning (particularly deep learning neural networks) are promising techniques to complement the OSM, and evaluate the consequences of directly using the OSM with missing roads to predict air pollution. On the other hand, OSM provides rich labels to perform supervised machine learning algorithms (e.g. to train a neural network).

This project expects to develop or apply advanced deep learning algorithms to extract transport networks from satellite imagery, using OSM. A software is developed to interactively complete OSM using satellite imagery. As the research matures, you will collaborate with environmental modelers to evaluate the consequences of using OSM for global air pollution mapping.

Keywords: OpenStreetMaps, deep learning, remote sensing, road extraction, software development

Spatially varying coefficients modelling in large-scale air pollution mapping

Atmosphere behaves less coherently over larger areas with respect to the dispersion and emission processes. Same traffic loads in a certain area may lead to different NO2 concentration levels due to different types of fuel (e.g. ethanol vs. petrol), car (e.g. electric car), and filter systems used in cars. The Tropomi measured column density relates to surface concentrations differently under different meteorological conditions. The relationship between NO2 and wind also differs under different city morphology. An effective city topology may contribute to efficiently modelling Spatially Heterogeneous response-covariate Relationships (SHR). A commonly-used method is GWR (Geographically Weighted Regression) but GWR does not have a formal uncertainty quantification. Bayesian SVC models coefficients as GP and fully quantify uncertainty in predictions but a computationally efficient solution is needed for modelling with large data volumes. The task of this study is to develop GWR and Bayesian SVC methods in modeling SHR and compare the models.

Keywords: geostatistics, spatial varying coefficients, Bayesian modeling, air pollution

Reading Material:
Gelfand, Alan E., Hyon-Jung Kim, C. F. Sirmans, and Sudipto Banerjee. "Spatial Modeling with Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98, no. 462 (2003): 387-96. Accessed August 2, 2021. http://www.jstor.org/stable/30045248.

Agent-based human space-time activity modeling for air pollution exposure assessment

Conventional air pollution exposure assessment methods assess air pollution exposure as air pollutant concentrations at front door locations. This approach ignores human activity patterns and consequently may lead to over- or under-estimation of air pollution exposure. Assessing air pollution exposure considering activity patterns of individual persons remains to be a challenge as the detailed working location information and detailed activity pattern of the residents are commonly unknown when large numbers of individuals need to be considered. This study aims at integrating mixed-effect and machine learning models in ABM (agent-based modelling) methods to predict human activity patterns for exposure assessment. The methodology will be tested and evaluated for Switzerland using national travel survey data. This study is closely related to our ongoing *Mobi-Air* project with Swiss TPH.

Tasks:

1) Regression modeling of travel survey data, finding relationships between e.g. travel mode and travel distances, ethnics.
2) ABM modeling of human space-time activity. Comparing the model results for different population groups.
3) Linking to temporal air pollution maps for exposure assessment.

Keywords: regression analysis, agent-based modeling, space-time activity, exposure assessment, air pollution, environmental health

Reading material

Meng Lu, Oliver Schmitz, Ilonca Vaartjes, Derek Karssenberg, Activity-based air pollution exposure assessment: Differences between homemakers and cycling commuters, Health & Place, Volume 60, 2019, 102233, ISSN 1353-8292,
https://doi.org/10.1016/j.healthplace.2019.102233

Environment - mobility interaction using social media and remote sensing data

Would you choose a greener route to the University when the other routing conditions are the same? Human space-time behaviours adapt to changes in our environment. However, environmental effects on human activities are commonly ignored in mobility modelling, which is a key component in assessing air pollution exposure. Geo-coded social media data are can provide us a huge amount of information about space-time activities of citizens. High-resolution environmental variables are also becoming more available with advancements in remote sensing techniques and the trend of open science. These data can be used for understanding the relationships between human space-time activities and our environment and contribute to more precise modelling of human space-time activity and subsequently exposure assessment. The development is also important for studying the unintended consequences of net-zero actions rolled out in Europe. Through future scenario planning, we aim at understanding the socio-economic impacts of electric vehicles, clean air zone and traffic rerouting.

Tasks:

1) Process geocoded social-media data (e.g. twitter, facebook) and extract information (i.e. the activity and environment information).
2) Process remote sensing imagery (e.g. sentinel 2, worldview 3) and extract environment information (e.g. NDVI, building density). Existing data products can also be used.
3) Process meteorological data (e.g. precipitation, temperature)
4) Finding relationships between mobility and our environment.
5) Using the relationship to develop an agent-based model for mobility modelling and exposure assessment.
6) Scenario planning of electric cars, traffic-rerouting, clean air zone, and understand the changes in air pollution exposure.

Keywords: remote sensing, social media, regression analysis, agent-based modeling, exposure assessment, air pollution, health

Additional information:

The study area you can define yourself for your region of interest. If you do not have a preference, it would be good to choose Netherlands, or Germany, or several cities from these two countries. You will get more data if you choose a bigger study area.

Twitter data are available but you need to register an account and get permission. See below.

https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/twitter-data-in-python/

https://cran.r-project.org/web/packages/rtweet/vignettes/auth.html

Example codes and data

https://www.kaggle.com/kazanova/sentiment140/code

Reading material:
Social-media data for urban sustainability
https://static1.squarespace.com/static/552ec5f5e4b07754ed72c4d2/t/5be1dae270a6adf7bb5a6609/1541528292607/ilieva+and+mcphearson_+nature+sustainabilty+2018.pdf