Workshop Proposal
Harnessing Earth Observations and Data Science to Evaluate Socio-Environmental Issues

Basic Information

- Facilitators: Raghavan Narayanan (European Bank for Reconstruction and Development, EBRD), Dominik Naeher (University of Göttingen), Alexander Stepanov (EBRD), and Zélie Marçais (EBRD).
- Level: Introductory.
- Prerequisite for participants: None.
- Target number of participants: 20 to 40.

Contents

This workshop aims to demonstrate how recent advances in data science, earth observations, and computational resources could work together to facilitate impact evaluation of interventions addressing complex environmental and societal problems. We will discuss when and how evaluators could benefit from using these new opportunities with a particular focus on applying geospatial toolkit to identify the effects of development interventions. The participant evaluators will go through the entire chain of steps from data gathering to analysis to results visualisation to produce powerful near-real-time insights and present them in an intuitive and appealing way. As a working example throughout the workshop, we will evaluate interventions aiming to improve the quality of urban environment, including pressing topics like air pollution, land use, transport, economic activity, etc.

The workshop will consist of four sessions:

- **Session 1: Introduction into geospatial data processing and analysis.** This session will start with the discussion of geospatial data types available via earth observations, modelling, and crowdsourcing and ways to process them for capturing of environmental and economic changes. It will clarify when geospatial analysis is mostly applicable to evaluations.
The session will proceed with introduction into causal interference analytics, which can be used to estimate the effects of an intervention. It will particularly highlight methods applicable to geospatial data such as difference-in-differences and regression discontinuity design.
- **Session 2: Monitoring and benchmarking performance.** The section will follow up to show how geospatial data could be deployed before and during an intervention to assess its relevance, monitor its progress, and capture early results.

  As an example, we will leverage publicly available sources like Google Earth Engine and OpenStreetMap to access data relevant for impact evaluation of interventions targeting improvement of various components of urban environment. We will demonstrate Python code to extract and process data from these sources. Then we will build a composite index of urban liveability across multiple relevant factors. This index will be used to discuss cities' relative performance over time and against comparators. Using the frontier analysis, we will estimate cities' potential for improvement.

- **Session 3: Attributing observed change to the intervention.** The section will proceed to demonstrate power of geospatial data and methods of data science to identify the causal effects of an intervention. Using one of the causal inference methods discussed in Section 1, we will estimate the effect of a large-scale real-life multilateral intervention on city-wide outcomes.

- **Session 4: Communicating the results.** The session will present the results obtained at the previous steps in an intuitive and appealing way, targeting a non-technical audience and suggesting ways for evaluators to communicate their findings better. We will demonstrate how to create interactive online maps using Python and Google Earth Engine tools.

Delivery Method

The facilitators will take multiple actions to ensure dynamic and interactive nature of the workshop:

1. **Preparatory materials**: several weeks before the event, the participants will receive suggestions of free tutorials as an introduction to geospatial analysis in Python. Introductory reading to the causal inference will be provided as well.
2. **Interactive programming**: participants will receive a Google Colab notebook defining the structure of the target application. During the workshop, they will be guided through the critical parts of the code. The complete solution will be provided under the workshop.
3. **Gamification**: participants will split into groups representing various types of urban dwellers with different, often competing, interests. In Session 2, each group will identify its own set of factors that define a "liveable" city[*] and then use these factors to calculate the composite index and then conduct causal analysis in Session 3. The exercise will demonstrate how the choice and availability of input data may affect evaluation results.

Anticipated Outcomes

This workshop intends to have two types of outcomes, both in the short run and the long run.

In the short run, the participant evaluators are expected to:

- Receive introduction into the area of geospatial analytics and learn about various types of geospatial data.
- Understand econometric methods of utilising geospatial data in evaluations, which cover multiple stages of intervention implementation.
- Gain hands-on experience of using popular tools such as Python and Google Earth Engine to access, analyse, and visualize geospatial data.

In the long run, we believe this workshop will open up thinking to analyse environmental and social compliance not just in the evaluation context (or ex-post) but also in the ex-ante project preparation and evaluability assessment stages, and ultimately shape the way large-scale development interventions can be designed, targeted, shaped and evaluated over time.

---

[*] This discussion can take place online before the workshop, so the facilitators have time to prepare necessary input data.