

skills	data engineering, distributed systems, stochastic simulation, geospatial visualization, econometrics, timeseries analysis, Bayesian inference, machine learning
programming	Python (numpy, pandas, dask, matplotlib, sklearn, pymc, tensorflow, pytorch, opencv), Julia, Haskell, R (tidyverse, ggplot, rstan, glm), Java (guava, dropwizard, dagger), JavaScript (react, bootstrap), C/C++, Go, Mathematica, Stata, MATLAB, shell (bash, zsh)
systems	Spark, Kubernetes, Elasticsearch, Cassandra, Postgres/PostGIS, *nix, AWS, GCP, Docker, Slurm, L ^A T _E X, Git, ArcGIS, QGIS, Google Earth Engine
languages	Spanish, Marathi, French, Hindi/Urdu, German

EXPERIENCE

- Head of Data Science, COVID International Working Group** Jun 2020-Aug 2021
- Advised national and regional governments in India and Indonesia on COVID-19 lockdowns, economic recovery, and vaccine distribution policies.
 - Developed novel, computationally-efficient Bayesian estimation models for key epidemiological metrics.
 - Led team of data scientists to build cloud-hosted data pipeline to deliver daily-updated insights to policymakers.
- Principal Research Engineer, Mansueto Institute for Urban Innovation** Jan 2019-Jun 2020
- Led development and data analysis for the Million Neighborhoods Project, the first global map identifying slums through topological and graph-theoretic models of infrastructure access.
 - Created algorithms for extracting city block and cadastral parcel geometry from open-source geospatial data.
 - Responsible for scaling data analysis and processing to terabyte-scale and for onboarding new researchers.
- Forward-Deployed Engineering Lead, Palantir Technologies** Aug 2014-Aug 2018
- Developed new algorithms for detecting money-laundering and cybersecurity breaches used by financial institutions to prevent crimes and comply with international regulations.
 - Architected a geospatial risk analysis tool for natural catastrophe insurance pricing and led user interviews with underwriters in Latin America and Europe as tool was developed and rolled out.
 - Implemented custom database capabilities, including: statistically-inferred relational keys across datasets, in-memory subquery caching for distributed databases, and efficient object-relational mapping tools.
 - Oversaw technology deployment for multiple humanitarian engagements with the American Red Cross and Team Rubicon.
 - Led application engineering teams of up to 15 people for clients in finance, healthcare, and manufacturing.
 - Redesigned engineering interview and hiring process for Singapore, Paris, and London offices.

EDUCATION

- PhD, Information Systems & Management** 2021-
University of California, Berkeley - School of Information
advisor: Joshua Blumenstock
minor: Development Engineering
supported by: NSF Grant DGE-2125913 (Digital Transformation of Development)
- MS, Computational Analysis & Public Policy** 2018-2020
University of Chicago - Harris School of Public Policy
minor: Financial Analysis & Policy
- BS, Electrical Engineering & Computer Science, Materials Science & Engineering** 2010-2014
University of California, Berkeley - College of Engineering
minor: Physics

CONFERENCE PROCEEDINGS

Estimating Household-Level Economic Characteristics from High-Resolution Satellite Imagery

Soman, S., Constenla-Villoslada, S., Aiken, E., & Blumenstock, J.E. (2024).

World Bank Land Conference - Research Track. Washington, D.C., USA.

Can Strategic Data Collection Improve the Performance of Poverty Prediction Models?

★ spotlight presentation (top 10% of submissions)

Soman, S., Aiken, E., Rolf, E., & Blumenstock, J.E. (2023).

Eleventh International Conference on Learning Representations (ICLR),

4th Workshop on Practical Machine Learning for Developing Countries. Kigali, Rwanda. [10.48550/arXiv.2211.08735](https://arxiv.org/abs/10.48550/arXiv.2211.08735)

PEER-REVIEWED PUBLICATIONS

A Hierarchical Bayesian Model for Estimating Age-specific COVID-19 Infection Fatality Rates in Developing Countries

Pugh, S., Levin, A., Meyerowitz-Katz G., **Soman, S.**, Owusu-Boaitey, N., Zwi, B., Malani, A., & Fosdick, B. (2024).

Statistics in Medicine [10.1002/sim.10259](https://doi.org/10.1002/sim.10259)

Assessing the Burden of COVID-19 in Developing Countries: Systematic Review, Meta-Analysis, & Public Policy Implications

Levin, A., Owusu-Boaitey, N., Pugh, S., Fosdick, B., Zwi, A., Malani, A., **Soman, S.**, et al. (2022).

BMJ Global Health 7, no. 5 (2022): e008477 [10.1136/bmjgh-2022-008477](https://doi.org/10.1136/bmjgh-2022-008477)

Epimargin: A Toolkit for Epidemiological Estimation, Prediction, and Policy Evaluation

Soman, S., Loftus, C., Buschbach, S., Phadnis, M., & Bettencourt, L. (2021).

Journal of Open Source Software, 6(65), 3464 [10.21105/joss.03464](https://doi.org/10.21105/joss.03464)

Worldwide Detection of Informal Settlements via Topological Analysis of Crowdsourced Digital Maps

Soman, S., Beukes, A., Nederhood, C., Marchio, N., & Bettencourt, L. (2020).

ISPRS Int. J. Geo-Inf., 9 (10) [10.3390/ijgi9110685](https://doi.org/10.3390/ijgi9110685)

WORKING PAPERS

Vaccine Allocation Priorities Using Disease Surveillance and Economic Data

Malani, A., **Soman, S.**, Ramachandran, S., Chen, A., Lakdawalla, D. (2022).

NBER Working Paper, (No. w29682) [10.3386/w29682](https://doi.org/10.3386/w29682)

Systems Architecture for Real-Time Epidemiological Prediction and Control

Bettencourt, L. & **Soman, S.** (2020).

Mansueto Institute for Urban Innovation Research Paper, (24) [10.2139/ssrn.3748704](https://doi.org/10.2139/ssrn.3748704)

Adaptive Control of COVID-19 Outbreaks in India: Local, Gradual, and Trigger-Based Exit Paths from Lockdown

Malani, A., **Soman, S.**, Asher, S., Novosad, P., Imbert, C., Tandel, V., et al. (2020).

NBER Working Paper, (No. w27532) [10.3386/w27532](https://doi.org/10.3386/w27532)

PATENTS

Systems and Methods for Annotating Datasets

Javitt, G., Szufita, S., **Soman, S.**, Pandey, H., Dhulipalla, S., Shekhawat, V.

US Patent #11,176,116. Granted: 16 November, 2021.

Data Item Aggregate Probability Analysis System

Soman, S., Hoffman, D., Al Khafaji, S., Kowalik, J., Sanzovo, P., Punukollu G.

US Patent #10,691,756. Granted: 6 February, 2018.

TEACHING

UC Berkeley - INFO 206A: Introduction to Programming & Computation (Fall 2024, head TA)

UC Berkeley - INFO 206B: Introduction to Data Structures & Analytics (Fall 2024, head TA)