

EXPERTISE

COMPETENCIES	data pipelining, distributed systems, stochastic simulation, geospatial visualization, timeseries analysis, econometrics, Bayesian inference, machine learning
PROGRAMMING	Python, R, Java, Haskell, C/C++, Go, Stata, Mathematica, MATLAB, JavaScript
SYSTEMS	Spark, Kubernetes, ElasticSearch, Cassandra, Postgres/PostGIS, *nix, AWS, GCP
LANGUAGES	Spanish, Marathi, French, Hindi, German

EXPERIENCE

COVID International Working Group Head of Data Science	JUN 2020 - PRESENT CHICAGO, IL
• 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.	
Mansueto Institute for Urban Innovation Research Engineer	JAN 2019 - JUN 2020 CHICAGO, IL
• 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.	
Palantir Technologies Forward-Deployed Engineering Lead	AUG 2014 - AUG 2018 PALO ALTO, CA / SINGAPORE / PARIS / LONDON
• 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 & Europe as product 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.	
• Led teams of up to 15 people providing custom development for finance, healthcare, and manufacturing clients.	
• Redesigned engineering interview and hiring process for Singapore, Paris, and London offices.	
Flutter (acquired by Google) Software Engineer	MAY 2012 - MAY 2013 MENLO PARK, CA
• Built toolset to measure histogram-of-oriented-gradients (HOG) computer vision algorithm performance.	
• Deployed regression testing framework on Hadoop to reduce product testing time by 90%.	
• Added image compression to training pipeline to reduce support vector machine (SVM) training time.	

EDUCATION

M.S. University of Chicago (Harris School of Public Policy) Computational Analysis & Public Policy	2018 - 2020
B.S. University of California, Berkeley (College of Engineering) Electrical Engineering & Computer Science, Materials Science & Engineering	2010 - 2014

PUBLICATIONS

Systems Architecture for Real-Time Epidemiological Prediction and Control

preprint
[[link](#)]

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

Mansueto Institute for Urban Innovation Research Paper, (26).

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

peer-reviewed
+ published
[[link](#)]

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

ISPRS Int. J. Geo-Inf., 9(10).

Adaptive Control of COVID-19 Outbreaks in India:

working paper
[[link](#)]

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).

PATENTS

Data Item Aggregate Probability Analysis System

[[#US10691756B2](#)]

System for selection and filtering of datasets by geospatial index and by configurable risk score (lead inventor).

Systems and Methods for Annotating Datasets

[[#US20190179936A1](#)]

Application of the Flajolet-Martin algorithm to identify potential database joins in data warehousing systems.