Jeison Mesa

Statistician and data scientist. Time series, state-space models and machine learning in R and Python.
Colombia · jeikosd.github.io · github.com/Jeikosd · LinkedIn · Google Scholar

SUMMARY

Statistician and data scientist with more than ten years in R and Python. Five years at CIAT on state-space and time series models behind national seasonal forecasts and a shared 2017 UN award; since 2021 at Globant, taking models to production for clients, now designing LLM agent systems. Seven journal articles as co-author; co-director of an undergraduate thesis; university teaching.

EXPERIENCE

AI Engineer, Globant (client: Disney) Dec 2025 – Present · Remote, Colombia

- Work on agent design for multi-agent LLM systems, including prompt engineering and the contracts between agents.
- Designed and implemented a multi-agent system that reads a codebase and writes its documentation. Led the prompt engineering and set the split between deterministic and LLM steps, and added schema validation of agent outputs so that a bad output fails loudly instead of slipping through.
- Built a conversational intake agent for internal product requests. It pushes back on vague or solution-first input and flags overlap across teams.
- Built the ETL and LLM engine behind scheduled reports. Live data from project-management tools goes into one reporting model under extraction rules that keep field-level provenance, with pre-distribution checks before anything is sent out.

Tools: Python, LangGraph, Azure OpenAI, REST APIs, Docker, Git, CI/CD.

Data Scientist, Globant May 2021 – Dec 2025 · Remote, Colombia

Sports-industry client (Aug 2023 – Dec 2025)

- Built statistical and machine learning models for sports-industry products, mainly pricing and match analytics.
- Worked with time series, econometric methods and embeddings on historical sales and event data.

Augmented Staffing, internal to Globant (Feb 2022 – Jan 2023)

- Owned the end-to-end forecasting and clustering of internal staffing tickets, used to plan recruiting.
- Built a semantic search and recommendation tool that matched candidates to tickets on text embeddings, then added a seniority estimation model in PyTorch with an ordinal loss to improve the match. Pipelines ran on Kubeflow.

Media and telecommunications client (May 2021 – Dec 2021)

- Led the team on subscriber segmentation and directed the data science work on the churn model and the marketing campaigns.
- The churn model combined platform usage with sentiment from text, using custom neural networks in PyTorch. Its feature importance fed the segments.
- Retention and acquisition email campaigns drew on an econometric analysis of subscription pricing and a market basket analysis. Pipelines ran on AWS SageMaker, with data from Athena and S3.

Professor (part-time), Universidad del Valle and Universidad Católica Aug 2020 – Jun 2021 · Cali, Colombia

- Taught Time Series Analysis and Probability at Universidad del Valle, and Fundamentals of Statistics I and II at Universidad Católica.

Data Scientist, CIAT (International Center for Tropical Agriculture) Jan 2015 – Mar 2020 · Cali, Colombia

- Modelled climate variables and their effect on crops and other productive sectors with probabilistic time series models, state-space models and spatio-temporal stochastic processes.
- Developed the statistical models behind the AClimate Colombia seasonal forecast platform (paper).
- Studied near-term climate change impacts on major crops in Latin America and the Caribbean for the IDB and USAID. Set up multiple DSSAT and ORYZA runs from R for the region's five main production systems and combined them with machine learning and geostatistics.
- Mapped climate service hotspots worldwide for USAID (2017–2018), meaning places where agro-climatic information could help close yield gaps. GAMs and other machine learning, the whole process in R (AgMetGaps).
- Built an index with spatial machine learning models to support the breeding of climate-smart crops.
- Orinoco regional climate change plan (Meta, Casanare, Vichada and Arauca, 2015–2016): climate clusters, geospatial models automated in R, and yield simulation for eleven production systems in the Llanos.

Student Intern, CIAT Apr – Oct 2014 · Cali, Colombia

- Classified production systems with cluster analysis for Colombia's Fondo Adaptación, with CORPOICA, as part of a vulnerability analysis for areas hit by the rainy season.

Earlier work 2010 – 2013

Statistical consultant for engineering theses and statistical assistant at Universidad del Valle; taught statistics in R to research faculty at Universidad Santiago de Cali.

SKILLS

Text and LLMs: agents and multi-agent systems (LangGraph, LangChain), prompt engineering, embeddings and vector search, topic models, Hugging Face, OpenAI and Azure OpenAI, LangSmith, PromptLayer.

Time series and statistics: state-space models and Kalman filters, ARIMA, GARCH, long-memory models, Gaussian processes, regression and mixed models, GAMs, experimental design, bootstrap, geostatistics.

Machine learning: clustering, dimensionality reduction (PCA, kernel PCA, t-SNE, UMAP), gradient boosting (XGBoost), deep learning (PyTorch), anomaly detection, scikit-learn.

Programming: R (advanced), Python, SQL.

Cloud and data: PySpark, sparklyr, Polars, pandas, Parquet, raster and satellite data, Databricks, AWS (SageMaker, Athena, S3), GCP (Vertex AI, BigQuery), Azure (OpenAI, Functions, DevOps).

MLOps and engineering: Git, GitHub, GitLab, Docker, Kubeflow, CI/CD, DVC, Datadog, Linux, Quarto, R Markdown, LaTeX.

Domain: climate impact modelling, crop simulation (DSSAT, ORYZA), churn, customer segmentation, pricing.

Languages: Spanish (native), English (B2, upper intermediate).

EDUCATION

B.Sc. in Statistics, Universidad del Valle, Cali (2008 – 2014). Thesis: *Comparison of estimation methods for the fractional differencing parameter in conditionally heteroscedastic long-memory time series.*

THESIS SUPERVISION

Co-director of an undergraduate thesis in Statistics, Universidad del Valle (2023), *Spectral analysis of Colombian genres: La Cumbia* (in Spanish). Open access.

TALKS

Principal speaker on the thesis at the ESPOL statistics and informatics conference (Ecuador, 2014) and at a Universidad Nacional de Colombia and Universidad del Valle summer school (2013). Co-author at the AGU Fall Meeting (2016).

AWARDS

UNFCCC Momentum for Change award, 2017, as part of the CIAT climate services team. **Best Research Outcome**, CIAT, 2017.

SELECTED PUBLICATIONS

Citation counts are on Google Scholar. Full list on jeikosd.github.io/publications.

- Valencia JB, Guryanov VV, **Mesa-Diez J**, et al. (2024). Predictive assessment of climate change impact on water yield in the Meta River Basin, Colombia: an InVEST model application. *Hydrology*, 11(2), 25. DOI
- Valencia JB, Guryanov VV, **Mesa-Diez J**, et al. (2023). Assessing the effectiveness of the use of the InVEST annual water yield model for the rivers of Colombia. *Water*, 15(8), 1617. DOI
- Ordoñez L, Vallejo E, Amariles D, **Mesa J**, et al. (2022). Applying agroclimatic seasonal forecasts to improve rainfed maize agronomic management in Colombia. *Climate Services*, 28, 100333. DOI
- Ochoa-Muñoz AF, Peña-Torres JA, García-Bermúdez CE, Mosquera-Muñoz KF, **Mesa-Diez J** (2022). On characterization of sensory data in presence of missing values: the case of sensory coffee quality assessment. *Ingeniare*, 30(3), 564–573. DOI
- Valencia JB, **Mesa J**, León JG, Madriñán S, Cortés AJ (2020). Climate vulnerability assessment of the *Espeletia* complex on páramo sky islands in the northern Andes. *Frontiers in Ecology and Evolution*, 8, 565708. DOI
- Sotelo S, Guevara E, Llanos-Herrera L, et al., **Mesa J**, et al. (2020). Pronosticos AClimateColombia: a system for the provision of information for climate risk reduction in Colombia. *Computers and Electronics in Agriculture*, 174, 105486. DOI