Elías D. Niño-Ruiz, Ph.D.

COMPUTATIONAL SCIENTIST · ASSOCIATE PROFESSOR · DATA & AI CONSULTANT

elias.d.nino@gmail.com enino@uninorte.edu.co +57 300 539 7765 Barranquilla, Colombia

https://enino84.github.io/

Profile

Computational scientist working across two tracks that feed each other. In research, I build ensemble-based data assimilation methods that correct numerical forecasts using noisy observations, with 70+ peer-reviewed works and open-source packages behind them. In industry, I work as a consultant in data engineering, data science and applied AI: production data platforms on AWS, GCP and Azure, forecasting models, and LLM and agent systems shipped end to end.

Full publication record, software and teaching material: https://enino84.github.io/

Education

2011 – 2015 **Ph.D. in Computer Science and Applications**
Virginia Polytechnic Institute and State University, USA

Thesis: *Efficient Formulation and Implementation of Ensemble Based Methods in Data Assimilation*. Advisor: Prof. Adrian Sandu. GPA 3.63/4.0.

2009 – 2010 **M.Sc. in Industrial Engineering**
Universidad del Norte, Colombia

Thesis: A novel algorithm based on deterministic finite automata for the mono-objective symmetric travelling salesman problem. GPA 4.45/5.0.

2007 – 2009 **M.Sc. in Systems Engineering**
Universidad del Norte, Colombia

Thesis: MODS, a metaheuristic of deterministic swapping for multi-objective combinatorial optimization. GPA 4.28/5.0.

2002 – 2007 **B.Sc. in Systems Engineering**
Universidad del Norte, Colombia

Thesis: Numerical methods for multivariate optimization. GPA 4.02/5.0.

Academic Appointments

2018 – present **Associate Professor, Department of Computer Science**
Universidad del Norte, Barranquilla, Colombia

- Chair of the Department of Computer Science, 2018 to 2023.
- Director of the Applied Math and Computer Science Laboratory (AML-CS) since 2017.
- Undergraduate teaching: algorithms and complexity, optimization, data mining, computational solutions to engineering problems.
- Graduate teaching: data assimilation, optimization theory, non-linear optimization, data mining.
- Supervision of undergraduate and graduate students in data assimilation, data science and data engineering.

2016 – 2017 **Assistant Professor, Department of Computer Science**
Universidad del Norte, Barranquilla, Colombia

04/2020 – 10/2021 **Visiting Instructor, Data Assimilation (graduate)**
Universidad EAFIT, Colombia

08/2015 – 12/2015 **Instructor, Numerical Methods**
Virginia Tech, USA

06/2014 – 07/2014 **Engineering / Technical Scholar**
Lawrence Livermore National Laboratory, USA

06/2013 – 08/2013 **Givens Associate**
Argonne National Laboratory, USA. Covariance matrix design for time-dependent wind energy potential estimation.

07/2011 – 12/2015 **Research Assistant, Computational Science Laboratory**
Virginia Tech, USA. Data assimilation methods for weather forecasting in Python, Fortran and C/C++.

06/2007 – 06/2011 **Instructor, Department of Systems Engineering**
Universidad del Norte, Colombia

Industry Experience (Consulting)

Technical Lead

Data architecture and AI platform assessment

- Built a Python discovery toolkit probing the compliance, admin and analytics surfaces of two major generative-AI vendors, reporting coverage against the thirteen mandatory columns of a Silver-layer Snowflake model.
- Implemented credential-type detection by key prefix before the first API call, so a run of 401s is never misread as missing data; access diagnosis dropped from days to minutes.
- Designed an evidence status scheme separating what was verified live from what was inherited from earlier documentation, preventing pipelines being built on unvalidated assumptions.
- Identified the granularity mismatch that redefined the workstream's scope: vendor analytics APIs return daily per-user aggregates against an event-level fact table.
- Ran an independent pre-pilot assessment of nine high-priority capabilities in a 3,500-seat enterprise workspace, scoring with documented evidence and verifying vendor claims through tests designed to refute them.
- Built a proof-of-concept conversational intake agent on Salesforce Agentforce (pro-code), with guided requirement capture, standardised scope generation and PII guardrails, and established the reusable pattern for eight further agents.

Data Engineer and Senior Data Analyst

Linear TV audience research, Nielsen panel and big data

- Designed and built the prepared-table layer in Snowflake between the unified Nielsen views and business queries, cutting per-question scans of tens of millions of rows to a bounded, reusable dataset across a multi-network portfolio and the full competitive cable and broadcast universe.
- Implemented ~25 documented production queries for overnight summaries, series performance snapshots and competitive rankers, each parameterised with dynamic defaults, no pinned dates and its own validation case.
- Resolved Nielsen integrity problems silently doubling metrics: panel and big-data overlap at the currency cutover, duplicate with/without-VCR rows per telecast, and double-registered ACM streams.
- Implemented broadcast-day chronology anchored at 6:00am so post-midnight telecasts sort within their own night, and established that a Nielsen month is not a calendar month.

- Validated against the client's official rankers and overnight summaries at telecast level, documenting every difference by its methodological cause rather than adjusting it; surfaced a ranker averaging flat instead of weighting by minutes and an inherited telecast floor dropping series silently.
- Built a resolution layer turning natural-language questions into verified parameters, handling Nielsen's 25-character program-name truncation and free-text season codes, with an explicit contract for when the answer does not exist.

Technical Services Lead

Research information management platform, higher education

- Lead an eight-person technical services team owning the data-loading platform behind a research information management product used by 200+ institutions.
- Led the technical discovery that reverse-engineered three undocumented subsystems writing to one Postgres backend: 204 client-specific PHP autoloads, 88 Matillion pipelines and a Python desktop migration tool.
- Audited 839,221 files and 1.13 TB across 108 tenants, surfacing roughly 1 TB of reclaimable storage and 37 institutions dormant 180+ days whose pipelines were about to be migrated at zero business value.
- Escalated production risks found in code review, including an unexplained 894 GB bulk load and a hardcoded database password, each converted into a tracked remediation item.
- Designed the target architecture: a batch-first FastAPI REST service on EKS with idempotency keys and audit logging, targeting a threefold throughput gain, plus a config-driven Databricks pipeline replacing 88 hand-maintained ones.
- Authored the ten-week modernisation plan across five sprints, with shadow-mode cutover, equivalence QA sign-off and rollback runbooks built into every migration ticket.

Staff Data Scientist

Consumer research and audience analytics

- Modelling on survey data at scale in Snowflake: 8.6 billion observations, ~105M users, ~3,600 syndicated questions.
- Documented the ten core warehouse tables and the publisher chain, verifying every relationship and match rate against the information schema; the resulting ERD and catalogue of known traps became the team's mandatory filter standard.
- Built Naive Bayes and KNN answer-imputation models with mutual-information feature selection, validated on accuracy, log-loss and Brier score.
- Attacked the 98.8% sparsity of the user-by-question matrix with publisher-based imputation, extended with a Naive Bayes posterior, evaluated by total variation distance across decoding strategies and publisher levels.
- Built a brand lift pipeline using KNN-imputed treatment and multiple imputation with Rubin's rules, producing lift estimates with correct variance, p-values and confidence intervals.
- Built a census-weighted audience segmentation pipeline (truncated SVD, KMeans by silhouette, raking to population margins), with segment labelling automated across four LLMs in parallel under consensus selection.
- Forecasting of unit sales with neural sequence models, entity and sentiment extraction from open text, and fully parallelised batch processes for large scenario sweeps.

Curriculum Engineer, AI and LLM Education

Online technical education

- Authored ungraded labs on agentic workflows, reflection and planning, powered by GPT and Claude models, for a widely used AI course catalogue.
- Designed agent pipelines in FastAPI with planning, execution and reflection loops for complex task automation.
- Built and deployed agent applications including research assistants, email simulators and customer-segment builders, with tools for SQL execution and document retrieval.
- Developed a research agent that analyses and summarises scientific documents through natural language, using external tools and contextual memory.
- Integrated hosted LLMs into FastAPI services and deployed interactive demos for hands-on learner testing.

Head of Data

E-commerce and retail analytics

- Led the data engineering and data science function across multiple product lines.
- Enriched product catalogues with local LLMs (Mistral, Phi, TinyLlama, Gemma 2, LLaMA 2) for metadata and descriptions.
- Modelled product, customer and feature relationships in Neo4j for graph-based recommendation.
- Built AI agents for product recommendation and audience construction using local and hosted models.
- Developed LLM systems that personalise dialogue and lift conversion across audience segments.
- Near-real-time data lake enrichment with Kinesis Streams and Firehose; storage on S3 and Google Cloud Storage.
- Warehousing on Redshift and BigQuery; transformation layers in dbt; pipelines in AWS Glue, GCP Dataflow and Python.
- Image feature extraction via vision APIs to enrich catalogues and improve DeepFM, Apriori and graph neural network models.
- Customer lifetime value models with scikit-learn and Keras.
- Services published through FastAPI on custom domains; on-demand compute provisioned for model training.

Senior Data Engineer / Senior Data Scientist

Freight and logistics analytics

- Route and path analysis for freight markets using mapping APIs, OpenStreetMap and OSRM.
- Logistics visualisation and optimisation tools built with Folium, Dash and Python.
- ML models and pipelines in Python, Keras, scikit-learn, pandas, NumPy and BigQuery ML, with dbt for transformation.
- Reporting in Looker Studio; delivery through Docker, GitHub, REST APIs and cloud functions.

Senior Data Scientist

Media and streaming measurement

- Ingestion and processing of S3-hosted data into Snowflake, with staging areas for reliable retrieval.
- Models predicting impressions per stream, trained and deployed through Snowflake stored procedures and tasks.
- Dashboards built directly in the warehouse environment with Streamlit.

AI Prompt Engineer (two freelance contracts)

Model evaluation and alignment data work

- Evaluated and ranked model responses against defined criteria, with written justification comparing candidates and identifying failure modes.
- Generated alternative responses targeting specific loss categories, including deliberately flawed responses matching realistic failure scenarios.

Senior Data Engineer (part-time, three-month contract)

Database migration

- Data quality assessment and automated migration comparison between on-premise databases and Azure, using Azure Synapse, Python, SQLAlchemy and pyodbc.
- Dimensionality reduction techniques applied to make real-time record comparison faster and more precise.

Python Developer

Payroll systems

- Python applications generating tailored HTML and Excel payroll reports, on Azure Functions and SQL Server.
- Delivery tracked through Azure DevOps with version control in GitHub.

Data Engineer and Data Science Trainer

Applied data science training

- Led 40 project teams averaging eight members each, on three-month delivery cycles.
- Built Python applications for varied markets including footwear retail and energy.
- Stack: Python, Docker, REST APIs, Dash, AWS RDS, EC2 and S3, with GitHub for versioning.

Data Engineering Intern

Mining sector

- ETL for Oracle databases and reporting built on Microsoft Access and SQL.

Technologies

LANGUAGES	Python, C/C++, Fortran, MATLAB, SQL, Go, Bash.
LLMS & AI	OpenAI models, Anthropic Claude, Mistral, Phi, TinyLlama, Gemma 2, LLaMA 2, Hugging Face Transformers, spaCy, agent pipelines with planning and reflection loops.
ML	scikit-learn, Keras, pandas, NumPy, Prophet, DeepAR, DARTS, MLflow, BigQuery ML, DeepFM, graph neural networks.
DATA ENG.	dbt, PySpark, SQLAlchemy, AWS Glue, Kinesis, Firehose, GCP Dataflow, Azure Data Factory, Azure Synapse.
WAREHOUSES	BigQuery, Redshift, Snowflake (Snowpark, Cortex).
DOMAIN DATA	Nielsen MIT panel and big data, linear TV audience metrics, large-scale survey panels.
DATABASES	PostgreSQL, MySQL, SQL Server, SQLite, Neo4j.
CLOUD	AWS (S3, EC2, RDS, Lambda), Google Cloud Platform, Microsoft Azure.
SERVICES	FastAPI, Flask, Dash, Streamlit, REST APIs, Docker, cloud functions.
REPORTING	Looker Studio, Streamlit dashboards, Excel.
DELIVERY	GitHub, Jira, Confluence, Azure DevOps, Google Colab.

Certifications

- Microsoft Certified: Azure Data Fundamentals (DP-900).
- Microsoft Certified: Azure Fundamentals (AZ-900).
- Data Scientist certification, Correlation One.

Selected Publications

Twelve of 70+ works. Complete record: https://enino84.github.io/publications · ORCID 0000-0001-7784-8163. Google Scholar: 806 citations, h-index 15, i10-index 24.

1. **Niño-Ruiz, E. D.** (2026). PyTEDA-web: a FastAPI platform for interactive data assimilation benchmarking with real-time streaming and persistent experiment tracking. *SoftwareX*, 34, 102738.
2. **Niño-Ruiz, E. D.** (2025). TEDA: a lightweight Python framework for educational data assimilation. *SoftwareX*, 31, 102297.
3. Pérez Rivera, J. D., Turizo, D., **Niño-Ruiz, E. D.**, & Montoya, O. D. (2025). Improved Rosenbrock method with error estimator and Jacobian approximation using complex step. *Results in Applied Mathematics*, 27, 100629.
4. **Niño-Ruiz, E. D.**, Cabarcas Arrieta, D. S., Fernández Quiroz, G. R., & Quintero, N. (2025). Statistical package for computing precision covariance matrices via modified Cholesky decomposition. *SoftwareX*, 30, 102125.
5. **Niño-Ruiz, E. D.**, & Díaz-Rodríguez, J. (2024). A 4D-EnKF method via a modified Cholesky decomposition and line search optimization for non-linear data assimilation. *Atmosphere*, 15(12), 1412.
6. Popov, A. A., Sandu, A., **Niño-Ruiz, E. D.**, & Evensen, G. (2023). A stochastic covariance shrinkage approach in ensemble transform Kalman filtering. *Tellus A*, 75(1), 159–171.

7. **Niño-Ruiz, E. D.**, Consuegra Ortega, R. S., & Lucini, M. (2023). Ensemble based methods for leapfrog integration in the simplified parameterizations, primitive-equation dynamics model. *Quarterly Journal of the Royal Meteorological Society*, 149(751), 573–587.
8. López-Restrepo, S., **Niño-Ruiz, E. D.**, Guzmán-Reyes, L. G., Yarce, A., Pinel, N., & Heemink, A. W. (2021). An efficient ensemble Kalman filter implementation via shrinkage covariance matrix estimation. *Computational Geosciences*, 25(3), 985–1003.
9. **Niño-Ruiz, E. D.**, Guzmán-Reyes, L. G., & Beltrán-Arrieta, R. (2020). An adjoint-free four-dimensional variational data assimilation method via a modified Cholesky decomposition and an iterative Woodbury matrix formula. *Nonlinear Dynamics*, 99(3), 2441–2457.
10. **Niño-Ruiz, E. D.**, Sandu, A., & Deng, X. (2018). An ensemble Kalman filter implementation based on modified Cholesky decomposition for inverse covariance matrix estimation. *SIAM Journal on Scientific Computing*, 40(2), A867–A886.
11. **Niño-Ruiz, E. D.**, Sandu, A., & Anderson, J. (2015). An efficient implementation of the ensemble Kalman filter based on an iterative Sherman–Morrison formula. *Statistics and Computing*, 25(3), 561–577.
12. **Niño-Ruiz, E. D.**, & Sandu, A. (2015). Ensemble Kalman filter implementations based on shrinkage covariance matrix estimation. *Ocean Dynamics*, 65(11), 1423–1439.

Open-Source Software

- **TEDA**, a Python toolbox for teaching ensemble-based data assimilation: stochastic and dual EnKF, Cholesky and modified-Cholesky variants, B-localization, with Duffing, Lorenz-63 and Lorenz-96 test models. github.com/enino84/TEDA
- **PyTEDA-web**, a FastAPI platform for interactive data assimilation benchmarking with real-time streaming and persistent experiment tracking. github.com/enino84/pyTEDA
- **AMLCS-DA**, a data assimilation package in Python for atmospheric general circulation models. github.com/enino84/AMLCS
- **Modified Cholesky package**, precision covariance matrix estimation released as a standalone statistical package.

Editorial Service and Peer Review

- Early Career Advisory Board, *Control Engineering Practice*, Elsevier, 2021 to present.
- Guest Editor, special issues of the *International Journal of Artificial Intelligence* on machine learning methods for inverse problems (2021), combinatorial optimization methods for inverse problems (2018), and combinatorial optimization (2013).
- 90 peer reviews for 41 publications and grants, including *Quarterly Journal of the Royal Meteorological Society* (10), *Soft Computing* (8), *IEEE Access* (6), *Entropy* (4), *Pure and Applied Geophysics* (4), *Monthly Weather Review* (3), *Journal of Computational Physics*, *Geoscientific Model Development* and *IEEE Transactions on Cybernetics*.

Teaching and Outreach

- Author of a free graduate course on data assimilation, with full lecture notes and Python notebooks, used in the Ph.D. programme in Systems Engineering at Universidad del Norte. enino84.github.io/courses/intro_data_assimilation
- Keynote, EnKF Workshop 2021, NORCE Norwegian Research Centre.
- Keynote, 7th International Symposium on Data Assimilation (ISDA 2019), RIKEN R-CCS, Kobe, Japan.
- Invited talks for Asociación de Universidades Grupo Montevideo, Universidad EAFIT and Universidad del Norte.

Awards

- Best Workshop Paper Award, International Conference on Computational Science (ICCS 2017), Zurich, for *A surrogate model based on mixtures of Taylor expansions for trust region based methods*.