JACQUES BOOYSEN

SENIOR DATA SCIENTIST · CITY OF JOHANNESBURG, SOUTH AFRICA · +27835085567

⊙ DETAILS ⊙

City of Johannesburg, South Africa
+27835085567
booysenjacques@gmail.com

⊙ LINKS ⊙

Linkedin

Portfolio

GitHub

⊙ LANGUAGES ⊙

Afrikaans

English

⊙ SKILLS ⊙

AWS/Cloud

Machine Learning

Renewable Energy

R, Python

Rapid Prototyping

Apache Airflow

Energy Modelling

Geospatial Data

⊙ HOBBIES ⊙

Adventure Racing, Ultra Trail Running, Mountain Sports

👤 PROFILE

I am a scientist at heart, passionate about Data Science, problem solving and software engineering. I am responsible for helping utilities and other companies interpret and manage data and solve complex problems using expertise in a variety of data niches often employing computer science, serverless cloud computing , machine learning, optimisation, modelling, statistics and advanced visualisations.

Created renewable energy forecasting models and interactive tools for forecasting Solar PV, Wind and Hydro power plant outputs hourly using global circulation models and historic satellite data/models. I have also been involved in the development of various electricity load research and energy modelling projects in R.

My current role is technical lead and data scientist in the disruptive innovation business unit at Enerweb, I created frameworks that our team use to provide Data Science as a Service on our Python based in house developed SaaS technology stack named Enerflow. I am an AWS certified solutions architect, and our stack uses AWS micro services architecture and Apache Airflow, allowing us to run Jupyter notebooks in the cloud from MS Excel.

My strength lies in my ability to do rapid prototyping because of my strong software engineering background and anything is possible attitude, my passion for machine learning and constantly improving myself and thinking how we as a team can do things in a better way.

💼 EMPLOYMENT HISTORY

Senior Data Scientist at Enerweb, South Africa
January 2019 — Present

Developed energy time series meter data anomaly detection tool for an energy utility using k-means clustering and statistics, and a model to forecast the Southern African Power Pool (SAPP) day ahead pool price.

I have produced day ahead hourly renewable energy forecasting models for utility scale grid connected PV/Wind farms, including an interactive R shiny app applying the models bottom up for all IPP sites in South Africa to give an aggregated forecast for solar and wind plant generation. Also, a hydro power forecasting model based on an existing WEAP model downscaled to daily data using ERA5 weather reanalysis models.

Developed an AWS micro services Python framework for our Data Science stack deployed using the DEVOPS process and CICD. Infrastructure as code written using (AWS SAM). The Framework uses the AWS miroservices framework and Apache Airflow (MWAA), to run our DATAOPS/MLOPS workflows, also using papermill and quarto to run Jupyter notebooks in production.

Skills: AWS/Cloud · Machine Learning · Apache Airflow · Python · R · SQL

Data Scientist at Enerweb
January 2010 — Present

Optimal regional geo based load forecasting (GLF), load subclass development finding sweet spot between error and model complexity (BIC), also a spatial buffering algorithm to calculate domestic building density estimates per spatial LSM. See my published papers for more details.

Spatial interpolation of temperature using digital elevation model and comparing and

evaluating various modelling techniques, a market intelligence system in R shiny for network planning support and sentiment analysis using MongoDB and machine learning.

Various energy modelling projects using an electricity market model that was written in MESSAGE from the International Institute for Applied Systems Analysis (IIASA) based on the bottom up IRENA model applying linear & mixed-integer optimisation.

The projects included a model estimating the least cost energy flows within and between the interconnected SAPP-EAPP power pools with results shown over a medium to long term till 2040 and for different scenarios. Each country was modelled as a node interconnected by cross-border Tx Lines. Another project was an Inter-connector Impact Study of ANNA a 400kV line connecting Angola and Namibia, also a CRIDF project used downscaled Climate Projections from NASA Earth Exchange (NEX) with a similar energy model in a Monte Carlo Simulation.

For the above energy modelling projects, I was the primary modeler on the ANNA interconnector feasibility study project, the modelling assistant on the SAPP-EAPP project, and climate data engineer and modelling assistant on the CRIDF project. For all projects I architected and implemented R Shiny and Tableau user interfaces.

Various other research projects developed for Eskom research on our in house DIAS (R/Java) platform, projects ranging from non-techinal losses detection, DSM impact studies, market intelligence and optimal power plant maintenace analysis.

Skills: R · MESSAGE · Energy Modelling · Optimisation · Shiny · Tableau · SQL · MongoDB

Software Engineer at Enerweb
January 2002 — December 2009

Started as a Linux programmer and Back-end Specialist, doing various research projects in C and Perl, and later Java and J2EE.

I was also the solutions architect and main developer piloting the Virtual Power Station (VPS) scada solution for Eskom, this involved a distributed system running on Linux servers over corporate GPRS/4G network, with integration with Schneider Momentum PLC and Frequency Relays. The initial solution was written in Perl and C, with my own C implementation of Modbus to talk to the PLC and SPABUS to talk to the Frequency Relays. I also developed an IVR solution using JBoss and Linux asterisk. Also did some PLC programming. The initial solution has been re-written in C# by our dev team for productization and are currently part of the range of company products.

Eskom decided to embark on a self developed, enterprise wide data warehouse, for financial, technical, HR and environmental data. This solution had to provide a "single version of truth", across all divisions, and had to be developed specifically according to Eskom business processes and technical standards. Immense technical complexity was overcome to produce a world leading technical solution. I was responsible for and developed a generic recursive roll-up engine as Java stored procedures that could navigate and extract multidimensional roll-ups over cubes and complex hierarchies in the data model.

Wrote a X.25 Protocol gateway using Linux C/Visual Basic and Open VPN, assisting Eskom Engineers to connect to substation equipment from the Eskom WAN.

I also wrote a custom Linux authenticating firewall system for Eskom Telecoms, enabling control and secure access to the operational telecoms network via software that could only use RS232 ports.

Skills: Linux · Perl · C · PHP · PL/SQL · Oracle · SCADA · PLC · Java · J2EE

✿ EDUCATION

B.Sc (Hons) Computer Science, University of Johannesburg
January 1998 — December 2001

Bachelor of Science (BSc) Applied Mathematics, University of Johannesburg
January 1998 — December 2001

★ PUBLICATIONS

Domestic building density estimates for network planning,
April 2014

Domestic Use of Energy Conference (DUE) - see LinkedIn

Regional electricity load profile subclasses for network planning
August 2013

Industrial and Commercial Use of Energy Conference (ICUE) - see LinkedIn

★ CERTIFICATIONS

AWS Certified Solutions Architect – Associate,
August 2022 — August 2025

R Programming
February 2015 — Present

Coursera Verified Certificates X9HT2EW4WW www.coursera.org

The Data Scientist's Toolbox
December 2014 — Present

Coursera Verified Certificates 2NZPCAYE53 www.coursera.org

✪ COURSES

Water Evaluation And Planning System (WEAP) Intro, SEI — Stockholm Environment Institute
October 2020 — October 2020

◀ REFERENCES

Schalk Heunis - Managing Executive: Big Data, AI & RPA (Vodacom)
schalk.heunis@vodacom.co.za

PhD Electrical Engineering

Marcus Dekenah – Load Research Specialist (MD Consulting)
marcus@mdekenah.co.za

NHD Elec. Eng (HC), M Dip.Tech. Eng, Bsc Elec Eng (Cum Laude), MBLII

★ PROFESSIONAL MEMBERSHIPS

The South African Institute of Electrical Engineers , Johannesburg
April 2019 — Present

Member and presenter at the load research chapter, Presenter at SatRday 2018: Practical applications using R for spatial data visualisation,creation and manipulation

Key:	AWS/Cloud	R / Machine Learning	Python	ML/Data/Dev-Ops
Years of Experience	4 years (2019)	12 years (2010)	4 years (2019)	1 year (2022)
Notes	Mar 2019: Training at AWS Popup Loft Johannesburg Been using AWS at various levels, for HPC for the last 4 years. I am an AWS certified solutions architect since 2022, use AWS CloudFormation and develop microservices daily.	Started programming R in 2010 to implement a large quantity and range of research projects using machine learning in R for Eskom Research. Used AWS sagemaker AutoML to evaluate different models for renewable energy forecasting.	Developing of research and SaaS using Python on AWS lambda and Jupyter notebooks in vscode. Used boto3 to create sagemaker notebooks that mounts EFS.	Apache Airflow Ploomber Azure (Devops)
Tools/Libraries	Lambda, SQS, TexTract, RDS, EC2, VPC, CloudFormation, MWAA, S3, EFS, AIM, etc.	dplyr, raster, ggplot2, shiny, mlr, random forests, rgdal, clustering etc.	pvlib, pandas, scikit-learn, pylint, pytest, boto3, re, xgboost	airflow, papermill, ploomber