

EDUCATION AND TRAINING

Research Fellow DFCI/Harvard University, Boston, MA August 2023-current

- **Mentor:** Giovanni Parmigiani, PhD
- **Project:** Developing Data Mining Algorithm using Bayesian Discrete Matrix Factorization Approach for Application to Genomics Data

**PhD. in Probability and Statistics Centro de Investigación en Matemáticas, AC,
Guanajuato, Gto, México March 2018**

- **Thesis Title:** Problems in Statistical Genetics: Classification and Testing for Network Changes.
- **Advisors:** Graciela González Farías , PhD & Terence P. Speed, PhD

Visiting PhD candidate Walter and Eliza Hall Institute of Medical Research/ University of Melbourne VIC, Australia August 2016-July 2017

Master of Science in Statistics Egerton University, Njoro, Kenya December 2008

- **Thesis Title:** Efficiency Evaluation in Modelling of Stock Data using ARCH and Bilinear Models.
- **Supervisors:** Dankit Nassiuma, PhD & Ali Islam, PhD

Bachelor of Education Science, Egerton University Njoro, Kenya July 2004

- **Major in Mathematics & minor in Chemistry**

SKILLS

- **Data Analytics:** R , PYTHON & SAS
- **Operating Systems:** MS Windows, Linux/Ubuntu, macOS
- **Word Processing:** LATEX, MS Office, Libre Office
- **Version Control:** Git, Github
- **Machine Learning:** Application of various machine learning algorithms.
- **Bioinformatics:** Unix/Shell Command prompt, Bulk RNA-seq analysis, DGE Analysis.
- **Education:** Pedagogy and Curriculum Development

SHORT COURSES AND OTHER RELEVANT TRAININGS

Harvard University

Boston, MA, USA

Workshop series by Harvard Chan Bioinformatics Core (HBC).

Feb 2024–May 2024

- Introduction to the command-line interface and high-performance computing (HPC).
- The Foundation - Basic Shell.
- Accelerate with Automation - Making your code work for you.
- Needle in a Haystack - Finding and Summarizing data from colossal files.
- Shell Tips and Tricks on O2.
- Introduction to RNA-seq using high-performance computing (HPC).
- Introduction Differential Gene Expression and Functional analysis.
- Introduction to single-cell RNA-seq data analysis.

Dana Farber Cancer Institute

Boston, MA, USA

DF/HCC and UMass Boston Research Education Core (REC) Program.

Feb 2024–May 2024

- Introduction to Mentoring: Fostering Scientific Identity
- Aligning Expectations
- Effective Communication: Communication Styles
- Effective Communication: Communicating with your Mentee
- Short Term Research Project Design
- Addressing Diversity & Inclusion

Karatina University

Karatina, Kenya

16 -20 Sept 2019

- Spatial Data Modelling: Theory and Practice

Bath, UNAM and CIMAT (BUC) Workshop Series

Guanajuato, Gto, M´exico

12-14 Nov. 2015

- Statistics in Environmental Research

University of Washington Seattle, W.A, USA.

6-17 July 2015

- Summer Institute in Statistics for Big Data (SISBID)
 - Accessing Biomedical Big Data
 - Visualization of Biomedical Big Data
 - Supervised Methods for Statistical Machine Learning.
 - Supervised & Unsupervised Methods for Statistical Machine Learning

CIMAT, AC

Guanajuato, Gto, M´exico,

9-11 Jun 2015

- A Short Course of Big Data & Machine Learning

Universidad Nacional Autónoma de México, CEPE

Taxco, Gro, México Jan- July 2013

- Spanish Language and Mexican Culture

Dana Farber Cancer Institute

Boston, MA, USA

Department of Data Science

Aug 2023 –current

Research Fellow

- I am working on a project to develop a novel methodology for data mining and machine learning methods mainly via boolean and ternary matrix factorization. This will be useful in revealing the hidden patterns in boolean/ternary types of data for instance the chromosomal alteration data for multiple myeloma disease.
- I am also involved in mentoring students in the data science/machine learning applications in genomics data especially the summer research project.

Bomet University College

Bomet, Kenya

Department of Mathematics & Computer Science.

Sept. 2020 - Aug. 2023

- **Senior lecturer in Statistics**

In this position, the duties and responsibilities were mainly: Research, Training and Teaching.

As a senior lecturer, I taught and assessed Mathematics and Statistics courses at both undergraduate and graduate levels, supervised student projects and theses, provided academic advising, developed curriculum, participated in research projects and academic leadership initiatives, and contributed to professional and community services, including fundraising and fostering institutional linkages.

- **Head of Department of Mathematics & Computer Science**

I was responsible for allocating duties to staff, preparing the departmental workplan, procurement plan, and budget, supervising teaching and learning activities, and attending

- **Coordinator: ERP- Master Timetable**

I coordinated and maintained the Master Timetable based on input from various departments and units, assigned courses, internships, and research activities, allocated appropriate teaching and research spaces, continuously updated the timetable database, and managed scheduling, rescheduling, and emerging issues.

Chuka University, Chuka, Kenya,

July 2012- August 2020

My career at Chuka University spanned around three phases. From July 2000 to August 2012, I was an assistant lecturer of Mathematics and Statistics in the Department of Business Administration. Then I was promoted to a lecturer in September 2012. From January 2013 to March 2018, I was on leave to pursue a PhD degree. I returned to Chuka University in March 2018 to continue as a lecturer.

▪ **Lecturer**

Sept 2013- Dec 2013, & March 2018 -August 2020

My responsibilities included developing and writing research proposals to attract grants, conducting applied research to address societal needs, and sharing findings through publications and presentations. I also provided consultancy in experimental design and statistical data analysis. In teaching, I taught Mathematics and Statistics at undergraduate and Masters levels, supervised and examined students, and coordinated all aspects of examination administration and moderation.

▪ **Assistant Lecturer**

July 2009-August 2012

As an Assistant Lecturer, I taught Mathematics and Statistics to undergraduate and diploma students, provided research and consultancy support to postgraduate students, and was responsible for supervising and examining students.

Short-Term Consultant (Advisory and Data Analytics), World Bank Group

Aug 2022-June 2023

ACP World Bank Firm Survey Kenya Investment

In this project, the World Bank purposed to map out how (Sub) National policy of the Government of Kenya affects the price competitiveness of its priority sectors via a private sector survey. The survey targeted organizations in Kenya with the aim of capturing organizations' historical performance, the effect of (Sub) National taxes, subsidies, levies, and fines on production costs and profitability, the main drivers for the existing investment decisions, and their biggest challenges related to government policies. I successfully supervised the whole

sample survey process from the beginning to the end. My specific roles and responsibilities included:

- Supervision of the designing and testing of the research tools.
- Advisory on the sample size determination and sampling procedures.
- Leader in tool testing and piloting.
- Supervision of data collection, cleaning and Analysis
- Participation in the stakeholders validation workshop.
- Writing and presentation of the final report.

The project concluded successfully, it was validated by the stakeholders and adopted by the World Bank.

SOFTWARE

I am currently building an R package for **Bayesian Boolean Matrix Factorization (BBooMF)**

TEACHING EXPERIENCE

MAT 104e: Basic Mathematics and Analytic Geometry: Backhouse, J. K, Houldworth, S.P.T and Cooper, B.E.D. (2002). Pure Mathematics 1. Cambridge: Cambridge University Press. Thomas, G.B., and Finney, R. L, (1995). Calculus and Analytic Geometry. New Jersey: Addison-Wesley Publishing Company.

MAT 113e: Differential Calculus: Thomas, G.B., and Finney, R. L, (1995). Calculus and Analytic Geometry. New Jersey: Addison-Wesley Publishing Company.

STA 100e: Probability and Statistics I. & STA 301e: Probability and Statistics II: Hogg, V.R, McKean J.W. & Craig. A.T. (2003). Introduction to Mathematical Statistics, 6th ed., Prentice Hall.

STA 302e: Multivariate Probability Distributions: Chartsfield, C. and Collins, A.J. (1980) Introduction to Multivariate Analysis, Chapman and Hall

STA 313: Theory of Estimation: Hogg, V.R, McKean J.W. & Craig, A.T. (2003). Introduction to Mathematical Statistics, 6th ed., Prentice Hall. Dennis Wackerly, William Mendenhall, and Richard L. Scheaffer. (2007). Mathematical Statistics with Applications. P W S Publishers.

STA 411e: Time Series Analysis: Shumway R. H. and Stoffer D. S. (2013). Time Series Analysis and Its Applications: with R Brockwell. P.J. and Davis R.A. (2009). Time Series : Theory and Methods, (2nd Edition).

STA 310: Advanced Probability and Statistics: Robert B. Ash 2008. Basic Probability Theory. Dover Publications. Sidney Resnick (1999). A Probability Path . Birkhauser, Boston

STA 410: Tests of Hypothesis: Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2). Duxbury Pacific Grove, CA.

STA 415: Design and Analysis of Experiments: Douglas C. Montgomery 2019. Design and Analysis of Experiments. 10th Edition. ISBN: 978-1-119-49244-3

MATH 141: Introductory Statistics: Introduction to Mathematical Statistics by R.V. Hogg & A.T. Craig.

Math 244: Introduction to Statistical Inference: Dennis Wackerly, William Mendenhall, and Richard L. Scheaffer, (2007). Mathematical Statistics with Applications, PWS Publishers 2007. Hogg, R.V and Elliot A. Tanis, (2005). Probability and Statistical Inference, 7th ed. Prentice Hall College.

MATH 443: Design and Analysis of Experiments I & MATH 444: Design and Analysis of Experiments II: Douglas C. Montgomery 2019. Design and Analysis of Experiments. 10th Edition. ISBN: 978-1-119-49244-3

MATH 442: Time Series Analysis: Shumway R. H. and Stoffer D. S. (2013). Time Series Analysis and Its Applications: with R Cryer J. D. and Chan K. S. (2010). Time Series Analysis with

Application in R. Springer. Brockwell P.J. and Davis R.A. (2009). Time Series : Theory and Methods, (2nd Edition).

MATH 456: Introductory Bayesian Inference: Bolstad. M. (2007). Introduction to Bayesian Statistics. John Wiley & Sons Inc
Lee, Peter m. (1994) Bayesian Statistics: An Introduction London : Edward Arnold.

MATH 344: Theory of Estimation: Hogg, V.R, McKean J.W. & Craig. A.T. (2003). Introduction to Mathematical Statistics, 6th ed., Prentice Hall. Dennis Wackerly, William Mendenhall, and Richard L. Scheaffer. (2007). Mathematical Statistics with Applications. P W S Publishers.

MATH 240: Probability and Statistics I & MATH 241: Probability and Statistics II: Hogg, V.R, McKean J.W. & Craig. A.T. (2003). Introduction to Mathematical Statistics, 6th ed., Prentice Hall.
MATH 449: Probability Theory: Robert B. Ash 2008. Basic Probability Theory. Dover Publications. Sidney Resnick (1999). A Probability Path . Birkhauser, Boston.

MATH 842: Measure Theory and Probability: Billingsley P. (2012). Probability and Measure, (Anniversary Edition). New York, John Wiley & sons, Inc.

MATH 848: Probability and Stochastic Processes: Durrett R. 2019: Probability: Theory and Examples. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. Achim Klenke 2020. Probability Theory. Springer.

COSC 811: Simulation and Modelling: Harry Perros (2008). Computer Simulation Techniques: The definitive introduction. Michael Pidd (2004), (Editor). Systems Modelling: Theory and Practice. John Wiley & Sons
Stewart Robinson (2004). Simulation: The Practice of Model Development and Use. John Wiley & Sons Ltd.

MATH 859: Financial Time Series and Risk Management: Tsay R. S. (2013). Multivariate Time Series Analysis: With R and Financial Applications. Wiley. Shumway R. H. and Stoffer D. S. (2013). Time Series Analysis and Its Applications: with R
Cryer J. D. and Chan K. S. (2010). Time

Series Analysis with Application in R. Springer. Brockwell P.J. and Davis R.A. (2009). Time Series : Theory and Methods, (2nd Edition). Springer.

MATH 846: Time Series Analysis. Tsay R. s. (2013). Multivariate Time Series Analysis: With R and Financial Applications. Wiley. Shumway R. H. and Stoffer D. S. (2013). Time Series Analysis and Its Applications: with R Examples. Springer.

STUDENTS MENTORSHIP AND SUPERVISION

- [1] JUMA SAMWEL KISIANG'ANI. Queuing Theory and it's Application to Modelling of Waiting Time: A Co-operative Bank of Kenya, Chuka Branch. Msc (Applied Statistics), Chuka University, Kenya Graduated in 2023
- [2] ERIC MUNENE NDEGE. Application of Asymmetric-GARCH Type Models To The Kenyan Exchange Rate And Balance of Payments. Msc (Applied Statistics), Chuka University, Kenya Graduated in September 2022
- [3] ONESMUS MBUKO TITUS. Modelling Of Foreign Exchange Volatility and it's Effects on International Trade in Kenya. Msc (Economics), Chuka University, Kenya Graduated in December 2021
- [4] KIMUTAI KOECH EMMANUEL. Singular Spectrum Analysis and Holt-Winters Triple Exponential Smoothing: An Application to Kenyas Industrial Input Price Index. Msc (Applied Statistics), Chuka University, Kenya Graduated in December 2021
- [5] MBURU S. NJOROGI. Vector Error Correction Model: Prediction of Bi-Directional Causality Between Gross Domestic Product and Wage Growth In Kenya. Msc (Applied Statistics), Chuka University, Kenya Graduated in December 2021
- [6] WANJUKI TEDDY MUTUGI. Forecasting Commodity Price Index of Food And Beverages In Kenya Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Models. Msc (Applied Statistics), Chuka University, Kenya Graduated in December 2021
- [7] LEONARD KIMTAI MASAI. Application of Response Surface Methodology in Modelling and Optimization of the Yields of Common Beans (*Phaseolus vulgaris* L.) using Animal Organic Manures. Msc (Applied Statistics), Chuka University, Kenya Graduated in October 2019

PUBLICATIONS

- [1] V. W. Lumumba, T. W. Mutugi, and **A. Wagala**, "Forecasting of Monthly Crude Oil Prices in Kenya Using Comparative Time Series Models", *Asian Journal of Probability and Statistics*, vol. 26, no. 9, pp. 97-109, 2024.
- [2] E. M. Ndege, D. K. Muriithi, and **A. Wagala**, "Application of Asymmetric-GARCH Type Models to The Kenyan Exchange Rates", *European Journal of Mathematics and Statistics*, vol. 4, no. 4, pp. 78-92, 2023.
- [3] K. K. Emmanuel, A. Wagala, and D. K. Muriithi, "Singular spectrum analysis: an application to Kenya's industrial inputs price index", *European Journal of Mathematics and Statistics*, vol. 3, no. 1, pp. 1-13, 2022.
- [4] M. S. Njoroge, G. G. Njoroge, and **A. Wagala**, "Vector Error Correction Model: Prediction of Bi-Directional Causality between Gross Domestic Product and Wage Growth in Kenya", *European Journal of Mathematics and Statistics*, vol. 3, no. 4, pp. 6-16, 2022.
- [5] O. M. Titus, **A. Wagala**, and D. K. Muriithi, "Foreign Exchange Rate Volatility and Its Effect on International Trade in Kenya", *European Journal of Humanities and Social Sciences*, vol. 2, no. 4, pp. 47-56, 2022.
- [6] T. M. Wanjuki, **A. Wagala**, and D. K. Muriithi, "Evaluating the Predictive Ability of Seasonal Autoregressive Integrated Moving Average (SARIMA) Models using Food and Beverages Price Index in Kenya", *European Journal of Mathematics and Statistics*, vol. 3, no. 2, pp. 28-38, 2022.
- [7] T. M. Wanjuki, **A. Wagala**, and D. K. Muriithi, "Forecasting Commodity Price Index of Food and Beverages in Kenya Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Models", *European Journal of Mathematics and Statistics*, vol. 2, no. 6, pp. 50-63, 2021.
- [8] K. L. Masai, M. M. Muraya, and **A. Wagala**, "Application of Response Surface Methodology in Optimization of the Yields of Common Bean (*Phaseolus vulgaris* L.) Using Animal Manures", *American Journal of Applied Mathematics and Statistics*, vol. 8, no. 2, pp. 64-68, 2020.
- [9] **A. Wagala**, "A Likelihood Ratio Test for Correlated Paired Multivariate Samples.", *Chilean Journal of Statistics (ChJS)*, vol. 11, no. 1, 2020.
- [10] **A. Wagala**, G. González-Farías, R. Ramos, and O. Dalmau, "PLS Generalized Linear Regression and Kernel Multilogit Algorithm (KMA) for Microarray Data Classification Problem", *Revista Colombiana de Estadística*, vol. 43, no. 2, pp. 233-249, 2020.
- [11] D. N. Nyongesa and **A. Wagala**, "Non Linear Time Series Modelling of the Diesel Prices in Kenya", *International Journal of Academic Research in Economics and Management Sciences*, vol. 5, no. 4, pp. 105-125, 2016.
- [12] M. K. G. Kiremu, N. Galo, **A. Wagala**, and J. K. Mutegi, "Stock price and volumes reaction to annual earnings announcement: A case of the Nairobi Securities Exchange", *International Journal of Business*,

Humanities and Technology, vol. 3, no. 2, pp. 101–111, 2013.

[13] **A. Wagala**, D. Nassiuma, A. Islam, and J. Mwangi, “Volatility modeling of the Nairobi Securities Exchange weekly returns using the ARCH-type models”, *International Journal of Applied Science and Technology*, vol. 2, no. 3, pp. 165–174, 2012.

[14] **A. Wagala**, A. S. Islam, and D. K. Nassiuma, “Efficiency Evaluation When Modelling Nairobi Security Exchange Data Using Bilinear and Bilinear-Garch (BL-GARCH) Models”, *International Journal of Applied Science and Technology*, vol. 2, no. 6, pp. 37–48, 2012.

LANGUAGES

- **English:** Advanced User
- **Spanish:** Intermediate (Con conversationally fluent)
- **Kiswahili:** Fluent in speaking, reading & writing

SCHOLARSHIPS AND FUNDING

- 2022 IMU-Simons Africa Fellowship May 2023
- Mexican Government Scholarship for Foreigners (SRE) Jan 2013 – July 2014
- Scholarship by University of Washington, Seattle, W.A, USA July 2015
- CONACyT, Mexico Aug. 2014–March 2018

PROFESSIONAL MEMBERSHIPS

- Statistical Society of Kenya (SSK) 2012–Current
- American Statistical Association (ASA) 2025–Current

REFERENCES

Available on Request