



RESEARCH PORTFOLIO

Janet van Niekerk

I am associate editor for *Bayesian Analysis* (IF 3.7, Q1) and *Statistics and Computing* (IF 2.0, Q1).

I have delivered international plenary talks, keynote addresses and several invited talks at international meetings, on foundational Statistics as well as Biostatistical applications.

I am a Bayesian statistician with a focus on computing and biostatistical applications.

Publications: 31

Citations: 420+

[Google h-index](#) : 12

[Scopus h-index](#) : 7

[Web of Science h-index](#) : 6

ORCID: 0000-0002-4334-2057

Table of Contents

1. Research self-evaluation	2
2. Summary of completed research.....	4
1. Prior Specification: Challenging Conventional Approaches.....	4
2. Advancing Computational Methods: From Exact to Approximate Inference.....	4
3. Development of INLA 2.0.....	5
4. Applications in Biostatistics and Epidemiology	5
3. Future and current research.....	7
1. Robust default priors.....	8
2. Complex survival analysis and joint models for precision medicine	8
3. Bayesian models for genomic and precision medicine.....	8
4. Epidemiology and environmental statistics	8
5. Bayesian machine learning.....	9

1. Research self-evaluation

I view myself as a Bayesian Statistician, with my passion for Bayesian statistics taking root in the second year of my undergraduate studies. I found this alternative approach to statistical inference to be more aligned with real-world challenges and more intuitive. Since then, my research has focused on Bayesian methods, particularly the specification of robust priors and the computational frameworks required for exact and approximate Bayesian inference, respectively.

During my PhD, I challenged the default use of the conjugate inverse-Wishart prior for covariance matrices. When challenging the use of conjugate priors, a tangible computational burden arises as the resulting posteriors are often not tractable. I subsequently developed a niche matrix-variate Gibbs samplers during my PhD. This work laid the foundation for prior development in various contexts, such as the development of robust contraction priors for survival and binary regression models. Based on this work which was published in *Journal of Multivariate Analysis*, I was invited as a speaker at the King Abdullah University of Science and Technology by Prof. Håvard Rue, a global leader in approximate Bayesian inference for latent Gaussian models, in 2017.

Based on my PhD work on developing and implementing non-conjugate yet effective matrix-variate priors, I have continued to make significant contributions to the field of Bayesian statistics. One of my key advancements involves the development of information-theoretic priors based on distance metrics designed to penalize model complexity. These priors, which provide robust default options with predictable and desirable properties, have been highly regarded by the Bayesian community. They fill a critical gap in the Bayesian workflow by offering practitioners reliable tools that were previously unavailable, as evident from citations of this work in top Statistics journals such as *Bayesian Analysis* (IF 3.7, Q1) and *Statistical Methods in Medical Research* (IF 3.0, Q1) and engineering journal *Composite Structures* (IF 6.3, Q1).

Another major focus of my research, initiated during my PhD, has been the computational aspects of Bayesian inference. To ensure that my proposed methodologies were practical, I implemented matrix-variate Gibbs samplers, which facilitated their application. However, recognizing the limitations of exact or sampling-based methods, such as convergence issues and high computational costs, I turned my attention to approximate Bayesian inference. My contributions span foundational work, applications, and novel statistical models in this area.

Notably, I developed a new Gaussian approximation technique that achieves the accuracy of sampling-based methods at the computational cost of simpler approximations. This innovation has significantly advanced the Integrated Nested Laplace Approximation (INLA) methodology, elevating its capabilities for high-dimensional problems. I was invited by Prof. Bob Carpenter to the Flatiron Institute in New York, in 2024, to collaborate on incorporating this work into the STAN software, the state-of-the-art exact Bayesian inference approach. This work has profound implications for Bayesian machine learning and other fields requiring efficient and accurate high-dimensional Gaussian approximations. Various projects of embedding this work in current inferential architectures are ongoing.

Furthermore, I contributed to the development of the next generation INLA framework. This framework empowers applied researchers to construct and infer more complex models for larger datasets. It has been successfully applied to non-linear generalized mixed effects models, spatial and spatio-temporal models, survival analysis, multivariate disease mapping, and more. Citations of this work can be found in Spatial Statistics (IF 2.1, Q1), Spatial and spatio-temporal Epidemiology (IF 2.1, Q1), Econometrics (IF 1.5, Q1), Journal of Environmental Management (IF 8.0, Q1), Methods in Ecology and Evolution (IF 6.3, Q1), amongst others.

In addition to my research contributions, I have supervised numerous postgraduate students (four completed PhD's and seven completed Masters), many of whom have published their research in high-impact journals. Mentoring the next generation of statisticians has been a fulfilling aspect of my career, and their success is a testament to the relevance and rigor of the projects we've undertaken.

Based on my current list of publications I have received 420+ citations. Seven of my publications (31.8%) are in the top 25% most cited documents worldwide and six publications (28.6%) have been published in the top 25% most cited journals (<https://www.scopus.com/authid/detail.uri?authorId=57219493915#tab=metrics>).

Moreover, my contributions have earned me numerous prestigious invitations to deliver short courses and participate in research visits at internationally esteemed institutions, such as the Centers for Disease Control and Prevention (USA), University of Lisbon (Portugal), University of Bordeaux (France), Universitas Airlangga (online available <https://www.youtube.com/watch?v=a-unDONKoRw>, Indonesia), University of Pretoria (RSA), University of South Africa (RSA), and others.

I have built an international network of collaborators and subsequently I was invited as the opening plenary speaker at the Workshop for Survival Analysis (<https://viwasa.softaliza.com.br/conferencias/>) in Brasilia in October 2024, for my work on efficient Bayesian inference for complex survival analysis. I have been an invited speaker to various international conferences like ISI 2019 (<https://www.isi2019.org/scientific-programme-2/> IPS97), IBC 2022 (<https://www.ibc2022.org/events/is22> IS16), IBC 2024 (<https://www.ibc2024.org/events/showcase>) and Geomed 2024 (<https://www.uhasselt.be/en/events-en/2023-2024/geomed2024/material> IS6), and to various research groups. I was also an invited speaker at SASA 2021 in the Biostatistics stream (<https://uctcmc.eventsair.com/the-62nd-annual-conference-of-the-south-african-statisticalassociation/programme>). I am invited to the 2025 O'Bayes meeting (<https://obayes25.aueb.gr/>) as a discussant of an invited paper.

The trajectory of my research, as well as its potential for future impact, is evident from my invited talks and courses, awards, citations and publications, and editorial board duties. I am excited for my future in research and hope to make a lasting tangible impact in society.

2. Summary of completed research

I have made significant contributions to theoretical, computational and applied aspects of Bayesian statistics, with particular focus on biostatistics and epidemiology. My research spans prior specification, computational methods, and applications of complex hierarchical Bayesian models, with impactful advancements in both exact and approximate Bayesian inference.

The majority of my publications are in Q1 (52%) and Q2 (17.4%) journals while three works (13%) are published in one of the top 25 ranked journals in Statistics and Probability.

My completed research can be summarized broadly into four main contribution areas.

1. Prior Specification: Challenging Conventional Approaches

Incorporating prior knowledge into Bayesian models is fundamental to the Bayesian workflow, typically through prior distributions on unknown parameters. Historically, priors have been classified as either informative or uninformative, with conjugate priors often used as defaults for informative settings and flat priors assumed in the absence of prior knowledge. My work has challenged these conventions, particularly the default use of conjugate priors and flat priors.

The traditional preference for conjugate priors originated in the pre-computing era due to their mathematical tractability. However, with modern computational tools, this default choice is no longer necessary. My research has demonstrated that conjugate priors can exhibit undesirable properties and unexpected poor behavior in certain cases (e.g., [1, 4, 5, 6]).

Additionally, I have shown that default flat priors can lead to over-parameterized models, especially for complex models like the Weibull survival model or the skew-normal distribution. For example, flat priors can prevent the model from converging to the simpler version (e.g., exponential or normal distributions), even when the data supports the simpler structure. To address this, I developed penalizing complexity (PC) priors based on the proposal by Simpson et al. (2017). These priors contract toward simpler models with exponential tails and provide robust defaults, favoring parsimonious models unless the data suggests otherwise (e.g., [3, 7, 8]). This approach ensures more reliable and interpretable posterior inference, particularly in complex modeling scenarios like skewed probit regression and survival joint models.

2. Advancing Computational Methods: From Exact to Approximate Inference

Posterior inference, the ultimate goal of Bayesian analysis, can be achieved through exact or approximate methods. During my PhD, I developed matrix-variate Gibbs samplers for covariance matrices, enabling the practical application of my methodologies (e.g., [1]). However, exact methods can be computationally expensive, particularly for high-dimensional models or large datasets. To overcome these limitations, I turned to approximate Bayesian inference and have since made groundbreaking contributions in this area.

The Laplace method's (in machine learning literature it is termed the Laplace approximation) Gaussian approximation is derived by determining the mode, and the Hessian matrix at the mode as an estimate of the curvature. For skewed or heavier-tailed densities the mode might not be the best estimate of the mean and a correction for this is thus needed. Other multivariate considerations cannot be deployed at scale (complex spatio-temporal models or huge data) and the idea was to push the Gaussian as far as possible.

One of my most notable developments is a low-rank Variational Bayes correction framework. This framework refines a computationally efficient but crude Gaussian approximation (derived via the Laplace method) to achieve the accuracy of sampling-based methods, but with negligible additional computational cost (e.g., [9]). This work provides the most accurate and scalable high-dimensional Gaussian approximation currently available, making it ideal for Bayesian machine learning and other fields requiring efficient inference. The framework has been shown to rival the accuracy of sampling-based methods at a fraction of the computational cost, sometimes achieving hundred-fold speedups.

3. Development of INLA 2.0

Building on my contributions to Gaussian approximations, I collaborated with Prof. Håvard Rue to reformulate a key step in the Integrated Nested Laplace Approximation (INLA) framework. INLA, introduced by Rue et al. (2009), is a popular methodology for approximate Bayesian inference in latent Gaussian models, with an R package boasting over 5,800 citations. My reformulated approach, termed INLA 2.0, leverages my work on Gaussian approximations ([9]) to provide even more efficient and accurate inference for latent Gaussian models (e.g., [10]), and has been cited 64 times in the first year. This innovation enables a more efficient application of INLA to complex models, including non-linear generalized mixed effects models, spatial and spatio-temporal models, and survival joint models. My recent work with a former PhD student has also introduced a novel gradient-based optimization method, further improving the performance of INLA (e.g., [11]).

4. Applications in Biostatistics and Epidemiology

Hierarchical Bayesian models are fundamental in biostatistics and epidemiology, particularly for survival analysis and joint modeling in clinical trials and observational data. In 2018, efficient Bayesian inference for such models was limited. I addressed this gap by developing a framework that reformulates many survival and joint models as latent Gaussian models, enabling inference through the INLA methodology (e.g., [12, 13, 14, 15]). I also co-developed the INLAjoint R package to facilitate user implementation of these models (<https://cran.rproject.org/web/packages/INLAjoint/index.html>).

These advancements have significant implications for public health, epidemiology, and clinical trial analysis. For example, the new INLA 2.0 framework supports efficient inference in spatio-temporal models, aiding in disease mapping, environmental research, and renewable energy studies (e.g., [16, 17, 18]). My work has been widely cited across disciplines, reflecting its broad applicability and impact.

References.

1. Bekker, A., Van Niekerk, J. and Arashi, M. (2017). Wishart distributions: Advances in theory with Bayesian application. *Journal of Multivariate Analysis*, 155, pp.272-283.
2. Rue, H., Martino, S. and Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2), pp.319-392.
3. Simpson, D., Rue, H., Riebler, A., Martins, T.G. and Sørbye, S.H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1), pp.1-28.
4. Van Niekerk, J., Bekker, A. and Arashi, M. (2017a). A gamma-mixture class of distributions with Bayesian application. *Communications in Statistics-Simulation and Computation*, 46(10), pp.8152-8165.
5. Van Niekerk, J., Bekker, A., Arashi, M. and Roux, J.J.J. (2015). Subjective Bayesian analysis of the elliptical model. *Communications in Statistics-Theory and Methods*, 44(17), pp.3738-3753.
6. Van Niekerk, J., Bekker, A. and Arashi, M. (2017b). Weighted-type Wishart distributions with application. *REVSTAT-Statistical Journal*, 15(1), pp.205-222.
7. Van Niekerk, J., Bakka, H. and Rue, H. (2021). A principled distance-based prior for the shape of the Weibull model. *Statistical and Probability Letters*, 174, pp.98-109.
8. Van Niekerk, J., and Rue, H. (2021). Skewed probit regression – identifiability, contraction and reformulation. *REVSTAT-Statistical Journal*, 19(1), pp.1-22.
9. Van Niekerk, J. and Rue, H. (2024). Low-rank Variational Bayes correction to the Laplace method. *Journal of Machine Learning Research* 25(62), 1-25.
10. Van Niekerk, J., Krainski, E., Rustand, D., & Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 107692.
11. Fattah, E. A., Van Niekerk, J., & Rue, H. (2022). Smart Gradient-An adaptive technique for improving gradient estimation. *Foundations of Data Science*, 4(1), 123-136.
12. Van Niekerk, J., Bakka, H., & Rue, H. (2023). Stable non-linear generalized Bayesian joint models for survival-longitudinal data. *Sankhya A*, 85(1), 102-128.
13. Van Niekerk, J., Bakka, H., & Rue, H. (2021). Competing risks joint models using RINLA. *Statistical Modelling*, 21(1-2), 56-71.
14. Alvares, D., Van Niekerk, J., Krainski, E. T., Rue, H., & Rustand, D. (2024). Bayesian survival analysis with INLA. *Statistics in Medicine*, 43(20), 3975-4010.
15. Burger, D.A., van der Merwe, S., van Niekerk, J., Lesaffre, E. and Pironet, A., 2024. Joint quantile regression of longitudinal continuous proportions and time-to-event data: Application in medication adherence and persistence. *Statistical Methods in Medical Research*, DOI <https://doi.org/10.1177/09622802241300845>.
16. Fattah, E. A., Krainski, E., van Niekerk, J., & Rue, H. (2023). Non-stationary Bayesian Spatial Model for Disease Mapping based on Sub-regions. *Statistical Methods in Medical Research*, 33(6), 1093-1111.
17. Alahmadi, H., Van Niekerk, J. & Rue, H. (2023). Joint quantile disease mapping with application to malaria and G6P deficiency. *Royal Society Open Science* 11(1), 230851.
18. Rustand, D., Van Niekerk, J., Rue, H., Tournigand, C., Rondeau, V., & Briollais, L. (2023). Bayesian estimation of two-part joint models for a longitudinal semicontinuous biomarker and a terminal event with R-INLA: Interests for cancer clinical trial evaluation. *Biometrical Journal*, 65(4).

3. Future and current research

The following papers are submitted (as of March 2025):

1. Dutta, S., Van Niekerk, J. & Rue, H. (2025). Skewing the Laplace method using Variational Bayes beyond symmetric approximations. *Journal of Graphical and Computational Statistics*.
2. Liu, Z., Rue, H. & Van Niekerk, J. (2025). Leave-group-out cross-validation for latent Gaussian models, *Statistics and Operations Research Transactions*.
3. Ye, X., Van Niekerk, J. & Rue, H. (2025). Principled priors for Bayesian circular models. *Journal of the American Statistical Association*.
4. De Bie, M., Van Niekerk, J. & Bekker, A. (2025). Spatio-temporal insights for wind energy harvesting in South Africa. *International Journal on Geomathematics*.
5. Van Niekerk, J. and Rue, H. (2025). Restricting feedback in joint models for robust inference. *STAT*.
6. Van Wyk, A., Van Niekerk, J., Arashi, M. & Bekker, A. (2025). Bayesian Variable Selection for Skew Normal Models. *Journal of Multivariate Analysis*
7. Freni-Sterrantino, A., Rustand, D., Van Niekerk, J., Krainski, E. T., & Rue, H. (2025). A Graphical Framework for Interpretable Correlation Matrix Models. *Statistical Methods and Applications*.

Currently I am busy writing two books with leading experts in the fields of natural hazard modeling (L. Lombardo and H. Tanyas) and approximate Bayesian inference (H. Rue):

1. Lombardo, L., Freni-Sterrantino, A., Van Niekerk, J., Tanyas, H., Krainski, E., Castro-Camilo, D., Opitz, T and Chiuchiolo, C. (2025). *Temporal and Spatial modeling of Landslide hazards and other natural disasters*. Wiley.
2. Rustand, D., Van Niekerk, J., Krainski, E.T. and Rue, H. (2025). *Advanced survival analysis with INLA*. CRC Press.

and an invited book chapter:

Van Niekerk, J. and Rue, H. (2025). Efficient spatio-temporal Bayesian modeling with INLA. In *Environmental Statistics: Innovative Methods and Applications*, CRC Press.

Efficient and principled Bayesian inference is a main aim of my research. I want to refine the INLA methodology to be even faster and (possibly) more accurate by developing a feedback mechanism for the reformulation of the latent field (based on work of the last three years and published) to the hyperparameter level. I am collaborating with colleagues and students from KAUST on this project.

Ongoing and future research projects can be classified broadly into the following five categories:

1. Robust default priors

I want to develop more contraction (penalizing complexity) priors to replace “non-informative” default priors in Bayesian models to ensure robust properties and contraction to more parsimonious models if the data necessitates it. This is a fundamental area that I have published in, and with a KAUST PhD student, Mr. Xiang Yi, we are currently developing contraction priors for directional data for robust Bayesian inference.

2. Complex survival analysis and joint models for precision medicine

In survival analysis, the survival probability under different treatment regimens is used to inform the treatment a patient should be described. These predictions rely on present (and past) biomarker data and thus need to be accurate and available almost immediately since the underlying disease mechanism changes while the patient waits for the result. As such, I want to keep working on fundamental models and efficient (near real-time) inference architectures to reach the goal of personalized medicine. Here, the development of differential privacy approaches will be important.

3. Bayesian models for genomic and precision medicine

The need to incorporate genomics information into precision medicine models is obvious. The efficient inference of these models, however, is not. Such high-dimensional covariate spaces pose many challenges and as such dimension-reduction techniques should be considered. These dimension reduction techniques cannot be inferred with the INLA methodology at this current time since they are inherently not latent Gaussian models. A current PhD student, Mr. Arno van Zyl, is developing a framework where Variational Bayes optimization can be combined with INLA for such conditional latent Gaussian models. This would allow fast inference and personalized prediction based on huge datasets.

4. Epidemiology and environmental statistics

Disease mapping is an important tool for policy-makers and researchers. Larger data in terms of higher resolution spatial data and richer temporal domains provide us with a challenge due to computational bottlenecks associated with identifiability constraints. The constraints are part of the well-known and often used intrinsic models such as random walk and CAR models for areal data. I plan to develop new spatio-temporal models (non-intrinsic), that would not need identifiability constraints. More complex structures, such as non-stationarity in the covariance, physical barriers (a current KAUST PhD student's focus, Ms. Martina Le-Bert Heyl) and different neighborhood structures, nonuniform range or smoother processes can also be considered.

5. Bayesian machine learning

Bayesian deep learning is a relatively new field with many possibilities. Based on my work correcting the Laplace method with Variational Bayes, I want to develop Bayesian machine learning frameworks, such as Bayesian convolutional neural networks with contraction priors. A current PhD student at UP, Mr. Ricardo Daniel Marques, is focusing on this topic.

In conclusion, there are many research opportunities in the field of approximate Bayesian inference and these aforementioned projects will generate various new research opportunities.