

Phillip John Paine, PhD.: Résumé

E-mail: pmxpp88@gmail.com

Based: Vancouver, British Columbia

Data Scientist with 6+ years of industry experience and 4 years post-doctoral research in statistics and probability. Industry work includes statistical modelling, BI reporting, and natural language processing at an enterprise insurance organisation, and time series forecasting and AI agent development at a supply chain startup.

Programming Languages: Python, R, Java and PySpark

Statistical Skills: Time Series Methods, Natural Language Processing and Bayesian Modelling

Data Science Tools: LangGraph+Langfuse, AWS and GCP, Docker and PostgreSQL

Professional Experience:

Alloy.ai

Senior Data Scientist in Forecasting and AI

Jan 2026 –

Data Scientist in Forecasting and AI

May 2022 – Dec 2025

Data scientist at a supply chain management startup, responsible for the full product lifecycle of forecasting models for demand planning, and more recently for developing analytics AI agent and the Agent Evaluation framework.

- **Architected and implemented the AI Agent Evals framework**, consisting of in-house LLM-as-a-judge pipeline and automatic trace classification. Enabling teams to move agent improvements from testing to production in weeks instead of months.
- **Developed the data analytics agent** used by vendors to get automatic summaries for their data across the AI products incorporating RAG, context management, and data citations and validations tooling.
- **Built, enhanced and maintained time series forecasting product** for vendor demand planning, implementing modelling features such as outlier detection, accounting for inventory out-of-stock periods, and an automated product grouping algorithm for forecasting new and low-velocity SKUs.
- **Designed and built the internal forecast model evaluation frameworks**, enabling faster experimentation and deployment of model improvements, as well as automated weekly monitoring of forecast accuracy for vendors

BCAA

Data Scientist – Underwriting Analytics

May 2020 – Apr 2022

Senior Data Analyst – Underwriting Analytics

May 2019 – May 2020

Analyst/Data scientist on the underwriting analytics team for residential insurance. Responsibilities included end-to-end delivery of data science projects using BI infrastructure, predictive modelling, and statistical forecasting in policy pricing, claims, and executive reporting.

- **Designed and deployed BI reporting infrastructure** across insurance products, combining PySpark-based data processing and hierarchical clustering, AWS Glue for data management, and PowerBI to deliver KPIs visibility to management and C-suite.
- **Developed a claims early detection model** using word-embedding encoders with CNN and LSTM layers in Keras-TensorFlow, deployed via AWS Lambda and Elastic Container Service. Providing the underwriting team with earlier identification of clusters of claims.
- **Improved homeowner policy pricing** by forecasting future policy volumes with an ARIMA model combined with k-Nearest Neighbours and GBMs to capture claims patterns, presenting business value to exec and finance teams via an interactive Shiny dashboard.

University of Sheffield, UK

Statistical Consultant

Sept 2016 – Dec 2018

Worked with applied science and engineering researches to provide short-term statistical consulting focused on uncertainty quantification. Example project:

Bayesian Analysis of Clonogenic Survival Assay Data

Created a dose-response curve with uncertainty contours from data on the irradiation of cancer cells in Python and R. A Bayesian hierarchical model was used to create a heat map around the dose-response mean curve quantifying the level of uncertainty for multiple sources of error. The outcome was a paper published in the Journal of Radiotherapy and Oncology.

University of Nottingham, UK

Post-Doctoral Fellow in Statistics

Sept 2014 – Sept 2016

The purpose of the work was to develop novel regression methods for manifold-valued data and landmark data. Applications of the work include predicting weather patterns and modelling vector-cardiogram signals.

Education:

University of Nottingham

PhD Statistics

Thesis: Parameter Estimation in Stochastic Differential Equations for Discretely-Sampled Observations

Thesis research was at the intersection of statistics, probability theory and mathematical finance.

Cardiff University

BSc Mathematics

First Class Honours

Other Experience:

University of Nottingham, UK

Lecturer for Topics in Statistics

Lecturer for a final-year undergraduate and masters class that covered hypothesis testing in sequential analysis, PCA and experimental design methods.