Jean-Baptiste Pettit

Palo Alto, CA linkedin.com/in/jbogp github.com/jbogp jb.pettit@gmail.com 650-283-8139

+ EDUCATION

2010 – 2014	**University of Cambridge** - PhD in statistics/bioinformatics
2009 – 2010	**INP-ENSEEIHT** (French Grande Ecole) - Specialized master's degree in computer science
2007 – 2010	**INP-ENSAT** (French Grande Ecole) - Master’s degree in life sciences engineering

+ PROFESSIONAL EXPERIENCE

2015-Present **Insight Data Science, Data Engineering Fellow** – Palo Alto, CA

- Designed and developed a data pipeline, API and website for news comments aggregation from multiple sources including RSS and Twitter Streaming feeds.

- Implemented “Big Data” pipeline to crawled news articles comments and Tweets using Spark for batch processing, Storm for Streaming, Hbase as a data store, Kafka for managing queues deployed on an AWS infrastructure.

2014-Present **ITK - Statistical modeling for agronomy, Developer** – Montpellier (France)

- Participated and delivered the first release of a Decision Support Tool for agriculture to one of the world's top crop science companies.

2010-2014
4 years **University of Cambridge, PhD student** – Cambridge (UK)

- Developed a Hidden Markov Random Field clustering methodology for Genomics data (10 GB of data), published in PloS Computational Biology. Open sourced on Github

- Created of a statistical spatial localisation methodology for single cell transcriptomics assays (100 GB of data). Accepted in Nature Biotechnologies (in press). Open sourced on Github

- Released a webGL/Javascript web application for scientific 3D data visualization (published in BMC Bioinformatics). Open sourced on Github

2010
6 months **European Bioinformatics Institute, Intern** – Cambridge (UK)

- Contributed in a JavaEE application for Systems Biology Format Conversion. Deployment of a JSP based web application for SBFC. Open sourced on SourceForge

2009
3 months **CNRS, Intern** – Sisyphe Laboratory – Paris (France)

- Modeled (ODE based) of N_2O emissions by agricultural soils. Computed GIS indexes in ArcGIS

+ SKILLS

Statistics	Bayesian hierarchical models, Markov chains/random fields models, generalized linear mixed models, ODE models, Mixture models, Multivariate analysis.
Programming	Functional (Scala), Object Oriented (JavaEE, Python), Imperative (C), Scientific (R, Matlab)
Tools	Hadoop Map Reduce, Spark, Spark Streaming, Pig, Hive, Kafka, AWS, Hbase, Cassandra, MySQL
languages	French (native), English (fluent), German (proficient), Spanish (basic), Arabic (basic)

+ SIDE PROJECTS

Blogging - Paid blogger for 1 year for the leading French Tech blog Presse-citron.net

- Blogging targeted at the Hacker News community (jbogp)

Twitter - Created a series of Twitter robots (@WhoTrendedit) analysing trending topics on Twitter and linked to a website (http://whotrended.it)

Freelance - Created the back-office employee/database/Ebay interaction system for a leading library in Paris