Stéphane Ghozzi

Researcher interested in supporting public health preparedness and response through automated analyses and visualizations: application and evaluation of machine-learning approaches, development and deployment of interactive dashboards. Moved from fundamental theoretical physics to biophysics and now infectious-disease epidemiology. Loves assembling interdisciplinary teams, comfortable as independent analyst and developer.

stephaneghozzi.com · orcid.org/0000-0002-3911-9573
gitlab.com/stephaneghozzi · twitter.com/stephaneghozzi
linkedin.com/in/stephaneghozzi · scholar.google.com/citations?user=uGVLwREAAAAJ

Experience

Research associate

Helmholtz-Zentrum für Infektionsforschung (HZI)
June 1, 2020 – present · 1 yr 10 mos
Brunswick (Braunschweig), Germany
helmholtz-hzi.de/en/research/research-topics/bacterial-and-viral-pathogens/epidemiology/our-research/ · sormas.org
Senior scientist, statistical methods and tools for public health, in the Epidemiology department (Krause group).

Surveillance Outbreak Response Management and Analysis System (SORMAS): Automated support for decision making by SORMAS users. Strategies to bring computations to data so as to allow for advanced analytics without compromising data privacy.

Effect of climate change on infectious disease spread: Statistical modelling of the impact of short and long-term environmental and demographic changes on Lyme disease spread in Germany.

Machine learning for multiplex serology.

Management and organization:
· contact person for topics of automated analyses and processes ;
· student support ;
· writing of grant applications.

Robert-Koch Institut (RKI)
April 15, 2016 – May 31, 2020 · 4 yrs 2 mos
Berlin, Germany
rki.de/signale-project
Machine learning, informatics, statistics and visualizations, Signale team, Department of Infectious-Disease Epidemiology.

Research and development:
· development of outbreak-detection algorithms and modelling of infection dynamics;
· performance evaluation and parameter optimization of the algorithms;
· natural language processing of online articles to support international infectious-disease surveillance;
· interactive visualizations of the results and other data for public-health professionals.

Management and organization:
· speaker for the team (four data scientists, two web developers, an average of two students);
· coordination of the international Topic Group Outbreaks under the Focus Group AI for Health of ITU and WHO: itu.int/en/ITU-T/focusgroups/ai4h/Pages/tg.aspx
· supervision of master theses;
· organization of workshops and hackathons;
· writing of grant applications.

World Health Organization (WHO)
May 1, 2019 – October 31, 2019 · 6 mos
Geneva, Switzerland
who.int/eios · who.int/emergencies/outbreak-toolkit
Machine learning and web-application development for epidemic intelligence and investigation of outbreaks of unknown origins.

In the Detection, Verification and Risk Assessment (DVA) and the Health Operations Monitoring and Data Collection (MDC) units of the Health Emergency Information and Risk Assessment (HIM) department within the WHO Health Emergencies (WHE) program of WHO.

Visual artist **self employed**
March 1, 2012 – April 14, 2016 · 4 yrs 1 mo
Berlin, Germany
Drawing, generative animation, photography, video, computer animation.

Selected projects and collaborations:
· 2014: videos exhibited during the *backup* festival, E-Werk, Weimar;
· 2005: twelve illustrations for *Trace.project*, a compilation album of original electronic music;
· 2005: videography for the dance piece *Entre-Deux* by Mirjam Fruttiger, Paris and Rome (including one-week invitation at the Villa Médicis);
· 2004: videography on the documentary *Manchay Tiempo* by Florence Blum and María Pía Medina-Luna (four-weeks filming in Peru);
· 2002: drawings and photographs published in the magazine *R de réel*.

Postdoctoral researcher **Institut für Theoretische Physik (THP), Universität zu Köln**
March 1, 2010 – February 29, 2012 · 2 yrs
Cologne, Germany
www.thp.uni-koeln.de/~lassig
Statistical and mechanical models of biological evolution, in the Lässig group.

Mathematical model and analysis of bacterial growth and gene expression, interpretation of experimental results.

Signatures of selection in DNA sequences and comparison with population genetics models:
· long-term influenza evolution;
· transcription-binding-site motifs in yeast.

Co-evolution signatures in protein sequences.

Teaching assistant

Universität zu Köln
September 1, 2010 – January 30, 2012 · 1 yr 5 mos
Cologne, Germany
Mathematics and statistical physics for Bachelor students.

UPMC Sorbonne Universités
October 1, 2005 – August 31, 2009 · 3 yrs 11 mos
Paris, France
Thermodynamics, optics and waves, mathematical methods for Bachelor students.

Doctoral researcher

Laboratoire de Physique Statistique (LPS), École normale supérieure
September 1, 2005 – December 31, 2009 · 4 yrs 4 mos
Paris, France
www.labos.upmc.fr/ljp/?article7 · www.lps.ens.fr
Theoretical and experimental biophysics, in the Chatenay group: Dynamics of gene regulatory networks.

Won a 60 k€ grant to fund the experimental project (over 3 years, used to buy apparatus and consumables): program "Interface physique, biologie et chimie : soutien à la prise de risque 2007-2009" of the CNRS.

Time series of gene expression levels, via fluorescent microscopy, of the lysis-lysogeny decision network of the bacteriophage Lambda:
· molecular biology (extraction of viral genes, insertion of genes coding for fluorescent protein, modification of bacterial genomes);
· microbiology (bacterial and viral cultures);
· automation and fluorescent microscopy;
· image analysis.

Mathematical analysis of noise statistics of bacterial gene expression.

Computer simulations of the dynamics and evolution of gene regulatory networks.

Volunteering Founding member

Celsius
August 1, 2007 – August 31, 2009 · 2 yrs 1 mo
Paris, France
Celsius was a think tank with the goal of developing the European project.

· Elaboration of background documents on technical themes;
· preparation, organization of, and follow-up on, two-days meetings in Madrid, Brussels, and Paris, each with more than 30 participants;
· development of publication strategies, print and online.

Internships	**Laboratoire de l'Accélérateur Linéaire (LAL), Université Paris-Sud**

February 1, 2005 – March 31, 2005 · 2 mos
Orsay, France
Theoretical particle physics, in the Davier group: Computation of fundamental physical quantities for high energy physics from particle collision experiments data.

Deutsches Elektronen-Synchrotron (DESY), Humboldt-Universität zu Berlin
January 17, 2003 – August 31, 2003 · 7 mos
Zeuthen, Germany
Theoretical particle physics, in the Jegerlehner group: Analysis of experimental data with ad hoc and first-principle mathematical models of elementary particles.

Laboratoire Kastler Brossel (LKB), École normale supérieure
June 1, 2002 – August 31, 2002 · 3 mos
Paris, France
Experimental quantum physics, in the Grynberg group: Construction of an optical atom trap.

Education

Certifications **Coursera**
January 1, 2016 – March 31, 2016 · 3 mos
Practical Predictive Analytics: Models and Methods
Machine Learning
Data Manipulation at Scale: Systems and Algorithms

Ph.D. **École normale supérieure**
September 1, 2005 – December 31, 2009 · 4 yrs 4 mos
Paris, France
Theoretical and experimental biophysics: Dynamics of gene regulatory networks. See "Experience" above for details.

B.Sc. & M.Sc. **École normale supérieure**
September 1, 2001 – August 31, 2005 · 4 yrs
Paris, France
Theoretical and mathematical physics: specialization in particle physics and statistical physics.

Competitive fellowship.

Activities and societies: President of the Photography and Video-Making student associations in 2002 and 2003:
· presenting, defending, and managing budgets;
· initiation and support for members.

M.A.	**École nationale supérieure des arts décoratifs**
	September 1, 2003 – August 31, 2004 · 1 yr
	Paris, France
	CGI, 3d animation, post-production.

Skills

Programming	R, Python, LaTeX, Mathematica, Matlab, Processing (Java), SQL, Perl, C++, Shell
Tools	Git, Jira, Confluence, Team Foundation Server (agile management: Scrum, Kanban), SQL Server Management Studio, Photoshop, Illustrator, After Effects, Microsoft Office, 3ds Max, Blender
Languages	French (native)
	German (professional)
	English (professional)
	Japanese (basics)

Selected projects

contrastR	R package for simulating epidemics on contact network under resource-constrained contact-tracing prioritization strategies: gitlab.com/stephaneghozzi/contrastR
sormasdatagen	R package for generating synthetic data for the outbreak management software SORMAS: gitlab.com/stephaneghozzi/sormasdatagen
asmodee-trendbreaker-evaluation	R scripts for evaluating the ASMODEE algorithm on time series and NHS pathways: gitlab.com/stephaneghozzi/asmodee-trendbreaker-evaluation
wotva	Prototypical R Shiny dashboard and analyses for the investigation of outbreaks of unknown origins: gitlab.com/stephaneghozzi/wotva
eios-anomaly	Python notebooks for simple NLP and machine learning to identify epidemiologically relevant texts: gitlab.com/stephaneghozzi/eios-anomaly

infectious-disease-severity-dashboard	R Shiny dashboard to compute and visualize the severity of infectious diseases following the PISA recommendations of WHO: gitlab.com/stephaneghozzi/infectious-disease-severity-dashboard
signale-logo	Processing scripts to generate patterns used for the graphic identity of the Signale team of the Robert Koch Institute: gitlab.com/stephaneghozzi/signale-logo
correlation-dance	Processing scripts to visualize the movement correlations between dancers from motion capture data: gitlab.com/stephaneghozzi/correlation-dance

Publications

[1] A. Abbood and S. Ghozzi, "General Framework for Evaluating Outbreak Prediction, Detection, and Annotation Algorithms," *medRxiv*, 10.1101/2022.03.16.22272469, 2022. https://www.medrxiv.org/content/10.1101/2022.03.16.22272469v2

[2] B. Silenou, C. Verset, B. Kaburi, O. Leuci, J. Dörrbecker, S. Ghozzi, C. Duboudin, and G. Krause, "SORMAS-Stats: A Novel Statistical Application for Real-time Estimation of Epidemiological Parameters of Communicable Diseases using Contact-tracing Data," *Preprint*, 10.2196/preprints.34438, 2021. https://preprints.jmir.org/preprint/34438

[3] T. Jombart*, S. Ghozzi*, D. Schumacher, T. J. Taylor, Q. J. Leclerc, M. Jit, S. Flasche, F. Greaves, T. Ward, R. M. Eggo, E. Nightingale, S. Meakin, O. J. Brady, Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, G. F. Medley, M. Höhle, and W. J. Edmunds, "Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 10.1098/rstb.2020.0266, 2021. https://doi.org/10.1098/rstb.2020.0266

[4] M. Becker, M. Strengert, D. Junker, P. D. Kaiser, T. Kerrinnes, B. Traenkle, H. Dinter, J. Häring, S. Ghozzi, A. Zeck, F. Weise, A. Peter, S. Hörber, S. Fink, F. Ruoff, A. Dulovic, T. Bakchoul, A. Baillot, S. Lohse, M. Cornberg, T. Illig, J. Gottlieb, S. Smola, A. Karch, K. Berger, H.-G. Rammensee, K. Schenke-Layland, A. Nelde, M. Märklin, J. S. Heitmann, J. S. Walz, M. Templin, T. O. Joos, U. Rothbauer, G. Krause, and N. Schneiderhan-Marra, "Exploring beyond clinical routine SARS-CoV-2 serology using MultiCoV-Ab to evaluate endemic coronavirus cross-reactivity," *Nature Communications*, 10.1038/s41467-021-20973-3, 2021. https://doi.org/10.1038/s41467-021-20973-3

[5] A. Abbood, A. Ullrich, R. Busche, and S. Ghozzi, "EventEpi — A natural language processing framework for event-based surveillance," *PLOS Computational Biology*, 10.1371/journal.pcbi.1008277, 2020. https://doi.org/10.1371/journal.pcbi.1008277

[6] O. Stojanović, J. Leugering, G. Pipa, S. Ghozzi, and A. Ullrich, "A Bayesian Monte Carlo approach for predicting the spread of infectious diseases," *PLOS ONE*, 10.1371/journal.pone.0225838, 2019. https://doi.org/10.1371/journal.pone.0225838

[7] B. Zacher, A. Ullrich, and S. Ghozzi, "Supervised Learning for Automated Infectious-Disease-Outbreak Detection," *Online Journal of Public Health Informatics*, 10.5210/ojphi.v11i1.9770, 2019. https://doi.org/10.5210/ojphi.v11i1.9770

[8] F. Eckelmann, S. Ghozzi, and A. Ullrich, "Dashboards as strategy to integrate multiple data streams for real time surveillance," *Online Journal of Public Health Informatics*, 10.5210/ojphi.v11i1.9701, 2019. https://doi.org/10.5210/ojphi.v11i1.9701

[9] N. Sarma, A. Ullrich, H. Wilking, S. Ghozzi, A. K. Lindner, C. Weber, A. Holzer, A. Jansen, K. Stark, and S. Vygen-Bonnet, "Surveillance on speed: Being aware of infectious diseases in migrants mass accommodations - an easy and flexible toolkit for field application of syndromic surveillance, Germany, 2016 to 2017," *Eurosurveillance*, 10.2807/1560-7917.ES.2018.23.40.1700430, 2018. https://doi.org/10.2807/1560-7917.ES.2018.23.40.1700430

[10] L. Perfeito, S. Ghozzi, J. Berg, K. Schnetz, and M. Lässig, "Nonlinear Fitness Landscape of a Molecular Pathway," *PLOS Genetics*, 10.1371/journal.pgen.1002160, 2011. https://doi.org/10.1371/journal.pgen.1002160

[11] S. Ghozzi, J. Wong Ng, D. Chatenay, and J. Robert, "Inference of plasmid-copy-number mean and noise from single-cell gene expression data," *Phys. Rev. E*, 10.1103/PhysRevE.82.051916, 2010. https://doi.org/10.1103/PhysRevE.82.051916

[12] S. Ghozzi, "Expression Dynamics of a Genetic Regulatory Network: the Lysis/Lysogeny Decision of Bacteriophage Lambda," Theses, Université Pierre et Marie Curie - Paris VI, Dec. 2009. https://tel.archives-ouvertes.fr/tel-00515109

[13] S. Ghozzi and F. Jegerlehner, "Isospin violating effects in e^+e^- vs. τ measurements of the pion form factor $|F_\pi|^2(s)$," *Physics Letters B*, 10.1016/j.physletb.2004.01.021, 2004. https://doi.org/10.1016/j.physletb.2004.01.021