Curriculum Vitae Jan Graffelman

1 Personal data

Name and surname:	Jan Graffelman
Address:	Department of Statistics and Operations Research Universitat Politecnica de Catalunya Campus Nord Building C5, 2nd floor, office 212A Carrer Jordi Girona, 1-3 08034 Barcelona, Spain
Telephone:	+34-93-4011739 (work)
Email:	jan.graffelman@upc.edu
Internet homepage:	http://www-eio.upc.es/∼jan/
Place of birth:	Doetinchem, the Netherlands.
Nationality:	Dutch.
Researcher ID:	L-8056-2014
Orcid identifier:	0000-0003-3900-0780
Scopus identifier:	8346689000
Web of Science (17/08/2021):	publications 39; citations 549; H index 13; (Core Collection)
Google scholar (17/08/2021):	citations 867; H index 16; i10 index 23

2 Education

Dutch Doctorandus degree in Biology (1988)		Rijksuniversiteit Groningen, the Netherlands.
	Specialization:	Theoretical Population Genetics.
	Secondary subjects:	Quantitative Genetics
		First degree in teaching.
PhD in Statistics (2000, with Honors)		Universitat Politecnica de Catalunya, Spain.

3 Professional positions

Current posts:

- Associate professor in Statistics (since 2002), Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.
- Affiliate associate professor of Biostatistics (since 2021), Department of Biostatistics, University of Washington (UW), Seattle, USA.; previously visiting associate professor since 2016.
- Accredited as full professor by the Spanish National Agency for Quality Assessment and Accreditation (ANECA).

Previous positions:

2000-2002 Assistant Professor. In Statistics, UPC.
1997-2000 Visiting Professor. In Mathematics and computer courses, Universitat Pompeu Fabra (UPF) Barcelona.
1994-1997 Assistant Professor. For computer courses, UPF, Barcelona.
1992-1994 Grant in supercomputing at the University of Pompeu Fabra.

1989-1992 Help desk coordinator of the Computer Centre of the University of Groningen (civil service).
1988 Graduate assistant for programming courses. University of Groningen, The Netherlands.

4 Academic responsibilities

- Head of the Doctorate program in Statistics and Operations Research at UPC (since April 2019).
- Principal investigator of a UPC research group in Compositional Data Analysis (since 2015).
- Organizer of the Statistics seminar at UPC (2009-2019).

5 Publications

Most recent work:

1. Stilp, A.M. et al. (2021) A System for Phenotype Harmonization in the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine (TOPMed) Program. *American Journal of Epidemiology.* **190**(10) pp. 1977–1992. 10.1093/aje/kwab115.
2. Graffelman, J. (2021) Maximum likelihood estimation of the geometric niche preemption model. Accepted for publication in *Ecosphere*. Preprint at doi: https://doi.org/10.1101/2021.01.27.428381.
3. Graffelman, J. (2021) Compositional Biplots: A story of false leads and hidden features revealed by the last dimensions. In: Filzmoser, P., Hron, K., Martín-Fernández, J.A., and Palarea-Albaladejo, J. (eds). *Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn.* Springer International Publishing, Cham. pp. 83–99. doi: 10.1007/978-3-030-71175-7_5.
4. Terré, M., Ortuzar, I., Graffelman, J., Bassols, A., Vidal, M., and Bach, A. (2021) Using compositional mixed-effects models to evaluate responses to amino acid supplementation in milk replacers for calves *Journal of Dairy Science* **104**(7) pp. 7808–7819. doi: 10.3168/jds.2020-20035. https://www.sciencedirect.com/science/article/pii/S0022030221005294
5. Graffelman, J., and Ortoleva, L. (2021) A network algorithm for the X chromosomal exact test for Hardy-Weinberg equilibrium with multiple alleles. *Molecular Ecology Resources* **21** pp. 1547–1557. https://doi.org/10.1111/1755-0998.13373.
6. Galván-Femenía, I., Barceló-i-Vidal, C., Sumoy, L., Moreno, V., de Cid, R. and Graffelman, J. (2021) A likelihood ratio approach for identifying three quarter siblings in genetic databases. *Heredity.* **126**(3) pp. 537–547. doi: 10.1038/s41437-020-00392-8
7. Egozcue, J.J., Graffelman, J., Ortego, M.I. and Pawlowsky-Glahn, V. (2020) Some thoughts on counts in sequencing studies *NAR Genomics and Bioinformatics.* **2**(4):lqaa094. doi: https://academic.oup.com/na
8. Graffelman, J. (2020) Statistical tests for the Hardy-Weinberg equilibrium. *Wiley StatsRef: Statistics Reference Online.* doi: 10.1002/9781118445112.stat08274.

First author articles in scientific journals:

1. Graffelman, J. (2020) Goodness-of-fit filtering in classical metric multidimensional scaling with large datasets. *Journal of Applied Statistics.* **47**(11) pp. 2011-2024. doi 10.1080/02664763.2019.1702929.
2. Graffelman, J., Galván Femenía, I., de Cid, R. and C. Barceló i Vidal (2018) A log-ratio biplot approach for exploring genetic relatedness based on identity by state. *Frontiers in Genetics.* Volume 10, article 341. doi 10.3389/fgene.2019.00341

3. Graffelman, J., Pawlowsky-Glahn, V., Egozcue, J.J. Buccianti, A. (2018) Exploration of geochemical data with compositional canonical biplots. *Journal of Geochemical Exploration* **194** pp. 120–133. doi: 10.1016/j.gexplo.2018.07.014

4. Graffelman, J. and Weir, B.S. (2018) Multi-allelic exact tests for Hardy-Weinberg equilibrium that account for gender. *Molecular Ecology Resources.* **18**(3) pp. 461–473. doi: 10.1111/1755-0998.12748

5. Graffelman, J. and Weir, B.S. (2018) On the testing of Hardy-Weinberg proportions and equality of allele frequencies in males and females at bi-allelic genetic markers. *Genetic Epidemiology* **42**(1) pp. 34–48. doi: 10.1002/gepi.22079.

6. Graffelman, J., Jain, D. and Weir, B.S. (2017) A genome-wide study of Hardy-Weinberg equilibrium with next generation sequence data. *Human Genetics* **136**(6) pp. 727–741. doi: 10.1007/s00439-017-1786-7.

7. Graffelman, J. and Weir, B.S. (2016) Testing for Hardy-Weinberg equilibrium at bi-allelic genetic markers on the X chromosome. *Heredity* **116**(6) pp. 558–568. doi: 10.1038/hdy.2016.20.

8. Graffelman, J., Nelson, S., Gogarten, S.M. and Weir, B.S. (2015) Exact Inference for Hardy-Weinberg Proportions with Missing Genotypes: Single and Multiple Imputation. *G3 (Genes, Genomes, Genetics)* **5**(11), pp. 2365–2373. doi: 10.1534/g3.115.022111.

9. Graffelman, J. (2015) Exploring Diallelic Genetic Markers: The HardyWeinberg Package. *The Journal of Statistical Software* **64**(3): 1–23. http://www.jstatsoft.org/v64/i03/paper
doi: 10.18637/jss.v064.i03

10. Graffelman, J., Sánchez, M., Cook, S. and Moreno, V. (2013) Statistical inference for Hardy-Weinberg proportions in the presence of missing genotype information. *PLoS ONE* **8**(12): e83316. doi: 10.1371/journal.pone.0083316.

11. Graffelman, J. and Moreno, V. (2013) The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Statistical Applications in Genetics and Molecular Biology.* **12**(4), pp. 433–448. doi: 10.1515/sagmb-2012-0039.

12. Graffelman, J. (2013) Linear-angle correlation plots: new graphs for revealing correlation structure. *Journal of Computational and Graphical Statistics*, **22**(1), pp. 92–106.
doi: 10.1080/15533174.2012.707850.

13. Graffelman, J. (2010) The number of markers in the HapMap project: some notes on Chi-square and Exact tests for Hardy-Weinberg equilibrium. *The American Journal of Human Genetics.* **86**(5): 813-823. doi: 10.1016/j.ajhg.2009.11.019.

14. Graffelman, J. and Morales-Camarena, J. (2008) Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* **65**(2):77-84. doi: 10.1159/000108939.

15. Graffelman, J., Balding, D.J., Gonzalez-Neira, A., and Bertranpetit, J. (2007) Variation in estimated recombination rates across human populations. *Human Genetics* **122**(3-4):301-310. doi: 10.1007/s00439-007-0391-6.

16. Graffelman, J. and van Eeuwijk, F. (2005) Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. *Biometrical Journal* **47**(6): 863-879. doi: 10.1002/bimj.200510177.

17. Graffelman, J. (2005) Enriched biplots for canonical correlation analysis. *Journal of Applied Statistics* **32**(2): 173-188. doi: 10.1080/02664760500054202.

18. Graffelman, J. and Tuft, R. (2004) Site scores and conditional biplots in canonical correspondence analysis. *Environmetrics* **15**(1): 67-80. doi: 10.1002/env.629.

19. Graffelman, J. and Aluja-Banet, T. (2003) Optimal representation of supplementary variables in biplots from principal component analysis and correspondence analysis. *Biometrical Journal* **45**(4): 491-509. doi: 10.1002/bimj.200390027

20. Graffelman, J. (2001) Quality statistics in canonical correspondence analysis. *Environmetrics* **12**(5): 485-497. doi: 10.1002/env.481

21. Graffelman, J. and Hoekstra, R. F. (2000) A statistical analysis of the effect of warfare on the human secondary sex ratio, *Human Biology*, **72**, 3, pp. 433-445.

22. Graffelman, J., Fugger, E. F., Keyvanfar, K. and Schulman, J. D. (1999) Human live birth and sperm sex ratio compared, *Human Reproduction*, **14**, pp. 2917-2920.

Co-authored articles:

1. Blay, N., Casas, E., Galván-Femenía, I., Graffelman, J., de Cid, R. and Vavouri, T. (2019) Assessment of kinship detection using RNA-seq data. *Nucleic Acids Research* 47(21) e136. doi: https://doi.org/10.1093/nar/gkz776

2. Puig, X., Ginebra, J. and Graffelman, J. (2019) Bayesian model selection for the study of Hardy-Weinberg proportions and homogeneity of gender allele frequencies. *Heredity* 123, pp. 549-564 doi: https://doi.org/10.1038/s41437-019-0232-0

3. Puig, X., Ginebra, J. and Graffelman, J. (2017) A Bayesian test for Hardy-Weinberg equilibrium of bi-allelic X-chromosomal markers. *Heredity* **119**(4):226–236. doi: 10.1038/hdy.2017.30

4. Galván-Femenía, I., Graffelman, J. and Barceló-i-Vidal, C. (2017) Graphics for relatedness research. *Molecular Ecology Resources*, **17**(6), pp. 1271-1282. doi: 10.1111/1755-0998.12674

5. Lidón-Moyano, C., Martín-Sanchez, J.C., Saliba, P., Graffelman, J. and Martinez-Sanchez, J.M. (2016) Correlation between tobacco control policies, consumption of rolled tobacco and e-cigarettes, and intention to quit conventional tobacco, in Europe. *Tobacco Control* **26**(2):1–4. doi: 10.1136/tobaccocontrol-2015-052482.

6. Laayouni, H., Montanucci, L., Sikora, M., Mele, M., Dall'Olio, G.M., Lorente-Galdos, B., McGee, K.M., Graffelman, J., Awadalla, J.P., Bosch, E., Comas, D., Navarro, A., Calafell, F., Casals, F., Bertranpetit, J. (2011) Similarity in recombination rate estimates highly correlates with genetic differentiation in humans. *PLoS ONE* **6**(3): e17913. doi:10.1371/journal.pone.0017913.

7. Elena Bosch, E., Laayouni, H., Morcillo-Suarez, C., Casals, F., Moreno-Estrada, A., Ferrer-Admetlla, A., Gardner, M., Rosa, A., Navarro, A., Comas, D., Graffelman, J., Calafell, F., Bertranpetit, J. (2009) Decay of linkage disequilibrium within genes across HGDP-CEPH human samples: most population isolates do not show increased LD. *BMC Genomics*, 10:338.

Chapters in books:

1. Graffelman, J. and Galván-Femenía, I. (2016) An Application of the Isometric Log-Ratio Transformation in Relatedness Research. In Martín-Fernández, J.A. and Thió-Henestrosa, S. (eds.) Compositional Data Analysis, Springer Proceedings in Mathematics & Statistics, Vol 187, Springer International Publishing Switzerland. Pages 75–84. doi: 10.1007/978-3-319-44811-4. ISBN 978-3-319-44810-7.

2. Galván-Femenía, I., Graffelman, J. and Barceló-i-Vidal, C.(2016) A Compositional Approach to Allele Sharing Analysis. In Martín-Fernández, J.A. and Thió-Henestrosa, S. (eds.) Compositional Data Analysis, Springer Proceedings in Mathematics & Statistics, Vol 187, Springer International Publishing Switzerland. Pages 63–73. doi: 10.1007/978-3-319-44811-4. ISBN 978-3-319-44810-7.

3. Graffelman, J. (2013) Factor Analysis, in Encyclopedia of Environmetrics, A.-H. El-Shaarawi and W. Piegorsch (eds), John Wiley & Sons Ltd: Chichester, UK. doi: 10.1002/9780470057339.vaf001.pub2.

4. Graffelman, J. (2011) A universal procedure for biplot calibration. In Ingrassia, S. Et al. (eds.) New Perspectives in Statistical Modeling and Data Analysis, Studies in Classification, Data Analysis and Knowledge organization. Springer-Verlag Berlin Heidelberg. Pages 195–202. DOI 10.1007/978-3-642-11363-5-22. ISBN 978-3-642-11362-8.

5. Graffelman, J. and Egozcue, J. J. (2011) Hardy-Weinberg equilibrium: a nonparametric compositional approach. In Pawlowsky-Glahn, V. and Buccianti A., editors, Compositional Data Analysis: Theory and Applications, pages 208-217, John Wiley & Sons, Ltd. ISBN: 978-0-470-71135-4

6. Graffelman, J. (2001) Factor Analysis. In El-Shaarawi, A.H. and Piegorsch, W. W., editors, Encyclopedia of Environmetrics. Volume 2, pages 763-767, Chichester. John Wiley & Sons, Ltd.

Book reviews:

1. Graffelman, J. and Martín-Fernández, J.A. (2019) Book review: Applied Compositional Data Analysis: With Worked Examples in R, P. Filzmoser, K. Hron, M. Templ (2018). Springer Series in Statistics, 280 pages. ISBN: 978-3-319-96420-1 *Biometrical Journal* **62**(1) pp. 250–252. doi: 10.1002/bimj.201900263

2. Graffelman, J. (2019) Book review: Compositional Data Analysis in Practice. Michael J. Greenacre (2018). London: CRC Press. 136 pages, ISBN: 978-1-138-31661-4. *Biometrical Journal* **61**(4): 1088-1089. doi: 10.1002/bimj.201900080

3. Graffelman, J. (2014) J. Gower, S. Lubbe and N. le Roux, Understanding Biplots, John Wiley and Sons, 2011, pp. 463, ISBN 978-0-470-01255-0. *Journal of Classification* **31**:129-133. doi: 10.1007/s00357-014-9153-z

4. Graffelman, J. (2010) Book review: Biplots in Practice, Michael Greenacre. BBVA Foundation, Rubes Editorial. *Statistics and Operations Research Transactions* **34**(2).

Miscellanea:

1. Graffelman, J. (2000) Use of the Moore-Penrose Inverse in Canonical Correspondence Analysis, *Econometric Theory*, **16**, 5 pp. 792-793. Solution 99.5.1.

2. Graffelman, J. (1999) Use of the Moore-Penrose Inverse in Canonical Correspondence Analysis, *Econometric Theory*, **15**, 5, p. 777. Problem 99.5.1.

3. Graffelman, J. (1999) The Justification of Multidimensional Scaling under Euclidean Conditions, *Econometric Theory*, **15**, 6, pp. 908-909. Solution 99.1.5.

4. Graffelman, J. & van de Velden, M. (1999) Upper bounds for the eigenvalues of the product of a symmetric idempotent and a non-negative definite matrix. *Econometric Theory*, **15**, 4, p. 631. Problem 99.4.4.

5. Graffelman, J. (1998) A Fundamental Matrix Result on Scaling in Multivariate Analysis, *Econometric Theory*, **14**, 5, pp. 693-694. Solution 97.5.3.

6 Seminars

1. Log-Ratio biplots for exploring genetic relatedness based on Identity by State. Presentation for the GAC Local Analysis Meeting at the department of Biostatistics, University of Washington, Seattle, 19th of August 2019.

2. Log-Ratio biplots for exploring genetic relatedness based on Identity by State. Population Genetics lunch seminar at the department of Genome Sciences, University of Washington, Seattle, 12th of June 2019.

3. Log-Ratio biplots for exploring genetic relatedness based on Identity by State. Presentation at the department of Biostatistics, University of Washington, Seattle, 7th of May 2019.

4. Weighting variants in aggregate association tests for rare variants. Presentation for the TopMed local analysis group, at the department of Biostatistics, University of Washington, Seattle, 27th of August 2018.

5. Statistical Methods for testing Hardy-Weinberg Equilibrium. Seminar at the department of Statistics, Universidad Carlos III de Madrid, Madrid, 21st of February 2018.

6. Accounting for Gender in Statistical Tests for Hardy-Weinberg Equilibrium. Seminar at the department of Biostatistics, University of Washington, Seattle, 25th of May 2017.

7. The Statistics of Hardy-Weinberg Equilibrium. Statistics Lunch Seminar, Department of Statistics and Operations Research, Universitat PolitÃ¨cnica de Catalunya, 24th of February 2016.

8. The effect of missingness at SNP markers on tests for Hardy-Weinberg equilibrium. Invited seminar at the department of Biostatistics, University of Washington, Seattle, 20th of August 2015.

9. Testing for Hardy-Weinberg equilibrium at genetic markers on the X chromosome. Invited seminar at the department of Biostatistics, University of Washington, Seattle, 10th of August 2015.

10. Accounting for missing data in exact inference on Hardy-Weinberg equilibrium. Invited seminar at the Barcelona Biomedical Research Park (PRBB), Barcelona 25th of May 2015.

11. Hardy-Weinberg equilibrium and genetic association studies. Seminar at the department of Biostatistics University of Washington, Seattle, 25th of July 2014.

12. Detecting family relationships in genomic databases. Invited seminar at the Servei d'EstadÃstica Aplicada of the Autonoma University, Barcelona, 30th of January 2014.

13. The Statistics of Hardy-Weinberg Equilibrium. Statistics seminar, Department of Statistics and Operations Research, Universitat PolitÃ¨cnica de Catalunya, 20th of December 2013.

14. Graphical tests for Hardy-Weinberg Equilibrium: a compositional approach. Invited seminar at CNIO, Madrid, February 2009.

15. Visualization of Multivariate Data in the Life Sciences. Invited seminar at TNO, Zeist, April 2008.

16. Hardy-Weinberg Equilibrium and the Ternary Plot. Invited seminar at the Department of Informatics and Applied Mathematics of the Universitat de Girona, January, 2008.

17. Linkage Disequilibrium in Tomato. Seminar held at the department of Plant Sciences and the department of Genetics of the Wageningen Universiteit & Researchcentrum in Wageningen, February 2005.

18. Multivariate analysis of sensorial, biochemical and genetic data of the Tomato. Work presented at the Department of Genetics of the Wageningen Universiteit & Researchcentrum in Wageningen, September 2004.

19. Quality statistics in canonical correspondence analysis. Invited seminar at the Department of Statistics of the Universitat de Barcelona, in Barcelona, June 2001.

7 Presentations at conferences

Invited presentations:

1. Group comparison with count-based compositional data. Contribution to the organized invited session on Analyzing compositional, distributional and relative abundance data of the 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020, http://www.cmstatistics.org/CMStatistics2020/) UK, virtually, 19-21 December 2020.

2. Recent methodological developments for a Hardy-Weinberg analysis of X-chromosomal variants. Contribution to the Host Region Special Invited Session of the International Biometric Conference (IBC 2018), Barcelona 8-13 July, 2018. Jointly with Bruce S. Weir, Xavi Puig and Josep Ginebra.

3. Compositional canonical biplots. Contributed to the invited session Compositional data analysis in modern biology and ecology of the International Conference of the Royal Statistical Society (RSS 2017), Glasgow 4-7 September, 2017.

4. Natural factors distorting sex ratio in man and domestic animals. Invited contribution to the conference "Sex ratio and sex selection" organised by ESHRE (European Society of Human Reproduction and Embryology) in Maastricht, December 1997.

5. Man-made factors distorting sex ratio in man. Invited contribution to the conference "Sex ratio and sex selection" organised by ESHRE (European Society of Human Reproduction and Embryology) in Maastricht, December 1997.

Ordinary presentations:

1. Principal balances for metabolite genome-wide association analysis in the GCAT cohort. Contribution presented at the 8th International Workshop on Compositional Data Analysis (CoDaWork2019) Terrassa, 3-8 of June 2019. Jointly with Iván Galván-Femenía, D. Torrents, L. Sumoy, V. Moreno and R. de Cid.

2. Compositional canonical correlation analysis. Contribution presented at the 8th International Workshop on Compositional Data Analysis (CoDaWork2019) Terrassa, 3-8 of June 2019. Jointly with Vera Pawlowsky-Glahn, Juan José Egozcue and Antonella Buccianti.

3. Multidimensional scaling for relatedness research: an application of the Aitchison distance in the GCAT population based cohort. Poster presented at the 7th International Workshop on Compositional Data Analysis, Abbadia San Salvatore, Siena, Italy, 5-9 of June, 2017. Jointly with Iván Galvan-Femenía, Carles Barceló-i-Vidal, Laura Sumoy, Victor Moreno, and Rafael de Cid.

4. Statistical tests for Hardy-Weinberg equilibrium at X-chromosomal genetic markers. Poster presented at the American Society of Human Genetics Conference 2016, Vancouver, Canada, 18-22 of October, 2016. Jointly with Bruce S. Weir.

5. Statistical tests for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. Poster presented at the European Human Genetics Conference 2016, Barcelona, Spain, 21-24 of May, 2016. Jointly with Bruce S. Weir.

6. Graphical tools for estimating family relationship. Poster presented at the European Human Genetics Conference 2016, Barcelona, Spain, 21-24 of May, 2016. Jointly with I. Galván-Femenía, R De Cid and C. Barceló-i-Vidal.

7. IBS.IBD.studies: an R package for relatedness research using microsatellites and SNP data. Contribution presented at the 15th Spanish Biometry Conference, Bilbao, Spain, 22-25 of September 2015. Jointly with Iván Galván.

8. The Mid p-value and Exact Statistical Inference on Hardy-Weinberg Equilibrium. Contribution presented at the 15th Spanish Biometry Conference, Bilbao, Spain, 22-25 of September 2015.

9. A compositional approach to allele sharing analysis. Contribution presented at the 6th International Workshop on Compositional Data Analysis, L'Escala, Spain, 1-5 of June 2015. Jointly with Iván Galván and Carles Barceló.

10. Microbial Denitrification and the Hardy-Weinberg law. Contribution presented at the 6th International Workshop on Compositional Data Analysis, L'Escala, Spain, 1-5 of June 2015. Jointly with Rutger de Wit.

11. On the asymptotic distribution of proportions of multinomial count data. Contribution presented at the 6th International Workshop on Compositional Data Analysis, L'Escala, Spain, 1-5 of June 2015. Jointly with Juan José Egozcue and Maribel Ortego.

12. An application of the isometric log-ratio transformation in relatedness research. Contribution presented at the 6th International Workshop on Compositional Data Analysis, L'Escala, Spain, 1-5 of June 2015. Jointly with Iván Galván.

13. Biplots for Compositional Regression. Contribution presented at the 6th International Workshop on Compositional Data Analysis, L'Escala, Spain, 1-5 of June 2015. Jointly with Raimon Tolosana-Delgado.

14. Accounting for missing data in exact inference on Hardy-Weinberg equilibrium. Contribution presented at the XXXV Congreso nacional de Estadística e Investigación Operativa y IX Jornadas de Estadística Pública (SEIO 2015), Universidad Pública de Navarra, Pamplona, Spain, May 26-29, 2015.

15. Uncovering family relationships in genomic databases. Contribution presented at the XXXIV Congreso nacional de Estadística e Investigación Operativa y VIII Jornadas de Estadística Pública (SEIO 2013), Universitat Jaume I, Castellón, Spain, September 11-13, 2013, jointly with Elena Zanetto.

16. Graphical tools for exploring Hardy-Weinberg equilibrium. Contribution presented at the 14th Spanish Biometry Conference, Ciudad Real, Spain, 22-24 of May 2013.

17. Statistical inference for Hardy-Weinberg equilibrium with missing data. Contribution presented at the R User Conference 2013, Universidad de Castilla-La Mancha, Albacete, Spain, July 10-12, 2013.

18. Visualizing correlation matrices with R. Contribution presented at the R User Conference 2013, Universidad de Castilla-La Mancha, Albacete, Spain, July 10-12, 2013.

19. Graphically testing bi-allelic markers for Hardy-Weinberg Equilibrium. Poster presented at the European Human Genetics Conference 2013, Paris, France, 8-11 of June, 2013.

20. On the asymptotic distribution of isometric log-ratio coordinates. Contribution presented at the 5th International Workshop on Compositional Data Analysis, Vorau, Austria, 3-7 of June 2013.

21. Exploring bi-allelic genetic markers with the HardyWeinberg package. Contribution presented at the IV Jornadas de Usuarios de R, Centro de Investigación en Epidemiologia Ambiental (CREAL), Barcelona November 15-16, 2012.

22. Inferencia estadística para el equilibrio de Hardy-Weinberg en estudios de genotipado con Missing Data. Contribution presented at the IV Jornadas de Usuarios de R, Centro de Investigación en Epidemiologia Ambiental (CREAL), Barcelona November 15-16, 2012, jointly with Milagros Sánchez Mayor.

23. On the use of generalized canonical correlation analysis in genetics. Contribution presented at the 20th International Conference on Computational Statistics (COMPSTAT 2012) Limassol, Cyprus, August 27-31, 2012

24. Biplots with correlations that are linear in the angle. Contribution presented at the XXXIII Congreso nacional de Estadística e Investigación Operativa y VII Jornadas de Estadística Pública (SEIO 2012), Universidad Rey Juan Carlos, Madrid, Spain, April 17-20, 2012.

25. Biplot calibration with the calibrate package. Contribution presented at the R User Conference 2011, University of Warwick, Coventry, United Kingdom, August 16-18, 2011.

26. Graphical tools for assessing Hardy-Weinberg equilibrium for bi-allelic genetic markers. Contribution presented at the R User Conference 2011, University of Warwick, Coventry, United Kingdom, August 16-18, 2011.

27. On the canonical correlation analysis of bi-allelic genetic markers. Contribution presented at the 9th Tartu Conference on Multivariate Statistics and the 20th International Workshop on Matrices and Statistics, Tartu, Estonia, 26th June-1st of July, 2011.

28. Testing Hardy-Weinberg equilibrium, a compositional approach. Contribution presented at the 4th International Workshop on Compositional Data Analysis, Sant Feliu de Guíxols, Girona, 9-13 of May 2011.

29. Statistical inference for Hardy-Weinberg equilibrium using log-ratio coordinates. Contribution presented at the 4th International Workshop on Compositional Data Analysis, Sant Feliu de Guíxols, Girona, 9-13 of May 2011.

30. New pictures for correlation structure. Contribution presented at the 6th CARME Conference in Rennes, France, 8-11 of February 2011.

31. Diagnostic biplots for linear models. Contribution presented at the LinSTat Conference in Tomar, Portugal, 27-31 of July 2010.

32. A modified Chi-square test for Hardy-Weinberg equilibrium. Contribution presented at the 12th Spanish Biometry Conference in Cádiz, 23-25 of September 2009.

33. A universal procedure for biplot calibration. Contribution presented at the 7th Meeting of the Classification and Data Analysis Group of the Italian Statistical Society, Catania, 9-11 of September, 2009.

34. A global test for Hardy-Weinberg equilibrium. Contribution presented at the XXXI congreso nacional SEIO y V Jornadas de Estadísitica Pública, Murcia, February 2009.

35. Hardy-Weinberg equilibrium and the Ternary Plot. Contribution presented at the Compositional Data Analysis Workshop (CodaWork '08), Girona, 27-30 of May 2008.

36. Biplots in canonical correlation analysis, Contribution presented at the XXX congreso nacional SEIO y IV Jornadas de Estadísitica Pública, Valladolid, September 2007.

37. Automatic calibration of axis in biplots and scatterplots, Contribution presented at the 11th Spanish Biometry Conference in Salamanca, June 2007.

38. Linkage disequilibrium in Tomato, Contribution presented at the 10th Spanish Biometry Conference in Oviedo, May 2005.

39. Correspondence analysis with quantitative supplementary variables. Contribution presented at the conference Correspondence Analysis and Related Methods in Barcelona, July 2003.

40. Biplots with supplementary data, Contribution presented at the 9th Spanish Biometry Conference in A Coruña, May 2003.

41. Describing the Distribution of Species Counts by Poisson Mixtures, Contribution presented at the 8th Spanish Biometry Conference in Pamplona, March 2001.

42. Use of the Zero-inflated Poisson for Describing the Distribution of Species Abundance, Contribution to the second Spanish STATA users meeting in San Lorenzo del Escorial, May 2000.

43. A Distance-based Approach to Canonical Correspondence Analysis, Contribution presented at the 7th Spanish Biometry Conference in Mallorca, 1999.

44. Using Environmental Information in the Correspondence Analysis of Abundance Data. Contribution presented at the 6th Spanish Biometry Conference in Córdoba, September 1997.

45. Testing multivariate models on simulated environmental data. Contribution presented at the XVIIIth International Biometrics Conference in Amsterdam, July 1996, jointly with Michael Greenacre and Reinhold Fieler.

46. Statistical Methods for North Sea Environmental Survey Data. Contribution to a workshop at the Polar Institute, Tromso, May 1996, jointly with Michael Greenacre and Reinhold Fieler.

47. La nota Paau y su relación con la nota Cou. Un modelo de regresión con coeficientes aleatorios para el estudio del efecto centro en la nota Paau. Contribution to the 5th Spanish Biometry Conference in Valencia in June 1995, jointly with Anna Cuxart.

8 Direction of PhD's and Master theses

Directed PhD's:

1. Galván-Femenía, I. (2020) Compositional methodology and statistical inference of family relationships using genetic markers.

2. Morales-Camarena, J.G. (2009) Statistical tests for Hardy-Weinberg equilibrium and linkage disequilibrium: graphical methods in the presence of multiple markers.

Directed Master theses in statistics:

1. Mona Thiele (2020) A compositional study of biochemical and haematological factors involved in calf lameness.

2. Leonardo Ortoleva (2020) A network algorithm for exact tests for Hardy-Weinberg equilibrium with X-chromosomal variants.

3. Sergio Hilario Martínez Mateu (2020) Multivariate analyis of molecular data related to autoimmune disorders.

4. Eva Iranzo Ribera (2020) A genome-wide association study of cattle feed efficiency.

5. Iban Ortuzar Fernández (2019) Analysis of calf growth with compositional mixed models.

6. Evelyn Bustos Espinoza (2018) Modelización estadística de la deserción de tarjetas de crédito.

7. Marina Española Seral (2017) Health status of Spanish workers in 2015.

8. Genny Díaz Rodríguez (2016) Principal component analysis of bi-allelic genetic marker data.

9. Guillermo Montero Arias (2015) Unit-length vector analysis (UVA).

10. Ana Micallef (2015) Segmentación de usuarios de Internet.

11. Cristina Lidón Moyano (2015) Tobacco Consumption and Governmental Tobacco Control in Europe. Jointly with José María Martínez.

12. Elena Zanetto (2014) Análisis de relaciones de parentesco mediante el escalamiento multidimensional de datos genéticos.

13. Iván Galván Femenía (2014) Allele-Sharing Analysis and Relationship Inference.

14. Maria Vila Casadesús (2013) Analysis of miRNA-mRNA interactions in alcoholic hepatitis.

15. Milagros Sánchez Mayor (2012) Inferencia estadística para el equilibrio de Hardy-Weinberg en estudios de genotipado con datos faltantes.

16. Susana Pérez Álvarez (2010) Tests of Genetic Association for Quantitative Traits Based on both Mean and Variance. Jointly with David J. Balding

Directed bachelor's thesis in statistics or mathematics:

1. Gisela Espigulé Pons (2019) Relacions entre indexos de Biodiversitat ecológica. Co-directed with Josep Fortiana.

2. Blanca Rius Sansalvador (2019) Estudio de asociación entre polimorfismos genéticos y fuerza muscular de estudiantes.

3. Oriol Ventura Ripoll (2016) Modelització de la biodiversitat d'espécies mitjan cant la sèrie geométrica.

4. Anna Golanó Baix (2016) Responsabilitat Social Corporativa en les PIMES catalanes [in Industrial Engineering].

5. Cláudia Masós Torrá (2015) Maximum likelihood estimation of linkage disequilibrium.

6. Júlia Densalat Ventayol (2014) A genetic association study of colon cancer.

7. Silvia Juanes Márquez (2014) Estimación del coeficiente de endogamia por máxima verosimilitud.

8. Augustín Boyero Chacon (2012) Comportament financer dels catalans.

9. Lidia Millán Muro (2011) Percepción de la Salud en Aragón: una aplicación del Análisis de Correspondencias a la Encuesta Nacional de Salud 2006.

9 Participation in PhD juries

1. Fatine Ezbakhe (30/10/2019) Decision analysis under uncertainty for sustainable development. Universitat PolitÃ¨cnica de Catalunya (jury member).

2. Marc Comas Cufí (24/10/2018) Aportacions de l'anàlisi composicional a les mixtures de distribucions. Universitat de Girona (jury president).

3. Alba Martinez (27/01/2011) Patent value models: Partial least squares path modelling with mode C and few indicators. Universitat Politécnica de Catalunya (jury member).

4. Raquel Iniesta Benedicto (23/11/2010) Mètode Bayesià per l'anàlisi d'Halotips en estudis d'Associació Genètica. Universitat Autònoma de Barcelona (jury member).

5. Hector Rene Alvarez Laverde (27/11/2009) Análisis de valoraciones atípicas en los estudios de ingeniería Kansei: Consideraciones estadística y prácticas. Universitat Politècnica de Catalunya (jury member).

6. Carlos Alberto Juárez Alonso (03/04/2005) Fusión de datos: imputación y validación. Universitat Politècnica de Catalunya (jury member).

10 Participation in evaluations committees

1. Member of the evaluation committee for a position for associate professor at the University of Barcelona (25-26 of February 2021)

11 Participation in publicly funded research projects

1. Project RTI2018-095518-B-C22 Transfer and methodological development of compositional data-analytic techniques for applied sciences and engineering. From 01/01/2019 to 30/09/2022. As principal investigator. Funding: 25.800€.

2. Project 5R01GM075091-03 Theoretical Population Genetics. From 12/09/2017 to 30/06/2021. PI: Bruce Weir, University of Washington. Funder: National Institutes of Health. As participating researcher. Funding: 351.234€

3. Project MTM2015-65016-C2-2-R Transferencia de métodos de datos composicionales a las ciencias aplicadas y la tecnología (2015-2017). As principal investigator. Funding: 43.500€.

4. Project MTM2012-33236 Métodos estadísticos en espacios restringidos (METRICS) (2013-2015). PI: Josep Antoni Martín Fernández, University of Girona. As participating researcher. Funding: 77.220€.

5. Project 2014SGR551 COmpositional and Spatial Data Analysis [COSDA]. PI: Josep Antoni Martín Fernández, University of Girona. As participating researcher. Funding: 43.000€.

6. Project CODA-RSS MTM2009-13272 Analisis estadistico de datos composicionales y otros datos con espacio muestral restringido. PI: Josep Antoni Martín Fernández, University of Girona and Juan José Egozcue, Technical University of Catalonia. As participating researcher. Funding: 92.444€.

7. Project MCEI (ECO2011-28875) Métodos para modelos de ecuaciones estructurales no lineales y aplicaciones a las ciencias sociales y del comportamiento. PI: Albert Satorra, University Pompeu Fabra. As participating researcher. Funding: 38.115€.

8. Project MCYT (SEJ2006-13537) Modelado Multivariante con Variables Latentes: Datos de Diseño Complejo, Causalidad y Aplicaciones. From 01/10/2007 to 30/09/2011. PI: Albert Satorra, University Pompeu Fabra. As participating researcher. Funding: 138.080€.

9. Project MCYT (SEC2003-04476) Analisis Multivariante de Datos Longitudinales. From 10/12/2003 to 09/12/2006. PI: Albert Satorra, University Pompeu Fabra. As participating researcher. Funding: 64.400€.

10. Project DGICYT (BEC2000-0983) Modelización estadística con variables latentes: datos de diseño complejo. From 19/12/2000 to 19/12/2003. PI: Albert Satorra, University Pompeu Fabra. As participating researcher.

11. Project DGICYT (PB96-0300) Modelización multivariante de datos longitudinales (1996-1999).

12 Foreign research visits

1. Department of Biostatistics, University of Washington, May 2019-August 2019 Seattle, United States.

2. Department of Biostatistics, University of Washington, May 2018-August 2018 Seattle, United States.

3. Department of Biostatistics, University of Washington, Sabbatical September 2016-September 2017 Seattle, United States.

4. Department of Biostatistics, University of Washington, July-August 2015 Seattle, United States.

5. Department of Biostatistics, University of Washington, July-August 2014, Seattle, United States.

6. Department of Epidemiology and Public Health, Statistical Genetics group, Imperial College, London, United Kingdom. August 2005-January 2006.

7. Laboratory of Genetics, Wageningen Universiteit en Researchcentrum, Wageningen, The Netherlands, August 2004-January 2005.

8. Universidad de Baja California Sur. La Paz, México. August 1997.

13 Referee activities

Collaboration as anonymous reviewer (see also publons.com):

Annals of Human Genetics (2014), Austrian Journal of Statistics (2020), Behavior Research Methods (2012), Biometrical Journal (2018, 2×), Biometrics (2019), Bio-informatics (2005), BMC Bioinformatics (2016), BMC Genetics (2012), Collegium Antropologicum (2008), Computational Statistics and Data Analysis (2007), Current Pharmaceutical Design (2017), Demographic Research (2019), Entropy (2019), Environmental and Ecological Statistics (2009), Frontiers in Genetics (2021, 2×), Frontiers in Plant Science (2021), Gigascience (2018,2019), Global and Planetary Change (2012), Human Reproduction (2000,2003, 2004,2005,2007), Hydrobiologia (2010), International Journal of Biomedical Data Mining (2013), International Journal of Environmental Research and Public Health (2017), International Journal of Molecular Sciences (2016), Journal of Agricultural, Biological and Environmental Statistics (2002), Journal of Applied Statistics (2018, 2019, 2020), Journal of Biopharmaceutical Statistics (2018, 2020, 2021), Journal of Computational and Graphical Statistics (2008,2009), Journal of Ecology (2006), Journal of Geochemical Exploration (2017,2021), Journal of Statistical Software (2011, 2001), Mathematical Geology (2005), Medicina Clinica (2014), Microbiome (2015), Molecular Ecology Resources (2017,201), NAR genomics and bioinformatics (2020), Paediatric and Perinatal Epidemiology (2013), PLOS ONE (2016,2014), Pychometrika (2008,2011,2013), Scientia Marina (2008), Statistics and Operations Research Transactions (2005), Statistics in Transition (2021), The American Statistician (2008), Theory in Biosciences (2015), Trends in Genetics (2020).

14 Developed Software

Author of R packages:

- calibrate (version 1.7.6; June 2020)
- Correlplot (version 1.0.2; October 2013)
- HardyWeinberg (version 1.7.2; March 2021)
- MLpreemption (version 1.0.1; February 2021)
- ToolsForCoDa (version 1.0.5; June 2020)

15 Main recent teaching activities

- Coordinator and professor for the course *Multivariate Analysis* of the Master program in Statistics and Operations Research (MESIO 2011-2016,2018-2021 UPC).
- Coordinator and professor for the course *Data Analysis* of the bachelor degree in Data Science and Engineering (FIB 2018-2021 UPC).
- Coordinator and professor for the course *Statistical Models and Stochastic Processes* of the bachelor degree in Bio-informatics (2017-2021 inter-university UPF-UPC-UB).
- Professor and past coordinator for the course *Estadística per les Biociències* of the Bachelor program in Statistics (2013-2016,2018-2021, UPC).

- Past coordinator and professor for the course *Estadística* of the Bachelor program in Mathematics (2011-2016, UPC).
- Professor of the course *Bio-informatics and Statistical Genetics* of the Master in Innovation and Research in Informatics, and previously of the Erasmus Mundus Master in Data Mining and Knowledge Management (MIRI and DMKM, 2011-2021, UPC).
- Past professor of the course *Estadística* of the Bachelor's degree in Industrial Technology Engineering (2013-2016, UPC).

16 Short courses taught

- *Multivariate Analysis for Genetic Data.* Three-day online course at the 26^{th} Summer Institute in Statistical Genetics (SISG), University of Washington, Seattle, 21-23 of July, 2021.
- *Multivariate Analysis for Genetic Data.* Three-day online course at the 25^{th} Summer Institute in Statistical Genetics (SISG), University of Washington, Seattle, 29-31 of July, 2020.
- *Compositional Data Analysis.* One-day course at the International Biometric Conference in Barcelona, 8th of July, 2018.
- *Curs d'exploració i modelització estadística de dades massives.* IRTA, Caldes de Montbui, Spain, 18-19 of April, 2018.
- *Biplots in Practice.* Course at the BBVA foundation in Madrid, 4th of October, 2010. Jointly with M. Greenacre.
- *Correspondence Analysis in Practice using R.* Course at the Universitat Autònoma de Barcelona, 15, 16, 17th of June, 2010. Jointly with M. Greenacre.