

University of Groningen

## Statistical modelling of spatio-temporal dependencies in NGS data

Ranciati, Saverio

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2016

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Ranciati, S. (2016). *Statistical modelling of spatio-temporal dependencies in NGS data*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



university of  
 groningen

# Statistical modelling of spatio-temporal dependencies in NGS data

**PhD thesis**

to obtain the degree of PhD at the  
 University of Groningen  
 on the authority of the  
 Rector Magnificus Prof. E. Sterken  
 and in accordance with  
 the decision by the College of Deans.

This thesis will be defended in public on

Friday 10 June 2016 at 12.45 hours

by

**Saverio Ranciati**

born on 6th of June, 1988  
 in Perugia, Italy

**Supervisor**

Prof. E.C. Wit

Prof. C. Viroli

**Assessment committee**

Prof. E.R. van den Heuvel

Prof. M. van der Wiel

Prof. D. Husmeier

Prof. A. Montanari

# Contents

<b>Contents</b>	<b>iii</b>
<b>Chapter 1: Introduction</b>	<b>2</b>
1.1 Statistics in biological systems . . . . .	2
1.2 Thesis outline . . . . .	4
1.3 Chapter 2: Mixtures and graphs . . . . .	5
1.4 Chapter 3: Spatial dependency and beyond . . . . .	6
1.5 Chapter 4: Time dynamics and complex systems . . . . .	8
<b>Chapter 2: Spatio-temporal model for multiple ChIP-Seq experiments</b>	<b>10</b>
2.1 Introduction . . . . .	12
2.2 Model and Methods . . . . .	14
2.3 Simulation study . . . . .	19
2.4 Genome-wide assessment of differential roles for p300 and CBP in transcription regulation . . . . .	21
2.5 Conclusion and discussion . . . . .	24
Bibliography . . . . .	25
<b>Chapter 3: Mixture model with multiple allocations for clustering spatially correlated observations in the analysis of ChIP-Seq data</b>	<b>26</b>
3.1 Introduction . . . . .	28
3.2 Methods . . . . .	31
3.2.1 Model-based clustering with mixture model . . . . .	31
3.2.2 Multiple Allocation Mixture model . . . . .	32
3.2.3 Multiple Allocation Mixture model for ChIP-Seq Data . . . . .	35

3.2.4	Modelling spatial correlation with a CAR structure . . . . .	37
3.2.5	Conditional Auto-Regressive Multiple Allocation Mixture ( <i>CAR-MAM</i> ) . . . . .	39
3.3	Results . . . . .	39
3.3.1	Simulation study – Multiple Allocation Mixture ( <i>MAM</i> ) model . . . . .	39
3.3.2	Simulation study – <i>MAM</i> with Conditional Autoregressive model ( <i>CAR-MAM</i> ) . . . . .	41
3.3.3	p300 protein binding ChIP-Seq experiment . . . . .	42
3.4	Conclusions . . . . .	44
	Bibliography . . . . .	45

<b>Chapter 4: Bayesian Smooth-and-Match estimation of ODEs' parameters with quantifiable solution uncertainty</b>	<b>47</b>	
4.1	Introduction . . . . .	49
4.2	Model formulation . . . . .	51
4.2.1	Tools and notation . . . . .	51
4.2.2	Prior, likelihood and posterior distributions . . . . .	52
4.2.3	Mimicking the data: relationship with other methods . . . . .	54
4.2.4	Bayesian Smooth-and-Match . . . . .	56
4.3	Simulation studies . . . . .	57
4.3.1	Logistic population growth . . . . .	58
4.3.2	Lotka-Volterra . . . . .	63
4.3.3	HIV viral fitness . . . . .	67
4.3.4	FitzHugh-Nagumo system for neuron electrical activity . . . . .	71
4.4	Conclusions . . . . .	73
	Bibliography . . . . .	75

<b>Appendix Appendix A: Appendix</b>	<b>78</b>
--------------------------------------	-----------





*From where we stand the rain seems random.  
If we could stand somewhere else,  
we would see the order in it.*

TONY HILLERMAN