

# Alberto Castellini

October 30, 2017

## Research interests (interessi di ricerca)

**General topics:** predictive modeling, data analysis, statistical learning, optimization.

**Detailed topics:** dynamic system modeling, time-series segmentation, activity recognition, feature selection, clustering, integrative analysis of large-scale data, inference of regulation mechanisms.

**Main methods:** regularized linear and logistic regression, bayesian networks and hidden Markov models, genetic algorithms, difference equations and metabolic P systems, neural networks.

**Applications:** situation assessment/activity recognition for aquatic drones, plant disease forecasting from sensor data, analysis and time-series segmentation of immunological data, predictive breeding of maize performance, modeling and simulation of metabolic systems, genomic sequence analysis/comparative genomics.

## Personal data (dati personali)

- Nationality and citizenship: Italian
- Languages: Italian (mother tongue), English (advanced)
- E-mail: alberto.castellini@univr.it

## Academic and Professional History (attività accademica e professionale)

- Research associate - postdoc, Department of Computer Science, University of Verona, Italy (06/2016 - Present).

Project: “INTCATCH: Development and application of novel, integrated tools for monitoring and managing catchments”.

Project: “EXPO-AGRI - EXtra-field Plant Observation for monitoring and forecast of AGRicultural Infections”.

- Statistician I, Statistical Analysis and Mathematical Modeling Group, Technical R&D, GlaxoSmithKline Vaccines, Siena, Italy (03/2016 - 05/2016).
- Research associate - postdoc, Bioinformatics group, Institute for Biochemistry and Biology, University of Potsdam/Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany (04/2014 - 01/2016).

Project: “Regression methods for integrative analysis of large-scale data”. Teaching activities listed above.

- Research associate - postdoc, Center for BioMedical Computing (CBMC), University of Verona, Italy (01/2010 - 03/2014).

Projects: *i*) “Modeling and simulation of metabolic systems”; *ii*) “Comparative genomics by k-mer based indices”. Teaching activities listed above.

- Business intelligence consultant at Sdg group, Verona, Italy (05/2012 - 03/2014). Projects at Eli Lilly, Benetton, Miroglio, Italcementi, Cattolica assicurazioni.

- Internship, Italian Space Agency, marketing unit, Rome, Italy (09/2006 - 03/2007). Project: “Analysis of Space Technologies”.

### Teaching (attività didattica a livello universitario)

Lecturer at the University of Potsdam (Germany), Master in Bioinformatics:

- Course: Network and Profile Data Analysis (academic years 2014-15, 2015-16). Topics: graph theory, network motifs, community structures, graph comparison, similarity measures, clustering algorithms, multiple hypothesis testing, time-series segmentation.
- Course: Selected methods and techniques for systems biology and informatics (academic years 2014-15, 2015-16). Topics: kinetic modeling, time-course simulation and linear stability, bifurcations and metabolic control analysis, genome-scale network reconstruction, constraint-based modeling, flux balance analysis, metabolic pathway analysis, chemical reaction network theory.

Lecturer at the University of Verona (Italy), Master in Bioinformatics and Medical Biotechnology:

- Course: Algorithms and Languages for Bioinformatics - module “Languages” (academic years 2011-12, 2012-13, 2013-14). Topics: introduction to the main elements of Java, Matlab (bioinformatics toolbox) and Python (biopython).
- Course: Algorithms and Languages for Bioinformatics (academic years 2010-11). Topics: descriptive statistics, hypothesis testing, correlation, regression and variable selection, time series analysis, alignment algorithms, web services in bioinformatics, introduction to the main elements of Java, Matlab (bioinformatics toolbox) and Python (biopython).

Teaching assistant at the University of Verona, Faculty of Mathematical, Physical and Natural Science:

- 2009-10: Programming II - laboratory (Laboratorio di Programmazione II), bachelor in bioinformatics.
- 2009-10: Programming I - laboratory (Laboratorio di Programmazione I), bachelor in computer science.
- 2008-09: Programming - laboratory (Laboratorio di Programmazione), bachelor in applied mathematics.
- 2008-09: Algorithms and data structures - laboratory (Laboratorio di Algoritmi e Strutture Dati), bachelor in bioinformatics.
- 2007-08: Informational Methods - laboratory (Laboratorio di Metodi Informazionali), bachelor in bioinformatics.

### Education (formazione universitaria)

- Master in Business Intelligence e Knowledge Management, University of Verona, Faculty of Economics. Granted by Center for BioMedical Computing (11/2010 - 11/2011). Thesis: *Business Intelligence and Data Mining in BioMedical Research*.

- Ph.D., Computer Science, Graduate School of Sciences Engineering and Medicine, University of Verona, Italy. Supported by a 3-year state scholarship. (01/2007 - 12/2009).  
Thesis: *Algorithms and Software for Biological MP Modeling by Statistical and Optimization Techniques* (with Java software *MetaPlab*). Supervisor: V. Manca.
- Laurea in Informatica (Bachelor + Master in Computer Science) - 5 years (110/110 cum laude), University of Verona, Italy (09/2000 - 07/2006).  
Thesis: *Simulazioni computazionali di operazioni del "DNA Computing"*. Supervisor: V. Manca. Co-supervisor: G. Franco.

### Project activities (realizzazione di attività progettuale)

- *European project: "INTCATCH"* (Call: H2020 WATER 2015 <http://intcatch.eu>). Department of Computer Science, University of Verona, Italy (06/2017 - Present).  
*Description:* Monitoring of catchments by autonomous aquatic drones has recently gained increasing interest. In this context robotic boats must navigate rivers and lakes to acquire real-time data concerning important water parameters, such as dissolved oxygen and electrical conductivity. My activity in this project focuses on the development of tools for situation assessment and activity recognition from sensor data based on supervised and unsupervised methods. Multivariate time-series segmentation (e.g., Hidden Markov Models) and clustering (e.g., Gaussian Mixture Models) methods were employed, for instance, to support on-line decision making of drones/operators and off-line analysis of large datasets collected by drone sensors (*Castellini et al., 2017e*). These methods aim also to overcome some issues related to manual data labeling which is sometimes impracticable for real datasets.
- *Project: "EXPO-AGRI - EXtra-field Plant Observation for monitoring and forecast of AGRicultural Infections"* (<http://www.di.univr.it/?ent=progetto&id=4589>). Department of Computer Science, University of Verona, Italy (06/2016 - Present).  
*Description:* In agriculture, several plant pathogens, e.g., bacteria and fungi, can cause serious damages to crops, consumers and economy. Traditional greenhouse control systems react to threshold crossing events on each monitored parameter but they do not consider the global relationship between all these parameters and crop behavior. The aim of this project is to develop a greenhouse control system which directly maximizes crop yield and minimizes infection based on real time data acquired through a sensor network. The main focus of my research is threefold: the creation of mathematical models linking physical environmental parameters, plant grow, and disease development; the creation of prediction models of plant grow and disease development as a function of physical environmental parameters and weather forecast; the creation of a monitoring and control system fostering plant grow and reducing disease development with application to the binomia "Basil-Peronospora". Methodologies based on regularized regression and bayesian networks are being used to reach these targets (*Castellini et al., 2017b*).
- *Project: "Statistical modeling of immunological data"*. Department of Computer Science, University of Verona, Italy (01/2011 - Present).  
*Description:* Immunosenescence concerns the gradual deterioration of the immune system due to aging. Recent advances in cellular phenotyping have enabled key improvements in the analysis of biological mechanisms involved in this process during

the last decades. The goal of this project is to develop novel data-driven models for describing age-related changes in the network of relationships among different subpopulations of peripheral B lymphocytes. A dataset generated by the University Hospital of Verona (Italy) of about six thousands samples of patients between one day and ninety-six years has been analyzed by different statistical methods. My research activities involve the development of state-of-the-art and novel modeling methods, such as, difference equations/metabolic P systems and multivariate time-series segmentation for the analysis of these data (*Castellini et al., 2014a; Castellini et al., 2017c; Castellini et al., 2017d*).

- *Project:* “Regression methods for integrative analysis of large-scale data”. Bioinformatics group, Institute for Biochemistry and Biology, University of Potsdam/Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany (04/2014 - 01/2016).

*Description:* The problem is to pre-select the best plant crosses from a huge number of potential candidates, with the goal to increase the production of this crop to meet its rising demand for feed and biofuels. Regression and regularization techniques, such as LASSO and elastic net, were used and extended in order to generate models that integrate data from different biological layers with the aim to predict traits (e.g., biomass) of crosses from SNPs, metabolic and enzymatic profiles of parent plants. Prediction performance, model interpretability and generalizability are fundamental aspects in this context since a small number of biomarkers should be identified among a large set of candidates to predict the biomass of new plants across different environments (*Castellini et al., 2015d; Edlich-Muth et al., 2015e; Edlich-Muth et al., in prep.*).

From a mathematical and computational point of view this problem is challenging because it involves high-throughput datasets with thousands of variables, and a relative small number (usually hundreds) of observations. Model generation in these conditions is ill-posed and affected by instability, overfitting and multicollinearity. Various search strategies grounded in statistics, machine learning and optimization theory, can be used to select predictors and tune model parameters. Each methodology could select different biomarkers and generate models having different performance and interpretability. Interesting theoretical questions concern the analysis of the effect of data distribution, number of samples, data correlation structure and other parameters on model performance. Moreover, the transfer of knowledge from greenhouse experiments to field cultivation, and the development of network-based and multi-layer predictive models which reflect the multi-level structure of omics datasets, are other important challenges.

- *Project:* “Modeling and simulation of metabolic systems”. Center for BioMedical Computing (CBMC), University of Verona, Italy (01/2007 - 03/2014).

*Description:* This research line concerns the development of reverse engineering methods for the synthesis of metabolic and gene regulatory networks from time series. The final aim is to infer the topology and the regulative mechanisms of these networks from observed data (*Castellini et al., 2014a; Manca et al., 2013b; Castellini et al., 2010a; Castellini et al., 2009b; Castellini et al., 2008a*) and to simulate their dynamics in-silico with different conditions. Differential equations are a standard modeling framework in this context. I have collaborated to the development of a non-conventional and time-discrete approach based on difference equations, called Metabolic P systems.

Several techniques were used to infer metabolic flux regulation functions. Multiple linear regression was used together with stepwise procedures for selecting flux regulators (*Castellini et al., 2011a*). A machine learning approach based on artificial neural networks was developed for generating nonlinear models, wherein both backpropagation and bio-inspired optimization techniques (e.g., genetic algorithms, particle swarm optimization and some heuristics) were used for training the models (*Castellini et al., 2010b; Castellini et al., 2009c*). Finally, a methodology based on genetic algorithms and multiple linear regression was proposed which

deals with model generation as an evolutionary process (*Castellini et al., 2015b; Castellini et al., 2013a; Castellini et al., 2012b; Castellini et al., 2012c*). An open source Java software called MetaPlab (<http://mplab.sci.univr.it/>) (*Castellini et al., 2009a; Castellini et al., 2007a*) was released along with some plugins, such as, GeneticSynth (*Castellini et al., 2015a*), that provide support for specific tasks related to the model building process.

- *Project*: “Comparative genomics by k-mer based indices”. Center for BioMedical Computing (CBMC), University of Verona, Italy (01/2010 - 03/2014).

*Description*: This project concerns genomic sequence analysis and comparative genomics by means of dictionary-based indexes. The final goal is to discover statistical/informational properties and patterns in genomes (*Castellini et al., 2012a*). As sequences of symbols, genomes determine dictionaries constituted by k-mers (factors) occurring in them. The starting point of our analysis was the computation of dictionaries of k-mers, with  $k = 6, 12, 18$ , of given genomes. Some properties of such dictionaries, such as the distributions of k-mers, and their compared statistics guided the research along lines of development which emerged from the empirical evidence of computed data. A new (mathematical) perspective on genomes stemmed from this approach and a Java client-server application was developed to provide the high computational power required for this kind of analysis. Suitable data structures and ad-hoc parallel algorithms were implemented for systematically storing, querying, analyzing and comparing k-mer dictionaries and a web portal for accessing this data was designed. Analyses moved in various directions, among them, the investigation of k-repeat-sharing gene networks (networks of genes connected if sharing k-mers of a given length) using statistical and graph theoretical methods (*Castellini et al., 2015c*), and genome classification (*Castellini et al., 2011b*).

- Projects at Eli Lilly, Benetton, Miroglio, Italcementi, Cattolica assicurazioni. Sdg group, Verona, Italy (05/2012 - 03/2014).
- *Project*: “Analysis of Space Technologies”. Italian Space Agency, marketing unit, Rome, Italy (09/2006 - 03/2007).

## Participation in conferences and related talks (relatore a congressi e convegni)

Symbols \* and \*\* denote personal presentation of, respectively, papers or posters in the conference.

- MOD 2017\*: The Third International Conference on Machine Learning, Optimization and Big Data, 14-17 September 2017, Volterra, Italy. Presented paper (*Castellini et al., 2017d*).
- DataMod 2017\*: 6th International Symposium “From Data to Models and Back”, Satellite event of SEFM 2017, 4-5 September 2017, Trento, Italy. Presented paper (*Castellini et al., 2017c*).
- IPCAT 2015\*: Information Processing in Cells and Tissues, 14-16 September 2015, San Diego, USA. Presented paper (*Castellini et al., 2015d*).
- PLANT 2030 Status Seminar 2015, 04-06 March 2015, Potsdam, Germany.
- GECCO 2012\*\*: Genetic and Evolutionary Computation Conference, 7-11 July 2012, Philadelphia, USA. Presented paper (*Castellini et al., 2012c*).
- ICARIS 2012\*: The 11th International Conference on Artificial Immune Systems, 28-31 August, Taormina, Italy. Invited talk, paper (*Castellini et al., 2010b*).
- MIMOS 2012\*: Italian Movement for Modeling and Simulation, 9-11 October 2012, Rome, Italy. Invited talk, Ph.D. thesis (*Castellini, 2010c*).

- PRIB 2011\*\*: The 6th IAPR International Conference on Pattern Recognition in Bioinformatics, November 2-4 2011, Delft, The Netherlands. Presented poster (*Castellini et al., 2011b*).
- GECCO 2009\*: Eleventh annual conference on Genetic and evolutionary computation, 8-12 July 2009, Montréal, Québec, Canada. Presented paper (*Castellini et al., 2009c*).
- WMC 2009\*: Tenth workshop on Membrane Computing, 24-27 August 2009, Curtea de Argeş, Romania. Presented paper (*Castellini et al., 2010b*).
- QAPL 2009\*: Seventh Workshop on Quantitative Aspects of Programming Languages, 28-29 March 2009, York, UK. Presented abstract: *Artificial Neural Networks for synthesizing Metabolic P Systems*.
- WMC 2008\*: Ninth Workshop on Membrane Computing, 28-31 July 2008, Edinburgh, UK. Presented paper (*Castellini et al., 2009a*).
- Modern Software Solutions in Drug Discovery Design and Planning Methods, FQS Poland (Fujitsu), 7 March 2008 Basel, Switzerland.
- BWMC6\*: Sixth Brainstorming Week on Membrane Computing, 4-8 February 2008, Sevilla, Spain. Presented abstract (*Castellini et al., 2008b*).
- NICO 2007\*: International Workshop on Nature Inspired Cooperative Strategies for Optimization, 8-10 November 2007, Acireale, Italy. Presented paper (*Castellini et al., 2008a*).

#### Internal talks (seminari interni)

- Data integration and prediction. OPTIMAL consortium internal meeting (“Genetic and biomarker-based predictive breeding of corn cultivars”). Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, April 2015.
- Statistical learning approaches for integrative modeling of molecular phenotypes in the context of biomass heterosis and metabolic networks. Progress seminar. Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, March 2015.
- Towards genetic and biomarker-assisted predictive breeding of maize hybrids. Affinity seminar. Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany, December 2014.
- Business intelligence and data mining in biomedical research. Center for BioMedical Computing (CBMC), November 2011.
- Variable selection for linear regression models: stepwise regression. University of Verona, Department of Computer Science, November 2009.
- Metabolic P systems for systems biology. Graduate School of Information Science, Nagoya University, Japan, December 2008.
- Metabolic P systems for modelling biological processes. ASAP group, School of Computer Science, University of Nottingham, UK, October 2008.
- The MetaPlab plugin-based architecture. Infobiotics software meeting, ASAP group, School of Computer Science, University of Nottingham, UK, June 2008.
- Theories and applications of modelling by Metabolic P systems. Research Away Day (RAD’08), ASAP group, Crown Plaza hotel, Nottingham, UK, June 2008.

- Modelling biological systems by MP systems. University of Verona, Department of Computer Science, January 2008.
- Genetic Programming for the Automatic Synthesis of Metabolic Pathways. University of Verona, Department of Computer Science, January 2008.
- Comparing models in computational systems biology. University of Verona, Department of Computer Science, October 2007.

#### **Awards (premi e riconoscimenti per attività di ricerca)**

- PhD thesis “A. Castellini. Algorithms and software for biological MP modeling by statistical and optimization techniques”. Winner of *Premio MIMOS 2011* for the best research thesis about simulation - MIMOS: Italian Movement for Modeling and Simulation - [www.mimos.it](http://www.mimos.it) (€1500).
- Master thesis “A. Castellini. Business intelligence and data mining in biomedical research”. Winner of *premio Martello 2010/11* for the best project work, Master in Business Intelligence e Knowledge Management 2010/11, University of Verona (€2000).
- Paper “A. Castellini and V. Manca. Learning regulation functions of metabolic systems by artificial neural networks”, *GECCO 2009*. Nominated for the “best paper award” in the *Bioinformatics and Computational Biology track* (2nd position).

#### **Grants (borse di studio)**

- Travel and accommodation costs for the visit at University of California Riverside (UCR) and participation in the international conference IPCAT 2015, San Diego, California, US. Provided by Potsdam Graduate School, University of Potsdam (€3800, 2015).
- Travel costs for the participation in the international conference GECCO 2009, Montreal, Canada. Provided by GECCO committee and Graduate School of Sciences Engineering and Medicine, University of Verona (about €2000, 2009).
- Application costs for the participation in the “Stochastic modeling” course at the Business school of Lancaster university, UK. Provided by the Operational Research Society - Prof. Natalio Krasnogor (2009).
- Travel and accommodation costs for the visit at the Science of Complex Interaction Lab (SCI-LAB), department of Complex Systems Science, Nagoya University, Japan. Provided by CooperInt funds, University of Verona (€3000, 2008).
- State Scholarship, PhD program in Computer Science, University of Verona (3-years, 2007-2009).
- Grant for the internship at the Agenzia Spaziale Italiana (Roma). Provided by the University of Verona (€3000, 2006-2007).

#### **Research visits (visite di ricerca)**

- Visiting associate researcher: Algorithms and Computational Biology Lab, Dept. Computer Science & Engineering, University of California Riverside, (1.5 months, 08/2015 - 09/2015). Prof. S. Lonardi. Grant from University of Potsdam.

- Visiting Ph.D. student: Science of Complex Interaction Lab (SCI-LAB), Dept. Complex Systems Science, Nagoya University, Japan (1.5 months, 11/2008 - 12/2008). Dr. Y. Suzuki. Grant from University of Verona.
- Visiting Ph.D. student: Automated Scheduling Optimization and Planning (ASAP) group, School of Computer Science, Nottingham University, UK (5 months, 06/2008 - 11/2008). Prof. N. Krasnogor.
- Erasmus student: Computer Vision Group, Dept. Electrical Electronic and Computer Engineering, Heriot Watt University, Edinburgh, UK (3 months, 06/2005 - 09/2005). Prof. M. Trucco. Grant from University of Verona.

### **Student supervision (supervisione di studenti)**

- 02/2017 - 10/2017: Valeria Gottelli, master degree in computer science, master thesis (Supervisor: Prof. Alessandro Farinelli, Co-supervisor: Dr. Alberto Castellini). Title: Situation assessment for water drones based on statistical methods.
- 10/2016 - 07/2017: Giovanni Alberto Beltrame, bachelor degree in bioinformatics, bachelor thesis (Supervisor: Prof. Alessandro Farinelli, Co-supervisor: Dr. Alberto Castellini). Title: Situation assessment for autonomous water drones based on probabilistic graphical models.
- 10/2016 - 07/2017: Gianni Di Dio, master degree in computer science, master thesis (Supervisor: Prof. Alessandro Farinelli, Co-supervisor: Dr. Alberto Castellini). Title: Rule-based anomaly detection for precision agriculture.
- 06/2015 - 08/2016: Gianni Di Dio, master degree in computer science, course “Artificial intelligence” (Prof. Alessandro Farinelli). Project: regularized linear regression methods for predicting *Peronospora belbahrii* in sweet basil.
- 06/2016 - 08/2016: Nicolò Danzi, master degree in computer science, course “Artificial intelligence” (Prof. Alessandro Farinelli). Project: regularized linear regression methods for predicting *Peronospora belbahrii* in sweet basil.
- 03/2015 - 08/2015: Tom Janetzek, bachelor degree in molecular biology and physiology, bachelor thesis. Title: Clustering and feature selection analysis of time-course maize phenotypic data.
- 03/2015 - 07/2015: Alessio Milanese, Max Planck Institute of Molecular Plant Physiology (Germany), Erasmus+ program. Regularization techniques for reconstruction of metabolic networks.
- 01/2013 - 03/2014: Daniele Paltrinieri, master degree in bioinformatics and medical biotechnology, University of Verona, course “Languages for bioinformatics”. Project: MP-GeneticSynth plugin (<http://mplab.sci.univr.it/plugins/mpgs/index.html>) for MetaPlab. Extended and published in (*Castellini et al., 2015a*).
- 01/2012 - 12/2012: Luca Barbon, master degree in bioinformatics and medical biotechnology, University of Verona, course “Languages for bioinformatics”. Project: pipeline for genomic sequence analysis by dictionary-based indexes.
- 01/2013 - 12/2012: Emiliano Maresi, bachelor in bioinformatics, University of Verona, internship. Project: analysis of repeat-sharing gene networks.



- 01/2013 - 07/2013: Alessio Milanese, master degree in bioinformatics and medical biotechnology, University of Verona, course “Languages for bioinformatics”. Project: algorithms for symbolic regression of metabolic flux-regulation functions by genetic programming (Matlab).
- 07/2010 - 12/2012: Nicola Gaburro, bachelor degree in bioinformatics, University of Verona, internship. Project: tools for symbolic regression of metabolic flux-regulation functions by genetic programming (Matlab, GPLab). The project was extended to the bachelor degree thesis: “Programmazione genetica applicata alla regressione di flussi metabolici” (Supervisor Dr. Franco).
- 05/2012 - 12/2012: Alessio Milanese, bachelor degree in bioinformatics, University of Verona, internship. Project: algorithms for the analysis of repeat-sharing gene networks by means of statistical and graph-theory based techniques (Matlab). Extended and published in (*Castellini et al., 2015c*).
- 05/2011 - 12/2012: Mauro Zucchelli, master degree in bioinformatics and medical biotechnology, University of Verona, course “Algorithms and languages for bioinformatics”. Project: genetic algorithms and regression techniques for reverse-engineering of biological systems (Matlab). Extended and published in (*Castellini et al., 2015b; Castellini et al., 2013a; Castellini et al., 2012b*).
- 03/2012 - 04/2012: Marika Albrizio, bachelor degree in bioinformatics, University of Verona, internship. Project: algorithms for detection and analysis of palindromic strings in genomic sequences.
- 03/2012 - 04/2012: Cesare Centomo, bachelor degree in bioinformatics, University of Verona, internship. Project: analysis of repeat-sharing gene networks by gene ontology techniques.
- 05/2011 - 03/2012: Mirko Busato, master degree in bioinformatics and medical biotechnology, University of Verona, course “Algorithms and languages for bioinformatics”. Project: genetic algorithms and regression techniques for reverse-engineering of biological systems (Matlab). Extended and published in (*Castellini et al., 2013a; Castellini et al., 2012c*).
- 01/2012 - 07/2012: Matteo Denitto, master degree in bioinformatics and medical biotechnology, University of Verona, course “Algorithms and languages for bioinformatics”. Project: software tools for correlation analysis of immunological data (Matlab).
- 04/2011 - 09/2011: Sara Compri, Valerio Marino and Gabriele Tosadori, master degree in bioinformatics and medical biotechnology, University of Verona, course “Algorithms and languages for bioinformatics”. Project: genome classification by dictionary-based indexes (Matlab/Python). Poster published in (*Castellini et al., 2011b*).
- 05/2008 - 01/2009: Marika Rodegher, master degree in computer science, University of Verona, internship. Project: linear regression plugin for MetaPlab (Java, <http://mplab.scienze.univr.it/plugins>).
- 04/2008 - 07/2008: Federica Agosta, bachelor degree in computer science, University of Verona. Project: MetaPlab website (PHP/MySQL, <http://mplab.scienze.univr.it/>).

### **Program Committees of International Conferences (comitati di programma)**

- Genetic and Evolutionary Computation Conference (GECCO 2010, 2011, 2012, 2013, 2014): member of program committee.
- International Conference on Machine Learning, Optimization and Big Data (MOD 2017): session chair.

### **Activity as Referee (attività di referaggio)**

- Bioinformatics (Oxford Journals)
- BioSystems (Elsevier),
- IEEE Transactions on Neural Networks and Learning Systems,
- Genetic and Evolutionary Computation Conference (GECCO),
- Journal of Regenerative Medicine and Tissue Engineering,
- International Workshop on Natural Computing (IWNC),
- International Conference on Parallel Problem Solving From Nature (PPSN),
- European Conference on Artificial Life (ECAL),
- Letters in Drug Design and Discovery (LDDD).

### **Courses attended during PhD with exam (partecipazione a corsi durante dottorato)**

- Stochastic modeling, Business School of Lancaster University, UK. Supported by the Operational Research Society (30 hours, 23/03/09 - 27/03/09).
- Bioinformatics, master degree in biotechnology, University of Verona, Prof. Giorgetti (40 hours, 09/11/07 - 23/01/08).
- Summer school “Computazione evolutiva e vita artificiale (SeCeVita 07)”, University of Catania. Keynotes: Proff. Nicosia, Cagnoni, Maniezzo, Mirolli, Vanneschi, Serra, Della Cioppa, Tettamanzi, Arena, Nolfi, Maglino (35 hours, 31/08/07 - 04/09/07).
- Advanced algorithms for the analysis and visualization of DNA and protein sequences, Department of Computer Science, University of Verona, Prof. Zu-Guo Yu (20 hours, 16/07/07 - 26/07/07).
- Advanced algorithms, master degree in computer science, University of Verona, Prof. Posenato (40 hours, 2/04/07 - 08/06/07).
- Project Management, ISTUD Business School (32 hours, 09/09/06 - 17/09/06), Italian Space Agency.

### **Other courses attended (partecipazione ad altri corsi)**

- IBM Big Data Platform: Infosphere Data Explorer (5 days, 08/07/2013 - 12/07/2013). IBM, London. Supported by Sdg group.
- SAS: Information Retrieval Studio (crawling web, Twitter, Facebook), Text Miner/Enterprise Miner, Sentiment Analysis Studio, Enterprise Guide, Enterprise Business Intelligence (7 days, 2012-2013). SAS, Milan and Rome. Supported by Sdg group.

- Statistical software “Stata” (24 hours, 20/01/2010 - 24/02/2010). Club economia applicata, Faculty of Economics, University of Verona.
- Modern software solutions in drug discovery design and planning methods (8 hours, 07/03/2008), Fujitsu FQS, Basel, Switzerland.

### Technical skills (conoscenze software e di programmazione)

- Programming languages and tools for software development:
  - Good knowledge: R (3 years), Matlab (3 years), Java (7 years), SQL, RStudio, MySql, NetBeans.
  - Basic/Medium knowledge: Python/Biopython, C/C++, J2EE, SAS language, Twitter/Facebook API, XML/XSLT/SBML, PHP, JavaScript, Apache/Tomcat, GlassFish, IBM Netezza, Oracle database, PostgreSQL, MS Access, UML.
- Operating systems and software applications: MS Windows, Linux, Latex, MS Word, MS Excel, MS PowerPoint, Cytoscape, IBM Data Explorer (IBM Big Data Platform), SAS Information Retrieval Studio/Enterprise Guide/JMP/Sentiment Analysis Studio/Enterprise Miner, Stata, Weka, SAP Business objects, QlikView.

### Publications (pubblicazioni)

#### Papers in refereed international journals

1. *Castellini et al., 2015a*. A. Castellini, D. Paltrinieri, V. Manca. MP-GeneticSynth: Inferring biological network regulations from time series. *Bioinformatics*, Oxford University Press, 31(5):785-787, 2015.
2. *Castellini et al., 2015b*. A. Castellini, V. Manca, M. Zucchelli. An evolutionary procedure for inferring MP systems regulation functions of biological networks. *Natural Computing*, Springer, 14(3):375-391, 2015 (published online in 2014).
3. *Castellini et al., 2015c*. A. Castellini, G. Franco, A. Milanese. A genome analysis based on repeat sharing gene networks. *Natural Computing*, Springer, 14(3):403-420, 2015 (published online in 2014).
4. *Castellini et al., 2013a*. A. Castellini, M. Zucchelli, M. Busato, V. Manca. From time series to biological network regulations: an evolutionary approach. *Molecular BioSystems*, 9(2):225–233, The Royal Society of Chemistry, 2013.
5. *Manca et al., 2013b*. V. Manca, A. Castellini, G. Franco, L. Marchetti, R. Pagliarini. Metabolic P Systems: A discrete model for biological dynamics. *Chinese Journal of Electronics*, 22(4):717–723, 2013.
6. *Castellini et al., 2012a*. A. Castellini, G. Franco, V. Manca. A dictionary based informational genome analysis. *BMC Genomics*, 13(1):485+, 2012.
7. *Castellini et al., 2011a*. A. Castellini, G. Franco, R. Pagliarini. Data analysis pipeline from laboratory to MP models. *Natural Computing* 10(1):55–76, Springer, 2011.
8. *Castellini et al., 2010a*. A. Castellini, G. Franco, V. Manca. Hybrid functional Petri nets as MP systems. *Natural Computing* 9(1):61–81, Springer, 2010.

## Book chapters selected from refereed conference proceedings

9. *Castellini et al., 2017a.* A. Castellini, G. Franco and A. Vella. Age-related relationships among peripheral B lymphocyte subpopulations. In 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, 2017, pages 1864-1871, 2017.
10. *Castellini et al., 2015d.* A. Castellini\*, C. Edlich-Muth\*, M. Muraya, C. Klukas, T. Altmann, J. Selbig. Towards a graph-theoretic approach to hybrid performance prediction from large-scale phenotypic data. \*:These authors contributed equally to this work. In the 10th International Conference on Information Processing in Cells and Tissues, IPCAT 2015, San Diego, California, USA, September 14-16 2015, *Lecture Notes in Computer Science*, volume 9303, pages 271-285, Springer-Verlag, 2015.
11. *Castellini et al., 2014a.* A. Castellini, G. Franco, V. Manca, R. Ortolani, A. Vella. Towards an MP Model for B lymphocytes maturation. In 13th International Conference on Unconventional Computation and Natural Computation, UCNC 2014, University of Western Ontario, London, Ontario, Canada, July 14-18, 2014, *Lecture Notes in Computer Science*, volume 8553, pages 80-92, Springer-Verlag, 2014.
12. *Castellini et al., 2012b.* A. Castellini, V. Manca, M. Zucchelli. Towards an evolutionary procedure for reverse-engineering biological networks. In the 11th International Conference on Artificial Immune Systems, ICARIS 2012, Taormina, Italy, August 28-31, 2012, *Lecture Notes in Computer Science*, volume 7597, pages 271-285, Springer-Verlag, 2012.
13. *Castellini et al., 2010b.* A. Castellini, V. Manca, Y. Suzuki. Metabolic P system flux regulation by artificial neural networks. In 10th International Workshop on Membrane Computing, WMC 2009, Curtea de Arges, Romania, August 24-27, 2009, Revised, Selected and Invited Papers, G. Păun and M. J. Pérez-Jiménez and A. Riscos-Núñez and G. Rozenberg and A. Salomaa, editors, *Lecture Notes in Computer Science*, volume 5957, pages 196-209, Springer-Verlag, 2010.
14. *Castellini et al., 2009a.* A. Castellini, V. Manca. MetaPlab: A computational framework for Metabolic P systems. In 9th International Workshop on Membrane Computing, WMC 2008, Edinburgh, UK, July 28-31, 2008, Revised, Selected and Invited Papers, D. W. Corne, P. Frisco and G. Păun and G. Rozenberg and A. Salomaa, editors, *Lecture Notes in Computer Science*, volume 5391, pages 157-168. Springer-Verlag, 2009.
15. *Castellini et al., 2009b.* A. Castellini, G. Franco, V. Manca. Toward a representation of hybrid functional Petri nets by MP systems. In 2nd International Workshop on Natural Computing, Nagoya, Japan, December 10-13, 2007, Y. Suzuki, M. Hagiya, H. Umeo, and A. Adamatzky, editors, *Natural computing*, volume 1 of *Proceedings in Information and Communications Technology*, pages 28-37. Springer Japan, 2009.
16. *Castellini et al., 2008a.* A. Castellini, V. Manca, L. Marchetti. MP systems and hybrid Petri nets. In International Workshop on Nature Inspired Cooperative Strategies for Optimization, NICO 2007, Acireale, Italy, November, 8-10, 2007, N. Krasnogor, G. Nicosia, M. Pavone, D. Pelta, editors, *Studies in Computational Intelligence 129*, pages 53-62, Springer, 2008.
17. *Bianco et al., 2007a.* L. Bianco, A. Castellini. Psim: A computational platform for Metabolic P systems. In 8th International Workshop on Membrane Computing, WMC 2007, Thessaloniki, Greece, June 25-28, 2007, Revised, Selected and Invited Papers, G.

Eleftherakis and P. Kefalas and G. Păun and G. Rozenberg and A. Salomaa, editors, *Lecture Notes in Computer Science*, volume 4860, pages 1-20, Springer-Verlag, 2007.

### Papers at conferences and workshops, with refereed published proceedings

18. *Castellini et al., 2017b.* A. Castellini, A. Farinelli, G. Minuto, D. Quaglia, I. Secco, F. Tinivella. EXPO-AGRI: Smart Automatic Greenhouse Control. In *FoodCAS 2017 Electronics for Better Quality of Food, Torino, October 20, 2017*. To appear.
19. *Castellini et al., 2017d.* A. Castellini, G. Franco. Detection of age-related changes in networks of B cells by multivariate time-series analysis. In *MOD 2017 - The third International Conference on Machine Learning, Optimization and Big Data*, Volterra, Italy, September 14-17, 2017.
20. *Castellini et al., 2017e.* A. Castellini, G. A. Beltrame, M. Bicego, D. Bloisi, J. J. Blum, M. Denitto, A. Farinelli. Activity recognition for autonomous water drones based on unsupervised learning methods. In *4th Italian Workshop on Artificial Intelligence and Robotics (AIRO 2017), satellite event of the 16th International Conference of the Italian Association for Artificial Intelligence (AI\*IA 2017)*, Bari, Italy, November 14-17, 2017.
21. *Castellini et al., 2012c.* A. Castellini, V. Manca, M. Zucchelli, M. Busato. A genetic approach for synthesizing metabolic models from time series. In T. Soule and J. H. Moore, editors, *GECCO '12: Genetic and Evolutionary Computation Conference, GECCO '12, Philadelphia, PA, USA, July 7-11, 2012, Companion Material Proceedings*, pages 1415-1416, ACM, 2012.
22. *Castellini et al., 2009c.* A. Castellini, V. Manca. Learning regulation functions of metabolic systems by artificial neural networks. In F. Rothlauf, editor, *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation, July 8-12, 2009, Montreal Canada*, pages 193-200, ACM, New York, NY, USA, 2009 (nominated for the best paper award).
23. *Castellini et al., 2008b.* A. Castellini, G. Franco. On modelling signal transduction networks. In D. Diaz-Pernil, C. Graciani, M. Gutierrez-Naranjo, G. Păun, I. Perez-Hurtado, A. Riscos-Nunez, *Proceedings of Sixth Brainstorming Week on Membrane Computing, BWMC6, February 4-8, 2008, Sevilla, Spain*, pages 67-77, Fenix Editora, 2008.

### Posters and extended abstracts

24. *Castellini et al., 2017c.* A. Castellini, G. Franco. Multivariate time-series segmentation for immunological data analysis. In *6th International Symposium - From Data to Models and Back (DataMod 2017), satellite event of the 15th International Conference on Software Engineering and Formal Methods (SEFM 2017)*, Trento, Italy, September 4-5, 2017.
25. *Edlich-Muth et al., 2015e.* C. Edlich-Muth\*, A. Castellini\*, J. Selbig, Z. Nikoloski. Towards genetic and biomarker-assisted predictive breeding of maize hybrids. \*:These authors contributed equally to this work. In *PLANT 2030 Status Seminar 2015*, Potsdam, Germany, 2015.

26. *Castellini et al., 2011b*. A. Castellini, V. Manca, M. Bicego, S. Compri, V. Marino, G. Tosadori. Genome classification by dictionary-based indexes. *6th International Conference on Pattern Recognition in Bioinformatics, Delft, The Netherlands*, November 2-4, 2011.

### **Theses**

27. *Castellini, 2011c*. A. Castellini. Business intelligence and data mining in biomedical research. Master thesis, Master in Business Intelligence e Knowledge Management, Faculty of Economics, University of Verona, 2011 (winner of the best master thesis award).
28. *Castellini, 2010c*. A. Castellini. Algorithms and software for biological MP modeling by statistical and optimization techniques. PhD thesis, Department of Computer Science, University of Verona, 2010 (winner of MIMOS award).
29. *Castellini, 2006a*. A. Castellini. Simulazioni computazionali di operazioni del “DNA Computing”. Master thesis, Department of Computer Science, University of Verona, 2006.

### **Technical reports**

30. *Manca et al., 2009c*. V. Manca, A. Castellini, G. Franco, L. Marchetti, R. Pagliarini. Metaplab 1.1 user guide. Department of Computer Science, University of Verona (<http://mplab.scienze.univr.it>), 2009.
31. *Castellini, 2007b*. A. Castellini. Albero delle tecnologie. Italian Space Agency, Rome, 2007.
32. *Castellini, 2005a*. A. Castellini. Human body tracking using particle swarm optimization, Group of Computer Vision, Department of Electrical Electronic and Computer Engineering, Heriot Watt University, Edinburgh, UK. 2005.