



## French Digital Agriculture Convergence Lab

PhD positions advertised,

Autumn 2018

Agriculture has to cope with the triple issue, ie being competitive, preserving the environment and offering correct living conditions to farmers. Digital technologies, eg satellite images, smartphones, connected objects, Internet of Things, big data, modelling, simulation, High-Performance Computing may help to face the challenge... However, innovation in digital agriculture is not straightforward: technology adoption is an issue and the way they can revolutionize the global value chain is not known.

#DigitAg, the “Digital Agriculture Convergence Laboratory”, has been created in 2017 in Montpellier (with some labs in Toulouse and Rennes) with the ambition to address these questions, by putting together a substantial interdisciplinary taskforce of about 300 researchers from 30 laboratories in Montpellier, Toulouse and Rennes (France). It carries out research, higher education and innovation to support the development of digital agriculture and related digital economy in France and in the Southern countries. Interdisciplinary research deals with disciplines crossing biological, agricultural, digital, economic and social sciences. #DigitAg gathers 17 partners, both public (INRA, INRIA, IRSTEA, CIRAD, Montpellier University, Montpellier Supagro, AgroParisTech Montpellier, ACTA, SATT AxLR) and private (IDATE, Smag, Vivelys, Pera-Pellenc, Agriscope, Fruition Science, ITK, Terranis). It is financially supported during 7 years by the National Research Agency, with a prisionnal funding of 9,9 M€.

#DigitAg supports a dozen of PhD grants every year, from blue sky research to applied research, as well as around 25 grants for master internships per year. #DigitAg has the ambition to host foreign students and researchers, with dedicated calls for mobility.

Véronique Bellon-Maurel  
Director of #DigitAg  
contact@hdigitag.fr

More information on: [www.hdigitag.fr/en/](http://www.hdigitag.fr/en/)

## Doctoral thesis subjects:

- 1) Integrating metabolomic data by quantification of high throughput data. Application to perinatal mortality in pigs 3
- 2) Distribute durably with digital? The case of fresh agricultural and food products ..... 4
- 3) Leverage Multi-Source Remote Sensing data via machine learning to improve Crop Monitoring Systems: a Cross-comparison between France and Senegal ..... 5
- 6) Selecting pesticidal plants for animal and plant health in Africa using exploratory conceptual navigation ..... 6
- 7) Applying big data methodologies to improve the chemometrics local-PLS algorithms ..... 7
- 8) Combining Semantic Web and modelling approaches for organizing phenotyping data collected at different scales of plant organization ..... 8
- 9) Interactive exact optimisation for numerical services to agriculture ..... 9
- 10) Data analysis from the connected beehives of beekeepers ..... 10
- 11) Improving innovation processes in digital agriculture ..... 11
- 12) The Innovation System of Digital Agriculture facing the ecological transition ..... 12
- 13) Improving food security systems by linking heterogeneous data - The case of agricultural production in West Africa ..... 13
- 14) Quantification of pest regulation by generalist predators by analysis of image sequence to determine and quantify network of interactions, case of the banana weevil ..... 14



## 1) Integrating metabolomic data by quantification of high throughput data. Application to perinatal mortality in pigs

Contact: Nathalie Villa-Vialaneix

Email: [nathalie.villa-vialaneix@inra.fr](mailto:nathalie.villa-vialaneix@inra.fr)

Institute: INRA

Summary:

Metabolomics data are an affordable way to access fine phenotyping information. They can be acquired from simple blood (plasma), amniotic fluid or urine samples using NMR (Nuclear Magnetic Resonance) high throughput technic. However, due to a lack of method to properly link acquired spectra to metabolite quantification, they are largely underexploited. This PhD proposes to develop a new fast and efficient approach to use metabolomics NMR spectra for metabolite quantification. The results will be validated with direct quantification of metabolites from the ANR project PORCINET ANR-09-GENM-005. This project addresses the issue of perinatal mortality in pigs and the numerous datasets acquired during the project provide a rich framework to prove the efficiency of our approach for precision farming. In particular, metabolite quantification will be integrated with various phenotypes (morphology, allometry, and physiology), proteomic and transcriptomic data obtained in a complex experimental design (4 genotypes x 2 late gestational stages) in order to better understand the biological processes involved in perinatal survival of pigs.

The output of this PhD will be twofold: first, we intent to develop state-of-the-art method for metabolite quantification from NMR spectra. The method will be released for public use in the form of an R package and a Galaxy module. Second, on a biological point of view, we will provide biomarkers explaining biological processes involved in piglet survival. These biomarkers will permit the selection of pig phenotypes to decrease the early death of piglets that will improve pig production competitiveness and sustainability. Because of the easy and cheap acquisition of metabolic data, our approach has great potentials for selection and will be generic enough to adapt to other types of problems.

## 2) Distribute durably with digital? The case of fresh agricultural and food products

Contact: Sophie Mignon

Email: [sophie.mignon.mtp@gmail.com](mailto:sophie.mignon.mtp@gmail.com)

Institute: University of Montpellier

Summary:

Lifestyles are changing rapidly, both in the West and among the middle classes in the South. Many modes of distribution have developed. An extreme example is that of an American company, which offers fresh home-made food products (including the Rosette de Lyon!) in the heart of New York. This latter model is based on an intensive mobilization of digital resources, with advantages (it brings together producers and consumers, it facilitates access to products of the most isolated, it reduces waste, limits car journeys) but also obvious disadvantages. On the other hand, this new form of distribution creates new balances of power between the actors within the sector and new legal arrangements (contracts etc.). The rebound effects induced by the use of ICT in this context are unknown, including the sustainability of the production companies engaged in these processes. In addition to the traditional forms of mass retailing in most developing countries, it is likely that new forms based on the use of ICT will emerge rapidly. Indeed, these countries are subject to technological leaps, as we have seen for mobile telephony. New distribution systems could therefore be developed through ICTs, both in the North and the South, and in the future will represent an important part of the market for the distribution of fresh agricultural and food products. In the hope that this form of distribution will increase, this project aims to improve our knowledge about its links with digital resources and its consequences of all kinds.

### **3) Leverage Multi-Source Remote Sensing data via machine learning to improve Crop Monitoring Systems: a Cross-comparison between France and Senegal**

Contact: Dino Ienco

Email: [dino.ienco@irstea.fr](mailto:dino.ienco@irstea.fr)

Institute: Irstea

Summary:

With the perspective of a population of 9 billion by 2050 and the expected impacts of climate change on natural and cultivated ecosystems, ensuring global food security while promoting sustainable agriculture is among the major challenges for the future development of our society. Monitoring the cultivated areas over large areas is vital to maintain sufficient levels of agricultural production.

Many studies have highlighted the potential of remote sensing to monitor agricultural production over large areas. In this context, modern Earth Observation (EO) systems, such as the Sentinel mission (freely available high-resolution optical and radar images with acquisitions every 5 days) or new constellation of NanoSatellites (eg. Planet), open up new opportunities to monitor and analyse cultivated ecosystems. Moreover, the currently employed methods (eg. to estimate cultivated areas and/or crop yields, eg] are no longer adapted to deal with such new and increasing sources of information. For example, very few studies have proposed tools to i) explicitly model the temporal correlation present in time series of remote sensing data and ii) efficiently combine optical and radar time series information to ameliorate the agricultural monitoring.

The present thesis aims to propose new techniques for monitoring cultivated areas by combining different sources of satellite data produced by modern EO systems (optical and radar time series, Very High Spatial Resolution imagery) using machine learning techniques. The approaches proposed will be calibrated and tested using data collected on the field. The thesis will be developed approaching the questions (cultivated area and crop yields estimation) through two different study sites, one located in France and characterized by conventional agriculture, and one in Senegal and characterized by a smallholder farming system.

## 6) Selecting pesticidal plants for animal and plant health in Africa using exploratory conceptual navigation

Contact: Marianne Huchard

Email: [marianne.huchard@lirmm.fr](mailto:marianne.huchard@lirmm.fr)

Institute: Cirad

Summary:

Social demand is incrementally forcing farmers, livestock producers and fish farmers to adopt alternative solutions to the use of synthetic chemical pesticides and antibiotics for agricultural crop protection and animal health, with the aim of improving food safety and environmental health. Among others, the use of pesticidal plants, applied in the form of essential oil, aqueous extract or directly using plant organs, is a solution that can be mobilised by low-income farmers in Africa.

The PPAf knowledge base gathers knowledge on the use of plants in Africa for plant, animal and human health. In PPAf, each use is described using 35 types of information, i.e. target organism, protected organism, method of preparation and application, unintended effects on non-target organisms, etc. The purpose of PPAf is to support farmers, their advisors, local entrepreneurs or researchers in selecting plants of immediate interest for food security in agricultural production systems. The problematic is to have the tools that enable field actors to realize this selection in the knowledge base.

Exploratory conceptual navigation is the adopted solution, as it is adapted to the context of large masses of data in which users formulate general, potentially imprecise, and potentially inaccurate queries without prior knowledge of the data. The work planned in the thesis will be oriented towards the development and dissemination of knowledge extraction tools for end users. It will consist of (i) establishing the theoretical foundations of conceptual classification and navigation by Pattern Structures and Relational Concept Analysis, (ii) elaborating algorithms to build on-demand responses around points of interest in knowledge, and (iii) developing interaction methods that assist users to express and transform on-the-fly queries according to the browsing context.

## 7) Applying big data methodologies to improve the chemometrics local-PLS algorithms

Contact: Jean-Michel Roger

Email: [jean-michel.roger@irstea.fr](mailto:jean-michel.roger@irstea.fr)

Institute: Irstea

### Summary:

Near infrared spectrometry can provide huge amounts of data to digital agriculture. The main tool of chemometrics, used to analyze NIR spectra, is Partial Least Squares (PLS) regression. PLS allows building efficient predictive models from a large number of variables even if these variables are highly correlated. The method has proved its relevance for small homogeneous databases. Its extension to medium-sized bases (<10,000 individuals) is the "local-PLS": it determines a neighborhood of the individual to be predicted, and then realizes a usual PLS on this neighborhood. This method combines the power of the  $k$  nearest neighbors' method ( $k$ -NN) and the PLS. However, it is not able to process large databases (e.g. >50,000 individuals) or even >1 million of individuals that will appear in the near future to digital agriculture. The current local-PLS algorithms all use sequential  $k$ -NN algorithms for which calculation times become unrealistic; other algorithms must be considered. Paradoxically, very little research has been done on this challenge in chemometrics. Our idea is that algorithms of indexation used in big data, integrated in the local-PLS method, could lift this methodological lock. We propose to consider two algorithms of dimension reduction and fast neighborhood searches used by the Zenith Team of Lirmm-Montpellier for processing large data sets of time series (that have a similar data structure as the NIR spectra): the hashing (calculation of sketches) and the iSax (Symbolic Aggregate approxImation). The work will consist in two steps: (1) a "business as usual" integration of the two algorithms in the local-PLS algorithm, (2) an optimisation of the algorithms taking into account the chemometric specificity of the NIR spectra. The new algorithms developed in this thesis will improve the ability to predict physico-chemical variables from large heterogeneous NIRS data bases, and will find direct applications in many domains (plants, feed, soils, etc.).

## 8) Combining Semantic Web and modelling approaches for organizing phenotyping data collected at different scales of plant organization

Contact: Llorenç Cabrera-Bosquet

Email: [llorenc.cabrera-bosquet@inra.fr](mailto:llorenc.cabrera-bosquet@inra.fr)

Institute: INRA

Summary:

In recent years, plant phenomics has generated massive datasets obtained in the field or in phenotyping platforms, at organ, plant or canopy levels. Such multi-origin, multi-scale datasets generate a massive problem for designing ontology-based information systems. The thesis consists in combining Semantic Web and modelling approaches to facilitate the mapping of concepts and ontologies at different levels of organization, thereby developing a framework for organising phenomic datasets. The PhD student will work at the interface between a biology group that has pioneered phenotyping approaches and a Mathematics-computer group specialist of data management. (i) He/she will adapt Semantic Web methods for expressing relationships between traits in field and greenhouse in such a way that software agents can query and integrate data originating from different contexts (e.g. greenhouse and field) and sources (e.g. databases and Web APIs). This will involve exploring existing ontologies, using existing tools for ontology alignment and adapting querying mechanisms to the sources mentioned above. (ii) He/she will then develop ontology extensions from sets of equations that link variables at plant by expressing the operations as semantic properties able to link concepts at plant and canopy levels. In this way the “field” and “greenhouse” concepts will be automatically translated one to the other via the existing knowledge expressed as equations. Proof of concepts will be carried out at each step by using existing datasets, for testing to what extent the conceptual framework developed in the thesis allows assembling datasets from different phenomic platforms in greenhouse and fields. This will allow mapping concepts able to cross scales of organisation and to create dynamic links between traits, with the final aim to re-use knowledge acquired in different contexts. The thesis will focus on traits associated with canopy development.



## 9) Interactive exact optimisation for numerical services to agriculture

Contact: Rodolphe Giroudeau

Email: [rodolphe.giroudeau@lirmm.fr](mailto:rodolphe.giroudeau@lirmm.fr)

Institute: University of Montpellier

Summary:

The trend towards a precise, numerical, and data intensive agriculture brings forward the need to integrate in a unique decision support methodology optimization techniques that are efficient, interactive, robust and adaptable. We propose to develop a decision calculus strategy for the management of agricultural activity that combines the efficiency and precision of optimisation methods based on linear integer programming and heuristics, and the flexibility and modularity of constraint programming methods. With the perspective that custom decision support services should be offered to farmers, we make the hypothesis that historic data, as well as daily updates of informations such as meteorology, crop evolutions and traceability informations should be available. This hypothesis allows for studying strategies that combine off-line (back office) approaches for searching best production processes that meet farmer criteria, and on-line (front office) approaches to contextualize and adapt solutions. We also make the hypothesis that distributed computing platforms could collaborate on calculus and numerical studies will be conducted on this topic, using available data and web services.

In fine, the methodology should integrate stochastic features because of the uncertainties of agricultural production. A first characterization and evaluation of solutions for the current decision period can be based on models built from data historic. In a second step, it is planned to make use of probability estimators linked to meteorological forecast to offer robustness oriented interactivity to the user of decision support tool and decision maker.

## 10) Data analysis from the connected beehives of beekeepers

Contact: Fabrice Allier

Email : [fabrice.allier@itsap.asso.fr](mailto:fabrice.allier@itsap.asso.fr)

Institute: ACTA

Summary :

The proposed study focuses on the analysis of data from electronic and connected scales measuring real-time hive weight. The student of this multidisciplinary thesis will be led by the ZENITH unit of INRIA and co-supervised by technical (ACTA / ITSAP-Institute of honeybee) and scientific structures (INRA, Bees and Environment unit). Together, we will attempt to develop and test the temporal dynamics of a hive's weight to predict the colony's ability, influenced by biotic and abiotic factors, to exploit available nectar resources within a given foraging radius. This will include taking into account conventional meteorological data. We will use robust methods to analyze large time series of data to create a predictive indicator to objectively support beekeepers for the installation and monitoring of honey production apiaries. More specifically, we will develop a matrix profile index on multidimensional and heterogeneous time series, which will be adapted to the data processed in this project. This study will focus on data mining (axis 5) and modeling (axis 6). Finally, we aim to provide beekeepers with decision support tools based on predictive models that value the real-time data of the network of equipped hives in order to facilitate the management of their production. The results will be evaluated individually by the beekeeper, but a collective valorization is also envisaged by constituting regional indicators.

## 11) Improving innovation processes in digital agriculture

Contact: Leïla Temri

Email: [temri@supagro.fr](mailto:temri@supagro.fr)

Institute: Montpellier SupAgro

Summary:

Smart agriculture is likely to contribute to the improvement of sustainability. However its environmental and social impacts are sometimes denounced. In fact, with the appearance of controversial technologies, such as biotechnologies, nanotechnologies, but also ICT in general, the concept of responsible innovation has emerged at the beginning of the 2000's, and since gained attention in the world, especially in developed countries, such as those situated in Europe. Responsible innovation refers to the inclusion of the stakeholders into the innovation process, and to the anticipation of social and environmental effects an innovation may induce.

Given this, we may ask our self the following question: Does responsible innovation lead to a better adoption of digital agriculture innovations? To answer to this question, it will be necessary: 1) To clarify concepts and definitions (responsible innovation, digital agriculture). 2) To draw up an inventory of the advantages and negative impacts digital technologies may have, especially in agriculture. 3) To identify relevant case studies, in the field of digital agriculture innovations, to describe innovation processes, and characterize them in terms of responsibility, using qualitative methods. 4) To assess the adequacy between the innovation process and users adoption. 5) Finally, recommendations could be made for improvement of processes, and a method could be proposed. The case studies will be chosen in the field of smartphone apps for agriculture. Some of them will come from southern countries to enrich the analysis.

## 12) The Innovation System of Digital Agriculture facing the ecological transition

Contact : Jean-Marc Touzard

Email: [jean-marc.touzard@inra.fr](mailto:jean-marc.touzard@inra.fr)

Institute : INRA

### Summary :

The thesis, which is positioned in the multidisciplinary field of Innovation Studies (Economics, Management and Sociology of innovation), mobilizes the approach of innovation systems, in order to analyze the development of digital agriculture and its possible effects on the ecological transition in the French agriculture. The thesis will thus study the modalities of adoption of digital innovations by two forms of agriculture that seek to contribute to this transition, either by "optimization of conventional practices" or by "breaking through organic farming". The research combines i) a national mapping of the innovation system of digital agriculture (census of institutions, firms, networks contributing to these innovations) with ii) the analysis of local dynamics of innovation in 3 sectors marked by both the development of digital agriculture and the coexistence of the two models which aim at contributing to the greening of agriculture: milk, cereal crops and wine (survey on a sample of farms). Digital agriculture has never been analyzed by Innovation Studies as a global movement, transforming agricultural sectors, its institutions, communities of knowledge and practices: This is a major scientific issue in research on innovation in agriculture. The thesis will contribute directly to the DigitAg convergence institute (axis 1), to the priorities of INRA (# 3Perf) and to the research on innovation network on 'Innovation (RNI).

## 13) Improving food security systems by linking heterogeneous data - The case of agricultural production in West Africa

Contact: Agnès Bégué

Email: [begue@cirad.fr](mailto:begue@cirad.fr)

Institute: Cirad

### Summary :

This thesis aims at the improvement of Food Security Monitoring systems through the use of heterogeneous data, focusing on the management of agricultural production risks. While agroclimatic data (e.g., satellite imagery, climate information, etc.) has been widely used for this task, the use of data coming from different domains (i.e., household surveys, social media, press, business analyses) has often been neglected. Remote sensing data is widely used for real time monitoring of vegetative growth, but is not sufficient to explain complex food safety-risk phenomena. The aim of this thesis is twofold: (i) to define innovative data mining techniques that will be able to exploit this heterogeneous data context. To reach this goal, three phases have been identified: (a) automatic discovery of spatial features from heterogeneous data, (b) features linking (i.e., through the definition of new similarity measures between features), and (c) data mining (i.e., through the definition of new network analysis, clustering and deep learning techniques) ; (ii) to show how remote sensing data can be enriched by linking it to data from different domains in order to make it more suitable for food safety-risk analysis tasks. During this #digitAg thesis, we will focus on studies carried out in Burkina Faso by the MOISA unit (Elodie Maître d' Hôtel), by exploiting satellite (with vegetation and climate features), economic, and textual data. The analytical framework will be based on retrospective analysis, focusing on the crop failures of 2007 and 2011 in Burkina Faso as major cases of studies. We will possibly extend our study to other areas, using data collected in Senegal. This study will be carried out in collaboration with Louise Leroux (UPR AIDA), member of the steering committee of the thesis #digitAg. Given the interdisciplinary path at the basis of this work, the results of the analysis and the defined techniques are expected to generate significant interest in socio-economic, remote sensing, and computer science (data mining) fields.

## 14) Quantification of pest regulation by generalist predators by analysis of image sequence to determine and quantify network of interactions, case of the banana weevil

Contact: Philippe Tixier

Email: [philippe.tixier@cirad.fr](mailto:philippe.tixier@cirad.fr)

Institute: Cirad

### Summary:

In the context of sustainable development in agriculture, it is crucial to define integrated pest management methods. Among these, pest regulation by natural enemies represents a promising way. However, to date, there is no method (i) to identify with certitude the generalist predators that are involved in the biocontrol of pests and (ii) to quantify the effect of this biocontrol on the population dynamics of the pests (i.e. the effectiveness of biocontrol in the field). These knowledges are crucial for the development and the transfer of effective agroecological management practices.

We propose the use of in natura mesocosms coupled with a non-perturbing video measurement method of interaction networks to quantify regulation and predation. These mesocosms will correspond to existing banana production situations in Martinique, the study area of the GECO unit, and will cover a large range of plant biodiversity (planned or not). In each of these mesocosms, the population dynamics of the banana weevil *Cosmopolites sordidus* will be followed by capture-mark-recapture method in order to determine the magnitude of the regulation by the natural enemies that occurs. In each selected situation, we will establish a control corresponding to a plot where natural enemies are excluded. The video method will allow to know the identity of the predators involved in the regulations and to quantify the links of trophic and non-trophic interactions existing in the animal community (frequency, duration and type of interactions). To this end, methods of automated digital image analysis will be developed (e.g. automatic recognition of a pest, count of individuals), including machine learning and in particular convolutional neural network methods that are developed by the ICAR team (LIRMM unit). These methods should revolutionize our understanding of the functioning of agrosystems and, because of their genericity, can be applied to the study of most cultivated systems.