

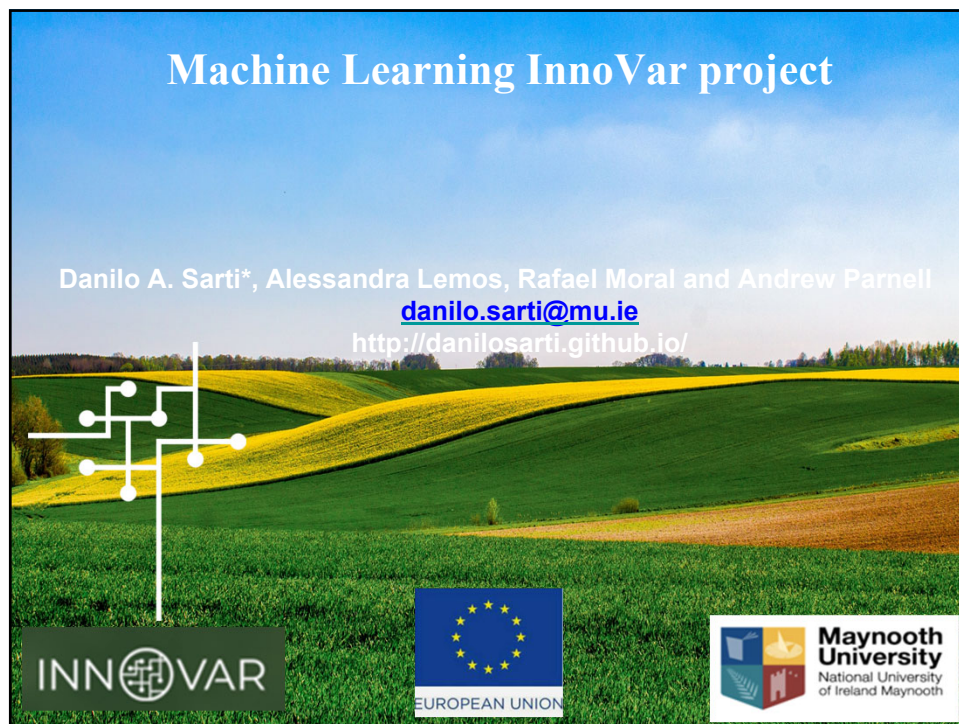
**Technical Working Party on Testing Methods and Techniques****TWM/1/25****First Session****Virtual meeting, September 19 to 23, 2022****Original:** English**Date:** September 19, 2022

---

**MACHINE LEARNING INNOVAR PROJECT***Document prepared by experts from Maynooth University**Disclaimer: this document does not represent UPOV policies or guidance*

The Annex to this document contains a copy of a presentation on “Machine Learning InnoVar project”, prepared by experts from Maynooth University, to be made at the first session of the TWM.

[Annex follows]



The slide features a background image of rolling green hills with a yellow field in the distance under a blue sky. A white circuit-like graphic is overlaid on the left side. At the bottom, there are three logos: INNOVAR, the European Union flag, and Maynooth University.

# Machine Learning InnoVar project

Danilo A. Sarti\*, Alessandra Lemos, Rafael Moral and Andrew Parnell  
[danilo.sarti@mu.ie](mailto:danilo.sarti@mu.ie)  
<http://danilosarti.github.io/>

**INNOVAR**  
**EUROPEAN UNION**  
**Maynooth University**  
National University of Ireland Maynooth

## AGENDA

- InnoVar project
- Machine Learning
- Machine Learning in InnoVar
- Achievements
- Next steps

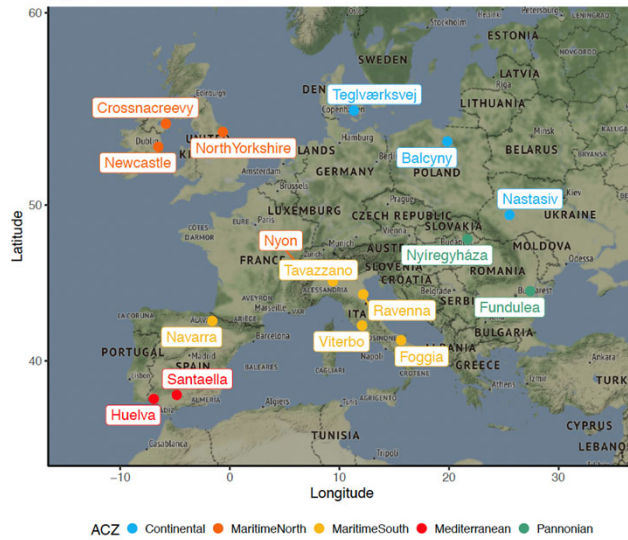
# InnoVar



- \* 21 Partners
- \* 15 Locations - 5 Agroclimatic zones
- \* Integration of new science into DUS and VCU testing processes
- \* Combine VCU and DUS characters
- \* Incorporate variety information into decision making on farm.
- \* *Triticum durum* and *Triticum aestivum*



Agro-climate Zones – ACZ



- \* Multi Environmental Trials
- \* Phenomics, DUS and VCU
- \* Genomics
- \* Weather
- \* Soil
- \* Drone images

Track ID	Track ID	Sample	SSTW ID
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12

SSTW ID	Sample	Drainage	Altitude
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12

Quantity	Type	SSTW ID
10	10	1
20	20	2
30	30	3
40	40	4
50	50	5
60	60	6
70	70	7
80	80	8
90	90	9
100	100	10
110	110	11
120	120	12

SSTW ID	Sample	Height	Yield
1	1	100	4
2	2	100	5
3	3	100	2.5
4	4	100	4
5	5	100	7
6	6	100	...
7	7	100	...
8	8	100	...
9	9	100	...
10	10	100	...
11	11	100	...
12	12	100	...

- \* Data collected will be stored in a database.
- \* Data will be used to create models and systems.
- \* **COVID** delays

## Machine Learning

Track ID	Track ID	Sample	SSTW ID
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12

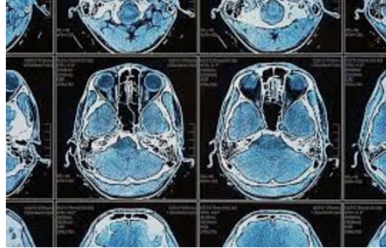
SSTW ID	Sample	Drainage	Altitude
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12

Quantity	Type	SSTW ID
10	10	1
20	20	2
30	30	3
40	40	4
50	50	5
60	60	6
70	70	7
80	80	8
90	90	9
100	100	10
110	110	11
120	120	12

SSTW ID	Sample	Height	Yield
1	1	100	4
2	2	100	5
3	3	100	2.5
4	4	100	4
5	5	100	7
6	6	100	...
7	7	100	...
8	8	100	...
9	9	100	...
10	10	100	...
11	11	100	...
12	12	100	...

Study of computer algorithms that can improve automatically through experience and use of data.

## Machine Learning Applications



Social Media  
Streaming platforms  
Precision medicine  
Precision agriculture  
Now: MET, DUS and VCU

## Machine Learning in InnoVar

Weather

Genomics

Phenomics

SOIL

```

# required by the copula package with d elements
z<-vector("list", length = n_soil)
for(i in 1:length(z)){
  z[[i]]$mean<-0
  z[[i]]$sd<-1
}
# now we update N, P, K and Clay
z[[6]][1]$mean=40## nitrogen 40ppm in average
z[[6]]$sd<-5
z[[7]][1]$mean=30## ## P
z[[7]]$sd<-3
z[[8]][1]$mean=40## K
z[[8]]$sd<-10
z[[9]][1]$mean=30## Clay
    
```

Data Collection

Computational Solutions

Decision Support Systems

Genomics

Weather

Phenomics

SOIL

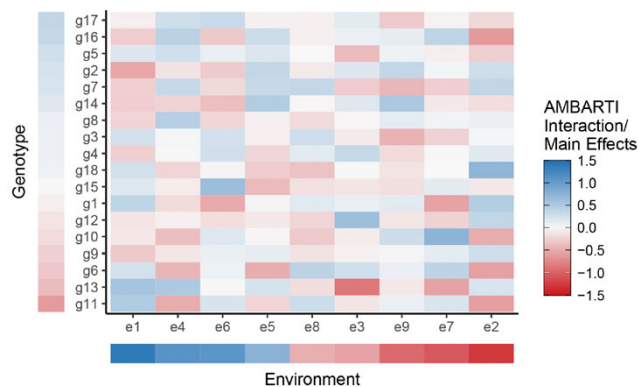
$$y_{ij}|x_{ij}, \theta \sim N\left(\mu + g_i + e_j + \sum_{t=1}^T h(x_{ij}, M_t, T_t), \sigma^2\right)$$

**Comprehend and predict the behaviour of DUS and VCU attributes considering interactions (G by E and G by E by T)**

## Achievements: innovaR

- R programming language package
- Developed to simulate a complete multi environmental trial (soil, genomics, and phenomics), considering dependence between attributes of VCU and DUS.
- Can be used to validate statistical models based on VCU and DUS data.
- Free and open source.
- It will be available at:  
<https://github.com/danilosarti/innovaR>

## Achievements: AMBARTI



BAYESIAN ADDITIVE REGRESSION TREES FOR GENOTYPE BY ENVIRONMENT INTERACTION MODELS

BY DANILO A. SARTI<sup>1</sup>, ESTEVÃO B. PRADO<sup>1,2</sup>, ALAN N. INGLIS<sup>1,2</sup>, ANTÓNIA A. L. DOS SANTOS<sup>1</sup>, CATHERINE B. HURLEY<sup>1</sup>, RAFAEL A. MORAL<sup>1</sup>, ANDREW C. PARNELL<sup>1,2</sup>

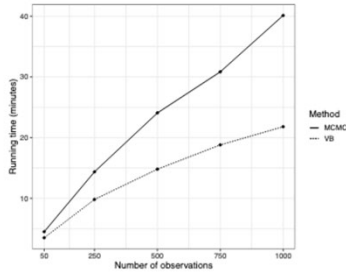
<sup>1</sup>Hamilton Institute, Department of Mathematics and Statistics, Maynooth University, Ireland  
<sup>2</sup>Insight Centre for Data Analytics, Maynooth University, Ireland

**How a DUS or VCU attribute behaves in Several Environments.**

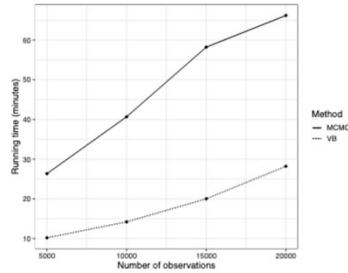
**Visual appeal. Could be an app for farmers.**

**Could we predict the performance of a genotype Considering some genomic attribute?**

## Achievements: Variational Inference



(a) Comparison of simulation times for  $y = \{100, 250, 500, 1000\}$ .



(b) Comparison of simulation times for  $y = \{5000, 10000, 15000, 20000\}$ .

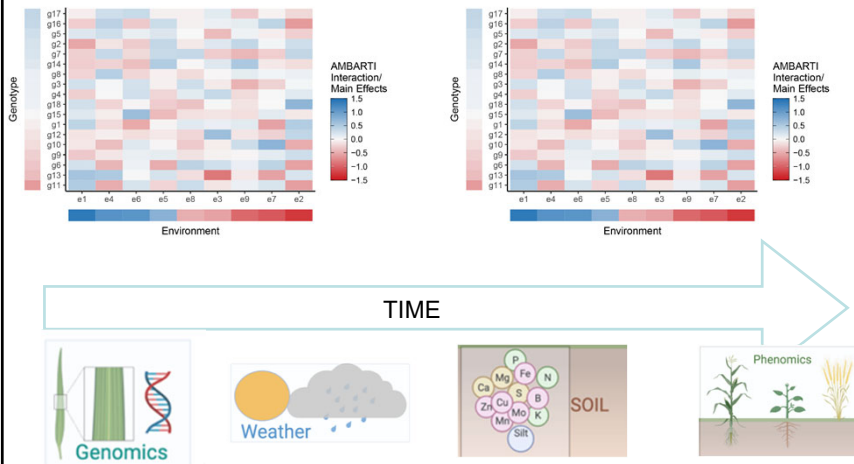
### VARIATIONAL INFERENCE FOR ADDITIVE MAIN AND MULTIPLICATIVE INTERACTION EFFECTS MODELS

BY ANTÔNIA A. L. DOS SANTOS<sup>1</sup>, RAFAEL A. MORAL<sup>1</sup>, DANILO A. SARTI<sup>1</sup>, ANDREW C. PARNELL<sup>1,2</sup>

<sup>1</sup>Hamilton Institute, Department of Mathematics and Statistics, Maynooth University, Ireland

<sup>2</sup>Insight Centre for Data Analytics, Maynooth University, Ireland

Other interactions: G by E by T by Soil  
How could such models incorporate genomic information ... epigenetic markers...



## Short Summary

- InnoVar has been collecting massive data from the soil, weather, genomics, and phenomics.
- The data will be used to develop machine learning algorithms and to create decision support systems.
- There is an opportunity for such models to be used to optimize DUS and VCU by allowing the prediction of the genotypes' attributes across environments and time, considering information on genomics and other sources. This could eventually speed up the process of regulation of new varieties.

[End of Annex and of document]