HAL
open science

Analyzing the impact of agro-environmental conditions on the dynamics of hazelnut yield in Chile

Simone Bregaglio, Laura Giustarini, Laure Hossard, Tommaso de Gregorio

To cite this version:

Simone Bregaglio, Laura Giustarini, Laure Hossard, Tommaso de Gregorio. Analyzing the impact of agro-environmental conditions on the dynamics of hazelnut yield in Chile. 9. International Congress on Hazelnut, International Society for Horticultural Science (ISHS). INT., Aug 2017, Samsun, Turkey. hal-01606428

HAL Id: hal-01606428

https://hal.science/hal-01606428

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyzing The Impact Of Agro-environmental Conditions On The Dynamics Of Hazelnut Yield In Chile

Simone **BREGAGLIO**, Laura **GIUSTARINI**, Laure **HOSSARD**, Tommaso **DE GREGORIO**

IX. International Congress on Hazelnut 15 – 19 August 2017, Samsun, Turkey

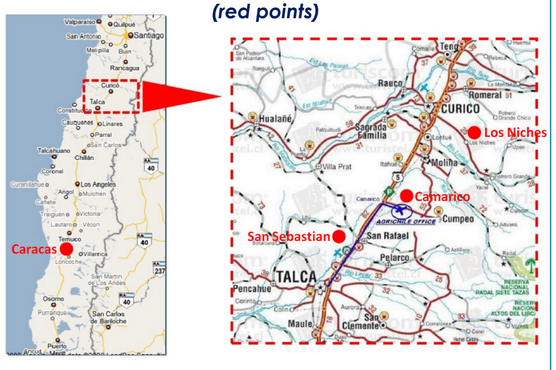
BACKGROUND

- **Worldwide hazelnut production** is steadily **increasing** (+59% from 1984 to 2014), even if **hazelnut cropping systems** markedly **differ in their performances**, with **average yield** ranging **from 0.64 t ha^{-1} in Turkey** - the first exporter worldwide – **to more than 2 t ha^{-1} in United States**.
- **The main reasons associated to this variability are related to the agro-environmental suitability for hazelnut cultivation and to the advancement in the agricultural management of the orchards.**
- This study presents a **statistical analysis** performed **on yield data** collected in **four Chilean orchards** of Ferrero Group, aimed at **understanding the contribution of agricultural and meteorological factors** to hazelnut **yield variability** in the area.

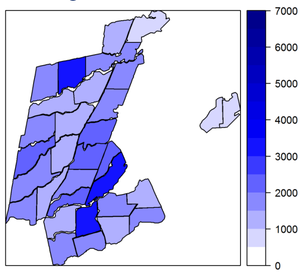
THE STUDY AREA

- The **statistical analysis** was carried out using **yield and production data** from **Chilean** hazelnut orchards placed in **Caracas**, **San Sebastian**, **Camarico** and **Los Niches**.
- Each **orchard is divided into *equipos***, which are plots interested by **the same irrigation system**.

Location of the four orchards (red points)



Yield map in the equipos of Camarico orchard in 2016. Unit: kg ha^{-1}



For **each equipo** in **each orchard**, available data were

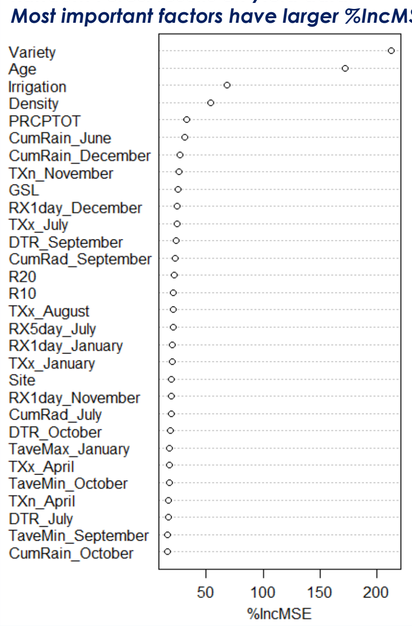
- ✓ **Hazelnut cultivar**;
- ✓ Percentage of **pollinator trees** (%);
- ✓ **Irrigation system**;
- ✓ **Planting density**;
- ✓ **Age** of the trees;
- ✓ **Surface** (ha)
- ✓ Annual **hazelnut yield** (t ha-1)
- ✓ **Production** (t)

THE WORKFLOW OF THE ANALYSIS

- **Database development** → *MS Access database storing **all the data** referred to the **four orchards**, queried by **R software to perform statistical analysis**.*
- **Generation and homogenization of climatic series** → *Climatic series in the period **1996-2016** were downloaded from **NASA Power** and **MSWEP**. **Quality check** to identify **discontinuity points** and **correction** with standard methods.*
- **Selection of relevant factors on yield variability** → *The machine learning technique **Random Forest** was applied using **yield as dependent variable**, and **agro-management** factors as **explanatory variables**.*
- **Description of yield dynamic** → *Fitting of **non-linear mixed effect logistic models** using the **tree age** as **explanatory variable** of hazelnut yield, **separately** for each main **variety** in the orchards.*
- **Inclusion of agro-meteorological factors** → *Hazelnut **yield data** were **divided by tree age** and **linear models** were fitted considering **agro-meteorological indicators**, **planting density** and **irrigation system**.*
- **Evaluation of goodness-of-fit versus expert opinion** → *The **outputs** of the linear models were **compared** with **projections from experts**, to evaluate and compare their **accuracy** in explaining observed **yield variability**.*

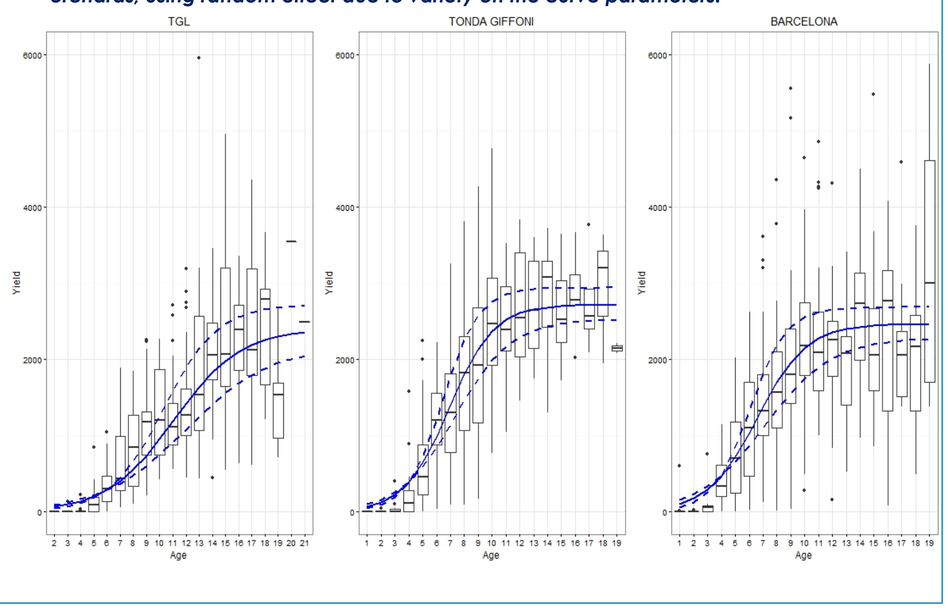
DESCRIPTION OF YIELD DYNAMICS

Random Forest results on yield data. Most important factors have larger %IncMSE



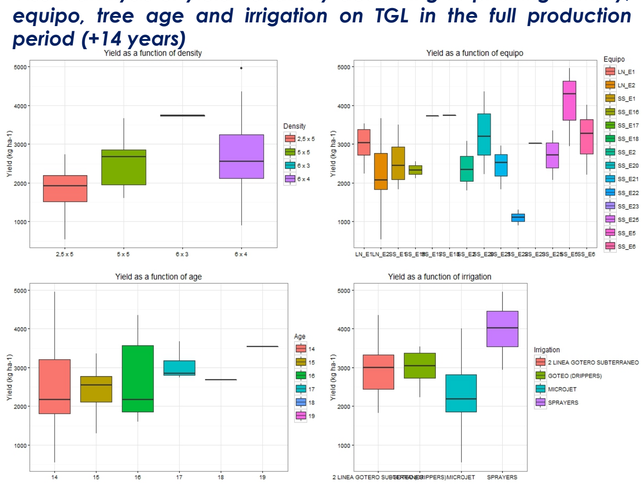
- The output of **Random Forest** (*left figure*) indicates that the **hazelnut variety** and the **tree age** are the **most important** factors **in explaining yield variability**.
- The **irrigation technique** and the **plant density** were ranked as **3rd and 4th factors** in explaining hazelnut yield, whereas **all the agro-environmental indicators** were **less correlated** to yield variability.
- **Three nonlinear growth functions** (logistic, Gompertz and Weibull functions) **were tested** for goodness of fit with yield dynamic data. The **choice of the best model** was driven by the **Akaike Information Criterion** (AIC).
- The **logistic model was selected** considering the effect of **each factor as separated** (**age, density, irrigation**) and acting **as a random effect** on the **three parameters** of the **logistic growth curve**.
- The **best model** resulted with **random effects due to variety on the three parameters of the logistic model** (*right figure*).

Logistic models fitted for the three main varieties present in the four Chilean hazelnut orchards, using random effect due to variety on the curve parameters.



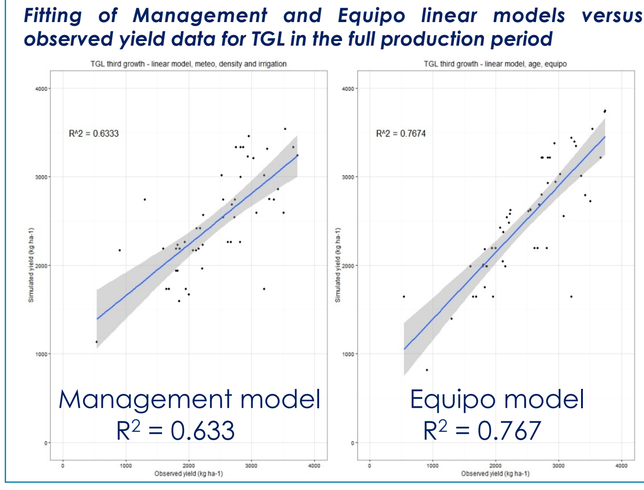
ANALYZING YIELD ACCORDING TO TREE AGE

Visual analysis of yield variability according to planting density, equipo, tree age and irrigation on TGL in the full production period (+14 years)

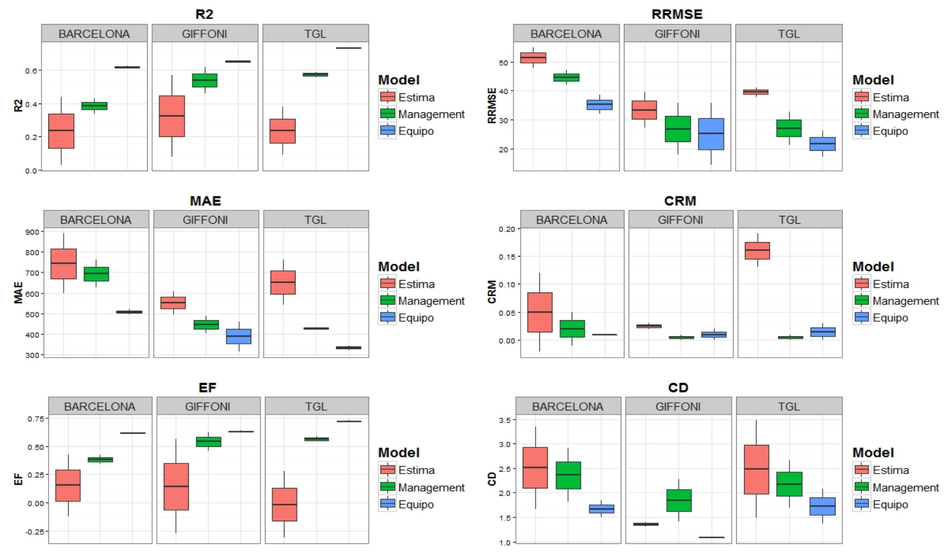


- Based on the **results** of the **previous step**, the hazelnut **yield data grouped by variety** were **divided** in **three growth periods** (0-5 year; 6-12 year; +10 year for Giffoni, + 11 year for Barcelona, +14 year for TGL)
- **For each growth period** and for **each variety**, **linear models** of hazelnut yield **were** then **fitted** including as **explanatory variables**
  1. ***Age, planting density, irrigation technique, the most correlated weather indicators** if not collinear → **Management model***
  2. ***Age, Equipo, the most correlated weather indicators** if not collinear. → **Equipo model***
- A **visual analysis** of the **yield variability** according to **management factors** was carried out, and **ANOVA** was performed **on the two linear models** to test the **significance of** the increase of **accuracy** moving from **Management** to **Equipo** model.

Fitting of Management and Equipo linear models versus observed yield data for TGL in the full production period



COMPARISON WITH EXPERT OPINION DATA



Evaluation indices used to compare the performances of Management and Equipo models with expert opinion data (Estima). MAE = mean absolute error, EF = modelling efficiency, RRMSE = relative root mean square error, CRM = coefficient of residual mass, CD = Coefficient fo determination

- The two linear models **Management** and **Equipo were** then **compared** with **expert opinion data** (**Estima** in the Figure above), to **evaluate** and **compare** their **accuracy in explaining observed hazelnut yield**.
- **Our models** allowed to **substantially improve the accuracy of forecasted hazelnut yield** with respect to Estima data. All the values of the **evaluation indices** are positive, with **better results obtained by Equipo mode**l.
- The **Management model** allowed to **reduce the relative root mean square error** in yield prediction **to 32.8%** with respect to **41.45% of the Estima data**.
- The **Equipo model** allowed to **further reduce the root means square error** in yield prediction to **27.35%.**
- These **improvements were larger for TGL** (**Estima MAE = 652 kg ha^{-1}**), with **MAE of the Management model = 427.62 kg ha^{-1}** and **MAE of Equipo model = 333.18 kg ha^{-1}**.

KEY FINDINGS

- This project allowed to develop **a database of historical hazelnut yield data** related to **the four production sites** of Camarico, Los Niches, San Sebastian and Caracas, which **can be further extended with new data**, in order to **recursively run the statistical procedures** presented here.
- **Historical weather series** referred to the four sites **were generated and can be reused for other applications**, including the **computation** of new **agrometeorological indicators** or as the **source of input data for biophysical modelling**.
- The **logistic models** fitted for TGL, Tonda Giffoni and Barcelona **allowed to differentiate their dynamic growth** as a function of **tree age**, whereas the **linear models** separately fitted for the three growth periods allowed to **reduce the errors in prediction** as compared to expert based assessments.