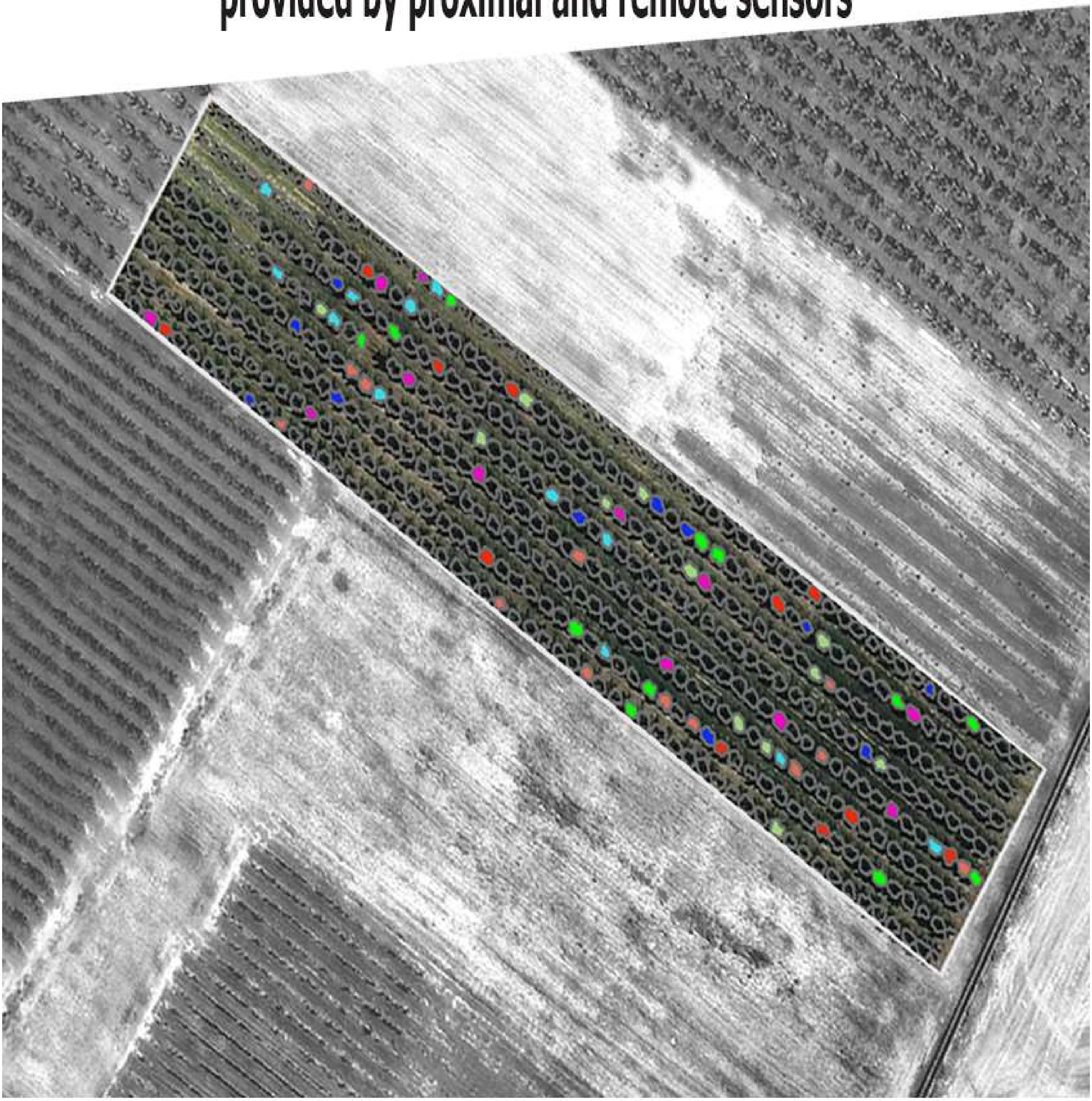
Assessing soil and canopy spatial variability in fruit orchards to improve management and sampling by using auxiliary information provided by proximal and remote sensors

Asier Uribeetxebarria Alonso de Armiño

http://hdl.handle.net/10803/666921

Assessing soil and canopy spatial variability in fruit orchards to improve management and sampling by using auxiliary information provided by proximal and remote sensors



Universitat de Lleida

PhD Thesis by

Asier Uribeetxebarria Alonso de Armiño

May 2019

Grup de Recerca en AgròTICa i Agricultura de Precisió

Research Group in AgroICT & Precision Agriculture

Uxueri eta gurasoei, Jose eta Iosune, bidai honetan nire makulu izatearren

Universitat de Lleida

TESI DOCTORAL

ASSESSING SOIL AND CANOPY SPATIAL VARIABILITY IN FRUIT ORCHARDS TO IMPROVE MANAGEMENT AND SAMPLING BY USING AUXILIARY INFORMATION PROVIDED BY PROXIMAL AND REMOTE SENSORS

Asier Uribeetxebarria Alonso de Armiño

Memòria presentada per optar al grau de Doctor per la Universitat de Lleida
Programa de Doctorat en (Ciencia y Tecnología Agraria y Alimentaria)

Directors

Jaume Arnó Satorra

José Antonio Martínez-Casasnovas

2019

SUMMARY

The aim of this Thesis is the analysis of spatial variability in fruit orchards of the central area of the Ebro Valley. In the first section, soil variability and peach trees (*Prunus pérsica* (L.) vigour were analysed with the goal to obtain site-specific management areas. In the second section, auxiliary information provided by different type of sensors was used to stimate quantitative (kg/tree) and qualitative parameters (fruit firmness and refractometric index) by using advanced sampling techniques.

During this Thesis, data corresponding to two agricultural campaigns (2015 and 2016) were collected. Data of 2015 correspond to the experimental plot that the IRTA (Agrifood Research and Technology Institute) has in Gimenells (Lleida). The most representative feature in this area was the presence of a petrocalcic horizon at a different depth. In 2016 data was collected in a commercial orchard of Utxesa (Lleida). The most outstanding characteristic of the plot was the presence of slightly saline soils and the land transformation accomplished in the 1980s. Regarding data adquisition, both plots were characterized by different sensors (contact and remote) employed in precision agriculture. Finally, soil variability was measured using the soil apparent electrical conductivity (ECa) with a resistivity on-the-go sensor, Veris 3100. In addition, a RGB camera and a multispectral camera, and the OptRX reflectance sensor were used to determine tree vigour. To obtain geometrical parameters of trees, a terrestrial LiDAR sensor was used. At the same time,, and based on the earlier delimited zones and the use of information from previous sensors, different sampling schemes were applied to estimate soil properties and quantitative and qualitative crop parameters.

The use of the ECa, together with a nonlinear, smoothness-constrained algorithme and multivariate techniques such as the ISODATA and a multivariate analysis of variance (MANOVA) allowed delimiting homogeneous management zones, based exclusively on soil variability. The procedure followed allowed to determine with precision the properties that most influenced ECa variability. The greater knowledge over origin of the variability allowed the recommendation of actions to improve orchard management. In a subsequent study, the relationship between soil properties and crop vigour was investigated by incorporating the Normalised Difference Vegetation Index (NDVI) to the analysis. Unexpectedly, tree vigour did not show a correlation with soil ECa. The origin of this lack of relationship could be drip irrigation, which favours the creation of a humid micro-habitat around the bulb. The greatest discordances between ECa and NDVI maps occurred in the places where old terraces were located. The works transform the former plot structure to new larger plots altered soil properties, creating discontinuities. Because of that, two types of actions to improve plot management were proposed.

The efficiency of stratified sampling (StRS) and ranked sampling (RSS) was compared versus simple random sampling (SRS) to estimate qualitative and quantitative fruit parameters. Both techniques, StRS and RSS, reduced necessary sample size by increasing the efficiency. The best results of StRS were obtained using the NDVI as auxiliary information, allowing diminishing yield estimation work up to 17%, without decreasing the accuracy. The stratification by ECa or by combining NDVI + ECa did not improve the results obtained with the NDVI.

Finally, RSS sampling was more efficient than SRS for all sample sizes (N = 4 untill N = 12). The use of RSS in conjunction with a variable strongly correlated with the target variable (fruit load), allowing the reduction of the sampling effort by 40% in comparison to SRS. The selected variable was the projected nadir area of the tree canopy, obtained using a drone. The high resolution image made possible the delimitation of each tree with precision. The process of delimiting the canopy projected area was done manually. Automation of this process through the use of different algorithms opens the doors to a new research line.

Key words: Spatial variability, NDVI, ECa, peach, stratified sampling (StRS), ranked sampling (RSS), random sampling (SRS), MANOVA, land transformation, sensors, soil sensing, spatial analysis.

RESUMEN

Esta Tesis doctoral se centra en el estudio de la variabilidad espacial presente en parcelas frutícolas de la zona central del Valle del Ebro. En un primera parte, se ha analizado la variabilidad del suelo y el vigor de los melocotoneros (*Prunus pérsica* (L.) Stokes) con el fin de obtener zonas de manejo diferenciado. En una segunda parte, la información auxiliar aportada por diferentes sensores ha sido utilizada para estimar parámetros cuantitativos (kg/árbol) y cualitativos (dureza e índice refractométrico) mediante técnicas de muestreo avanzadas.

Los datos utilizados durante la tesis fueron recopilados en el transcurso de 2 campañas agrícolas, 2015 y 2016. Los datos del año 2015 correspondieron a la parcela experimental que el IRTA (Instituto de Investigación y Tecnología Agroalimentaria) dispone en Gimenells (Lleida), en la cual, la característica más representativa era la presencia en el suelo de un horizonte petrocálcico a diferente profundidad. Los datos del año 2016 fueron obtenidos en una parcela comercial en Utxesa (Lleida). La presencia de sales y los movimientos de tierra realizados en la década de 1980 son las características más reseñables de esta parcela. En cuanto a la adquisición de datos, ambas parcelas fueron caracterizadas mediante diferentes sensores (próximos y remotos) empleados en agricultura de precisión. La variabilidad del suelo fue medida mediante la conductividad eléctrica aparente (CEa) proporcionada por un sensor resistivo, Veris 3100 (Veris Technologies). Además del sensor óptico terrestre OptRx (AgLeader), el vigor de la vegetación fue caracterizado mediante una cámara RGB y otra multiespectral transportadas por una avioneta y dron. Por otra parte, los parámetros geométricos de los árboles fueron medidos (caracterizados) mediante un sensor LiDAR terrestre. En paralelo, y en base a zonificaciones previas y al uso de la información de los sensores anteriores, distintos esquemas de muestreo fueron aplicados para estimar propiedades del suelo y parámetros cuantitativos y cualitativos de la cosecha.

El uso de la CEa, junto a un algoritmo no lineal y técnicas multivariantes como el ISODATA y el análisis múltiple de la varianza (MANOVA), permitió delimitar zonas de manejo homogéneas, basándose exclusivamente en la variabilidad del suelo. El procedimiento empleado permitió determinar con precisión las propiedades del suelo que más influyeron sobre la variabilidad de la señal de conductividad. El mayor conocimiento del origen de la variabilidad hizo posible recomendar acciones para mejorar la gestión de la parcela. En un análisis posterior, la relación entre el suelo y la cosecha fue evaluada añadiendo la variable *Normalised Difference Vegetation Index* (NDVI). En contra de lo esperado, el vigor de los árboles no mostró relación con el CEa del suelo. El origen de esta falta de relación se atribuyó al riego por goteo, que favorecía la creación de un micro-hábitat húmedo (bulbo) alrededor de la zona radicular. Las mayores discordancias entre los mapas de CEa y NDVI se dieron en los lugares donde se localizaban las antiguas terrazas de parcelaciones anteriores. Los trabajos de remodelación de las antiguas terrazas alteraron las propiedades del suelo, creando discontinuidades en los patrones de variación espacial. Debido a ello, 2 tipos de actuaciones fueron propuestas para mejorar la gestión de las parcelas.

En cuanto a las técnicas de muestreo en parcela, el muestreo estratificado (StRS) y el *ranked set sampling* (RSS) fueron comparados en términos de eficiencia con el muestreo aleatorio simple (SRS) a la hora de estimar parámetros cualitativos y cuantitativos. Tanto el StRS como el RSS redujeron el tamaño de muestra necesario al aumentar la eficiencia. Los mejores resultados del StRS se obtuvieron empleando el NDVI como información auxiliar, permitiendo reducir el tamaño de muestra necesario un 17% en estimaciones de cosecha (kg/árbol). La estratificación mediante CEa o la combinación de NDVI+CEa no mejoraron los resultados obtenidos con el NDVI.

Finalmente, el muestreo RSS mejoró la eficiencia del muestreo respecto al SRS para todos los tamaños de muestra que fueron evaluados (*N=4* hasta *N=12*). El empleo de este método de muestreo requirió el uso de una variable auxiliar fuertemente correlacionada con la variable a estimar (carga de frutos), obteniendo así muestras que, con igual precisión, podían ser ahora de menor tamaño (un 40% más pequeñas que las muestras SRS). La variable seleccionada fue el área nadir proyectada de la copa del árbol. Esta área fue obtenida mediante un dron y, dada su alta resolución por pixel, posibilitó delimitar individualmente cada árbol con alta precisión. El proceso de delimitación de las copas fue realizado manualmente. La automatización de este proceso mediante algoritmos más precisos y robustos abre las puertas a una nueva línea de investigación.

Palabras clave: Variabilidad espacial, NDVI, ECa, melocotón, muestreo estratificado (StRS), ranked sampling (RSS), muestreo aleatorio (SRS), MANOVA, reparcelación, sensores, análisis espacial.

RESUM

Aquesta Tesi doctoral s'ha centrat en l'estudi de la variabilitat espacial present en parcel·les fructícoles de la zona central de la Vall de l'Ebre. En una primera part, i amb l'objectiu d'obtenir zones de maneig diferenciat, la variabilitat del sòl i el vigor dels presseguers (*Prunus pèrsica* (L.) Stokes) foren analitzats. En una segona part, la informació auxiliar aportada per diferents sensors fou utilitzada per estimar paràmetres quantitatius (kg/arbre) i qualitatius (fermesa i índex refractomètric) mitjançant tècniques de mostreig avançades.

Les dades utilitzades durant la Tesi foren recopilades en el decurs de 2 campanyes agrícoles, 2015 i 2016. Les dades de l'any 2015 correspongueren a la parcel·la experimental que l'IRTA (Institut d'Investigació i Tecnologia Agroalimentària) disposa a Gimenells (Lleida), en la qual la característica més representativa era la presència en el sòl d'un horitzó petrocàlcic a diferent profunditat. Les dades de l'any 2016 foren obtingudes en una parcel·la comercial a Utxesa (Lleida). La presència de sals i els moviments de terra realitzats en la dècada de 1980 són les característiques més destacables d'aquesta parcel·la. Pel que fa a l'adquisició de dades, ambdues parcel·les foren caracteritzades mitjançant diferents sensors (propers i remots) emprats en Agricultura de Precisió. La variabilitat del sòl fou mesurada mitjançant la conductivitat elèctrica aparent (CEa) proporcionada per un sensor resistiu, Veris 3100 (Veris Technologies). A part del sensor òptic terrestre OptRx (AgLeader), el vigor de la vegetació for caracteritzat mitjançant una càmera RGB i una altra multiespectral embarcades en una avioneta i drone. Per altra banda, els paràmetres geomètrics dels arbres foren mesurats (caracteritzats) mitjançant un sensor LiDAR terrestre. En paral·lel, i en base a zonificacions prèvies i l'ús de la informació dels sensors anteriors, diferents esquemes de mostreig foren aplicats per estimar propietats del sòl i paràmetres quantitatives i qualitatives de la collita.

L'ús de la CEa, juntament amb un algorisme no lineal i tècniques multivariants com l'ISODATA i l'anàlisi múltiple de la variància (MANOVA), va permetre delimitar zones de maneig homogènies, basant-se exclusivament en la variabilitat del sòl. El procediment emprat va permetre determinar amb precisió las propietats del sòl que més influïren sobre la variabilitat del senyal de conductivitat. El major coneixement de l'origen de la variabilitat va fer possible recomanar accions per millorar la gestió de la parcel·la. En una anàlisi posterior, la relació entre el sòl i la collita fou avaluada afegint la variable *Normalised Different Vegetation Index* (NDVI) a l'anàlisi. Contràriament a l'esperat, el vigor dels arbres no va mostrar cap relació amb la CEa del sòl. L'origen d'aquesta manca de relació s'atribuí al reg per degoteig que afavorí la creació d'un microhàbitat humit (bulb) al voltant de la zona radicular. Les majors discordances entre els mapes de CEa i NDVI es van donar en els llocs on es localitzaven les antigues terrasses de parcel·lacions anteriors. Els treballs de remodelació de les antigues terrasses alteraren les propietats del sòl, creant discontinuïtats en els patrons de variació espacial. Degut a això, 2 tipus d'actuacions foren proposades per millorar la gestió de les parcel·les.

Pel que fa a les tècniques de mostreig en parcel·la, el mostreig estratificat (StRS) i el *ranked set sampling* (RSS) foren comparats en termes d'eficiència amb el mostreig aleatori simple (SRS) a l'hora d'estimar paràmetres quantitatius i qualitatius. Tant el StRS com el RSS reduïren la mida de mostra necessari a l'augmentar l'eficiència. Els millors resultats del StRS s'obtingueren emprant l'NDVI com informació auxiliar, permetent reduir la mida de mostra necessària un 17% en estimacions de collita (kg/arbre). L'estratificació mitjançant CEa o la combinació de NDVI+CEa no varen millorar els resultats obtinguts amb l'NDVI.

Finalment, el mostreig RSS va millorar l'eficiència del mostreig en relació al SRS per a totes les mides de mostra que foren avaluats (N=4 fins N=12). La utilització d'aquest mètode de mostreig requerí l'ús d'una variable auxiliar fortament correlacionada amb la variable a estimar (càrrega de fruits), obtenint d'aquesta manera mostres que, amb igual precisió, podien ser ara de mida menor (un 40% més petites que les mostres SRS). La variable seleccionada fou l'àrea nadir projectada de la copa de l'arbre. Aquesta àrea fou obtinguda mitjançant un drone i, donada la seva alta resolució per píxel, va possibilitar delimitar individualment cada arbre amb alta precisió. El procés de delimitació de les copes fou realitzat manualment. L'automatització d'aquest procés mitjançant algorismes més precisos i robustos obren les portes a una nova línia d'investigació.

Paraules clau: Variabilitat espacial, NDVI, CEa, préssec, mostreig estratificat (StRS), *ranked sampling* (RSS), mostreig aleatori (SRS), MANOVA, parcel·lació, sensors, anàlisi espacial.

LABURPENA

Tesi honen funtsa, Ebro Bailaran landatzen diren mertxikondoen (*Prunus persica* (L.) Stokes) aldakortasun espazialaren ikerketa da. Lehen pausu batean lurreko aldakortasun espaziala eta fruitu arbolen indarra aztertu ziren, maneiu zona homogeneoak lortzeko asmoz. Bigarren pausu batean, sentsore ezberdinek igorritako informazio lagungarria erabili zen parametro kuantitatibo (fruituaren gogortasuna, °Baumè) eta kualitatiboak estimatzeko laginketa aurreratuen bidez.

Tesi honetan erabilitako datuak 2015 eta 2016 urteetan bildu ziren. 2015-eko datuak IRTAk Gimenells-en duen landa eremu esperimentalean bildu ziren. Bertako ezaugarri aipagarriena sakonera desberdinean garatzen den horizonte pretokaltzikoren presentzia da. Beste alde batetik, 2016. urteko datuak Utxesako landa eremu komertzialean bildu ziren. Lur zati honetan, gatz kontzentrazio nahiko altuak aurkitzen dira, horretaz gain, 80. hamarkadan bertako lurra mugitu zen terraza txikiak eraldatuz lursail handiak eratzeko. 2 lursail hauek Prezisiosko Nekazaritzan erabiltzen diren sentsoreen bidez karakterizatu ziren. Lurreko aldakortasuna ageriko konduktibitate elektrikoaren (ECa) bidez neurtu zen. OptRx sentsoreaz gain RGB kamera bat eta espektru anitz neurtzeko gaitasuna duen kamera bat erabili ziren zuhaitzen indarra neurtzeko. Mertxikondoen propietate geometrikoak LiDAR sentsore baten bidez eskuratu ziren.

ECa-k duen informazioa algoritmo ez lineal batez eta ISODATA eta MANOVA bezalako teknika multibarianteen bidez aztertu zen, lurreko maneiu zona homogeneoak lortzeko asmoz. ECa gehien baldintzatzen duten lur propietateak zeintzuk diren jakitea ahalbidetu zuen erabilitako prozedurak. Aldakortasunaren iturria ezaguna izanik, hau konpontzeko ekintzak proposatu ziren. Bigarren ikerketa batek lurraren eta uztaren arteko elkarrekintzak aztertzea zuen helburu, horretarako ECa-z gain NDVI-a erabili zen. Esperotakoaren aurka, bi aldagaien artean ez zegoela erlaziorik aurkitu zen. Erlazio falta honen sorburua tantaz-tantako ureztatzean egon daitekeela uste da, sustrai sisteman baldintza hezeagoak egotea eragiten baitu. ECa eta NDVI-aren arteko diskordantzia handienak terraza zaharrak zeuden guneetan aurkitu zen. Terrazen birmoldatze lanek lurreko propietateak eraldatu zituzten. Hau horrela izanda 2 ekintza mota proposatu ziren lursailen maneiua obetzeko asmotan.

Mertxikondoen parametro kualitatibo eta kuantitatiboak estimatzeko garaian, laginketa estratifikatua (StRS) eta *ranked set sampling* (RSS) laginketak, zorizko laginketa sinplearen efizientziarekin (SRS) erkatu ziren. StRS eta RSS laginketek SRSa baino efizienteagoak dira eta horren ondorioz beharrezko lagin tamaina murrizten dute hauen efizientzia handituz. StRS-aren emaitza onenak NDVI-a erabiliz lortu ziren. Esate baterako, zehaztazuna galdu gabe uztaren estimazioa %17-an obetu zuen. ECa edo ECa+NDVI bidezko estratifikazioak ez zuen NDVI-aren estimazioa hobetu.

Beste alde batetik, RSS-aren efizientzia SRS baino altuago zela ikusi zen azterturiko lagin tamaina guztietarako (*N=4*, *N=12*). Fruitu kopuruarekin erlazio estua zuen aldagai lagungarri baten erabilerak, esfortzua %40 batean murriztu zuen. Kasu honetan, erlazio estuena zuen aldagaia zuhaitz buruaren area proiektatua izan zen. Aldagai hau, dron batean kokaturiko RGB kamara bateaz lortu zen. Irudiak zuen erresoluzio altuak zuhaitz buruaren definizio zehatza egitea ahalbidetu zuen. Algoritmo trinko eta zehatzagoen bidezko automatizazioak ateak ireki dakioke ikerketa lerro honi.

Hitz gakoak: Aldakortasun espaziala, NDVI, ECa, mertxikondo, laginketa estratifikatua (StRS)*, ranked sampling* (RSS), zorizko laginketa (SRS), MANOVA, lur eraldaketa, sentsoreak, analisi espaziala.

AGRADECIMIENTOS

No podría empezar de otra manera el apartado de agradecimientos que agradeciendo a Jaume y José Antonio, los directores de esta tesis, la buena disposición y la ayuda brindada durante el transcurso de estos 3 años ante cualquier duda o dificultad que he tenido. He de remarcar de una manera especial el tiempo invertido en las correcciones de las diferentes secciones de la Tesis. Hasta recabar en Lleida nunca había pertenecido a un grupo de trabajo tan grande como el GRAP, donde compañeros de diferente origen académico se organizan "asamblеariamente" para trabajar conjuntamente hacia un mismo objetivo. Gracias por dejarme formar parte de este grupo. Bruno, gracias por abrirme las puertas de tu grupo de investigación en Montpellier, merci.

Muchos viajes se emprenden con el objetivo de buscar y acumular nuevas experiencias para enriquecernos. Gracias a la forma peculiar de ver las cosas de mi madre he conseguido contemplar la Tesis como una especie de viaje, que comienzo desde el confort de lo conocido (casa) y voy trabajando y descubriendo diferentes aspectos de mí mismo y las personas que me rodean, completando el saco de las experiencias, hasta culminar este viaje en una nueva zona de confort. Haberme inculcado esta forma de contemplar las cosas no se puede agradecer con palabras. Aún y todo, no ha sido siempre sencillo el camino, y no han sido pocas las veces en las que he pensado mientras conducía de regreso a Lleida "zertan zabitz zu, Lleidara bidian, zertarako balio dotzu honek danak" y no son pocos los kilómetros conducidos, 56 000 km. Aita, momento horretan zu etortzen zare nire burura, pertsona gutxi ezautzen baitittut zu bezain konstanteak, ekindako lanaz eta hartutako konpromezuaz, jo eta ke hauek bukatu arte, betik aurrera. Uxue, zuri gauza asko dauzkot esateko, baina guztiak laburtzen dittuen hitz bat bakarrik idatzikot zuretako, milesker.

Las personas que me conocen saben que el cocinar y el comer son aspectos clave para mí. Por ello invierto mucho tiempo en la cocina, pero si no pudiese disfrutar de una conversación y la compañía de los compañeros a la hora de la comida apenas me merecería recrearme en la cocina. Por ello querría agradecer encarecidamente los momentos compartidos con el grupo de Rius, moltes gracies.

Bueno Efren, no puedo evitar una sonrisa cuando recuerdo las palabras que me dijiste el día 4 de febrero del 2016 nada más conocernos "Kaixo, andas en bici no? que grupo llevas". Entonces pensé en que por lo menos alguien de Lleida compartía mi afición por la bici. Desde entonces hemos compartido muchas horas sobre la bici y hemos sufrido todo tipo de inclemencias climáticas. Desde los 38 C ° subiendo Larrau a los 3 C° bajando Errozate en camiseta. Gracias por acompañarme en estas aventuras alocadas.

Ezingo neuke atal hau amaitu Arrasateko lagunei eskerrak eman barik. Hiru urte hauetan ez dot denbora asko konpartitu zuokin, baina batera egondakoa kalitatezkua izan dala uste dot.

Mila esker guztioi!!!!!

INSTITUTIONAL ACKNOWLEDGEMENT

This thesis is directly linked to the research project AgVANCE: Photonic-based tools for a sustainable agronomic management and use of pesticides in tree crops in the framework of precision farming (AGL2013-48297-C2-2-R), funded by Spanish Ministry of Economy and Competitiveness. On the other hand, this thesis would not have been carried out without the predoctoral training scholarship of the Universitat de Lleida.

AGL2013-48297-C2-2-R

Chapter 1

General introduction

INTRODUCTION

1. Background and rationale of the research

This Thesis is framed into the research project "Photonic-based tools for a sustainable agronomic management and use of pesticides in tree crops in the framework of precision farming” " (AgVANCE project, AGL2013-48297-C2-R, 2014-2017); carried out by The Research Group in AgroICT & Precision Agriculture (GRAP) of the University of Lleida and financed by the Spanish Ministry of Economy and Competitiveness.

The relevance of this Thesis lays in the importance and specific weight that stone fruit production has in the economy of some Mediterranean countries. Particularly, Spain, Italy and Greece concentrate more than 90 % of peach (*Prunus pérsica* (L.) Stokes) production of the European Union (Eurostat, 2015). In Spain, deciduous fruit tree production is the main sub-sector contributing to the total agricultural production, with 31% of it. The main production area is located in Ebro valley, where the 68 % of this production is concentrated. With 11 000 ha dedicated to this type of orchard and with an annual production near to 190 000 tones, the province of Lleida is the most significant production area. In the mid of the past decade (2006-2007), there was a substantial decrease in the area dedicated to peach production in Spain. In 2 years it was reduced from 80 000 ha to 50 000 ha, 32% less (Seoane, 2016). From then until today, the cultivated area has remained, more or less, constant.

In this context, agriculture of societies such as the European, Australian, and North American are facing two significant challenges. On one hand, the area dedicated to agricultural production is gradually decreasing (bancomundial.org); on the other, each habitant consumes more resources. In developing countries, the same trend is observed. In this way, the first challenge of modern agriculture has as purpose to increase production without increasing the area under cultivation, to be able to satisfy the new requirements of modern societies. The second objective tries to pool the first objective with the new ecological conscience. In this way, fruticulture, besides being competitive economically on a global market, must be environmentally sustainable (Collete et al., 2011) and allow the development of small farmers. The key to face both milestones of modern agriculture lies in the management of crops. A key point to manage plots adequately is to adapt it to the spatial variability of the crop (Aggelopoulou et al., 2010, Bramley et al., 2011). In this way, inputs application and distribution of the resources will be carried out in concordance with crop needs. For this, is essential to know the crop phenological stage, since it allows to efficiently organize the necessary resources and to plan work peaks. Harvesting, is considered one of the most complex and expensive tasks of agriculture (Bonora et al., 2014) since it involves a lot of workforces and specialized machinery. Properly planning of harvest has a positive impact on the final costs, since this reduces the workforce. Another advantage of knowing the phenological state of the crop is that allows to harvest the fruits in his optimum moment. Sampling is an efficient option for this (Doraiswamy et al., 2010).

1.1.1. Spatial variability of soil and crops

The achievement of objectives in modern agriculture is hindered by the intrinsic spatial variability of crops (Arnó et al., 2012, Martínez-Casasnovas et al., 2012), and soil (Brevik & Fenton, 2003). The presence of spatial variability in the crops makes it inappropriate to manage homogeneously plots, making the farmers work more difficult. The spatial variability may have a natural or anthropic origin. Human activity can increase soil variability by land movements or interventions to adapt the land to new technologies and machinery (Martínez-Casasnovas & Ramos, 2009). As an example, Figure 1 shows the transformation suffered in an area of 1 km^2 located in the Ebro Valley (Utxesa, Lleida). The image of 1946 shows smaller plots (terraces) adapted to terrain morphology, where olive trees (*Olea europea* K.) and almonds (*Prunus dulcis* (Mill.) D.A. Webb) were cultivated in rainfed land. The image of 2016 shows a very different structure. Terraces were removed through land movements and small plots were redesigned in larger ones. Moreover, rainfed crops were replaced by irrigated crops, such as peaches, nectarines and Saturn peaches, among others. These land movements can concern crop development (Martínez-Casasnovas et al., 2010).

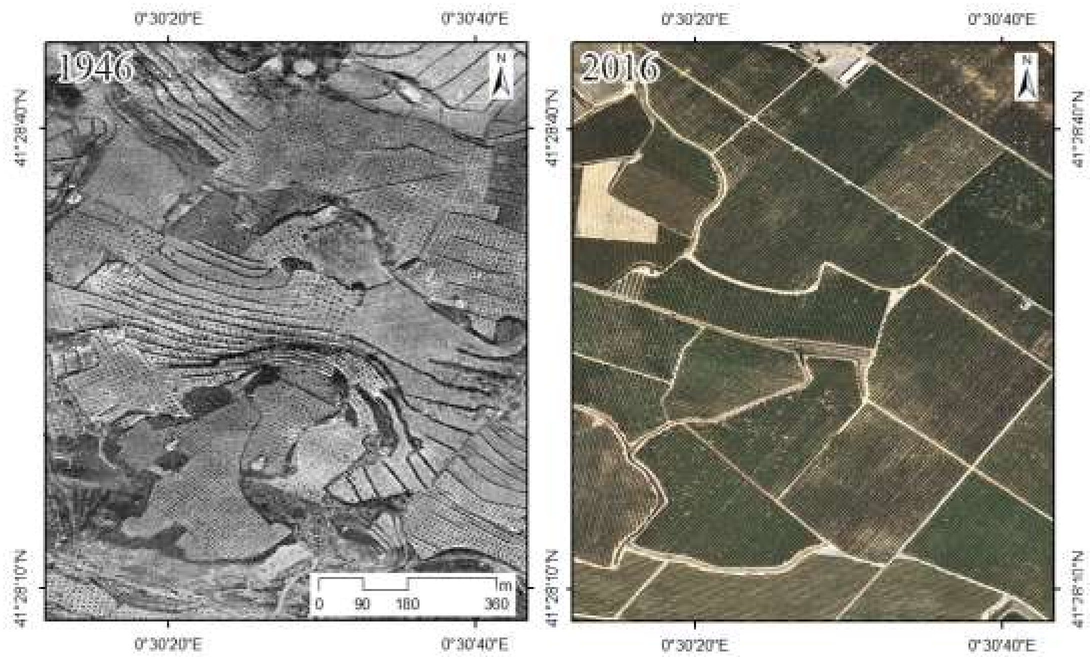


Fig. 1 Comparison of 2 orthophotos that show changes suffered in a representative area of the Ebro Valley during the last 70 years. Left: Orthophoto of 1946 from the area of Utxesa (Lleida) (American Flight Series A, 1945-1946). Right: Same area orthophoto of 2016, scale 1:2500 (Institut Cartogràfic i Geològic de Catalunya). This orthophoto allows seeing variability in crop development in the new plots.

Homogenization of crops is one strategy used by the agricultors to facilitate the management. To achieve this goal, genetically uniform trees are planted following regular patterns (Miranda et al., 2018). In spite of it, crops present spatial patterns, which they should be born in mind at the moment of being managed. Because of this spatial variability, differentiate management at plot level would be essential to get optimum yield (Whelan & McBratney, 2000). In this respect, Precision Agriculture (PA) is nowadays the new paradigm of agriculture which analyze synergies between plot and crop variability (Stafford, 2000). PA is a farming management strategy based on information and technology, which improves the efficiency of production through the adjustment of doses to the plot's intrinsic variability (Stafford, 2000). Differentiated management has as objective carrying out the appropriate intervention at the right time and in the right place (Bongiovanni & Lowenberg-Deboer, 2004). Following these principles, PA approaches previously mentioned milestones of modern agriculture. On one hand, by reducing the number of supplies, environmental impact of agriculture diminishes (Alamo et al., 2012). On the other hand, with a more efficient use of the resources, farmers can get better economic results. In addition, the management of plot zones with different characteristics allows the differentiated sale of the products, obtaining major benefits (Bramley et al., 2005). As a result, these products can get higher prices in the market. For example, Griffin & Lowenberg-DeBoer (2005) affirm that economic benefits of PA are directly correlated to machinery cost. Furthermore, in 68 % of the studied cases, economic benefits improved when managing the plot according to PA principles, in comparison with conventional management.

To carry out site-specific crop management (SCCM) it is necessary to know soil and crop intra-plot variability. In this respect, PA uses sensors for it. Those of photonic base are used to obtain information about crop geometry and vegetative status. Besides, resistivity or electromagnetic induction sensors measure soil apparent electrical conductivity (ECa), and from it other soil properties can be inferred (Corwin & Lesch, 2003). In both cases, data provided by the sensors must be geo-referenced so they can be analyzed and mapped by advanced geostatistical procedures, being the result maps that facilitate their subsequent interpretation.

Figure 2 shows an example of soil variability. In it, a temporal lapse of 11 years is shown. In the image of 2005, bare soil can be observed, with different tonalities that may indicate distinct soil properties (Francis & Schepers, 1997). Right image of Figure 2 shows the plot in 2016. The central zone shows a more intense green vegetation strip. This strip coincides with an area where in 2005 appears with a darker tonality, which probably correspond with deeper soils and/or with greater water retention capacity.

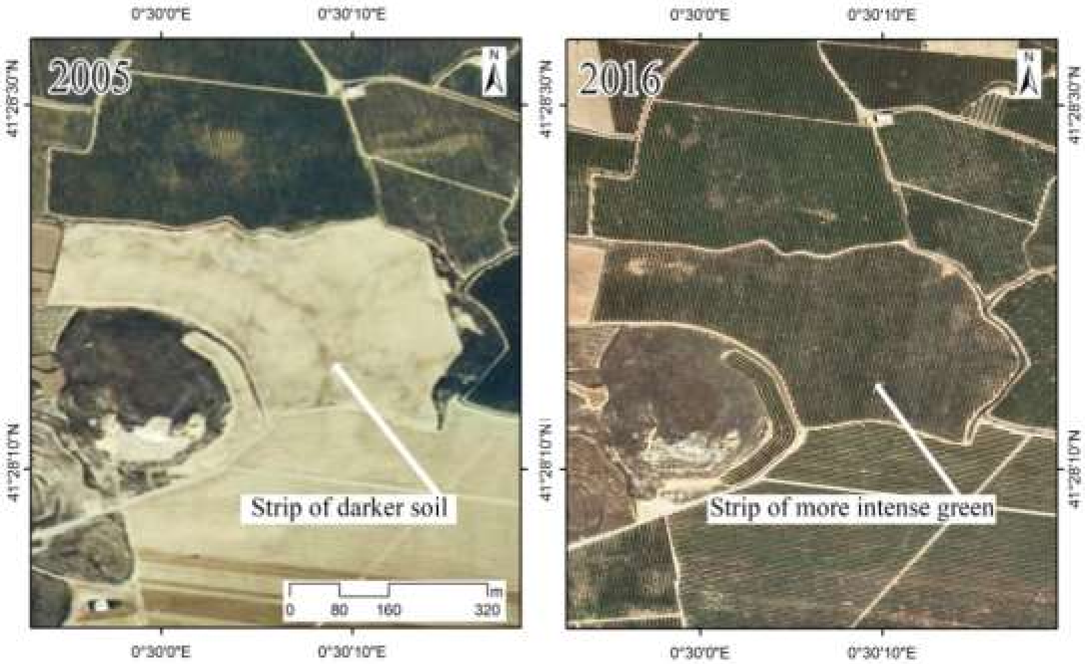


Fig.2 Comparison between 2 orthophotos of the same plot previous to be cultivated (left) and during cultivation (right). The orthophoto of 2005 shows bare soil with different tonalities, indicative of distinct properties. In the orthophoto of 2016, the variability of fruit trees vigour is observed.

Soil variability influences crop development (De Benedetto et al., 2013), therefore, in a specific plot, it is necessary to take several samples to capture that variability. Due to the high cost and effort required for intensive sampling, an appropriate number of samples is not usually taken (Peralta & Costa, 2013). With classical soil sampling techniques, spatial resolution required for PA is not achieved (Mertens et al., 2008). However, non-invasive sensors, such as the Veris 3100 resistivity sensor (Veris Technologies Inc., Salina KS, USA), or the EM38-MK2 electrical conductivity meter (Geonics Ltd, Mississanga, Ontario, Canada), can help to streamline and reduce the price of the soil sampling process. These sensors, in conjunction with a few number of soil samples, allow detailed mapping of soil properties (Corwin & Lesch, 2005), and with significant fewer sampling effort (James et al., 2003).

As mentioned above, the ECa signal depends on soil composition (Brevik & Fenton, 2002). Theoretically, soil properties that more affect ECa are salinity, moisture content and texture (Williams & Baker, 1982; Corwin & Lesch, 2005). Doolittle & Brevik (2014) compiled some works where these relationships were analyzed individually. Nevertheless, soil is an entity composed by a mixture of organic compounds, liquids, minerals, gases and alive organisms that interact between them. This makes difficult to determine the contribution of each soil property to the ECa variability. This makes inadvisably to determine the contribution of each property to the signal through a univariate analysis. To solve this problem, multivariate techniques can be used. For example, Morari et al. (2009) and De Benedetto et al. (2012) used multivariate geostatistics to predict clay, sand and gravel concentration from ECa. However, and despite not being the most appropriate technique to test differences between specific management zones, univariate statistics are still used. Thus, from an agronomic point of view, it would be interesting to analyze the usefulness of multivariate statistics to differentiate soil properties and to analyse the contribution of each of them to ECa variability.

High concentration of calcium carbonate ($CaCO_3$) is common in Ebro Valley soils. Moreover, some areas can develop a petrocalcic horizon, which can limit the growth of the root system. Despite this, there are few studies where relation between the concentration of calcium carbonate and ECa signal variability is analysed. Then, knowing how $CaCO_3$ concentration affects the ECa signal is a pending research issue. In the same way, something similar happens with land transformations carried out to enlarge plotss to favour the mechanization of agricultural works. These operations are widespread since 1970s-80s (Martínez-Casasnovas & Ramos, 2009), being able to affect crops' growth by altering the spatial distribution of soil properties. In this respect, ECa could be an option to identify the areas where soil movements have been carried out.

1.1.2. Sampling for yield stimation

A key element in fruit production management is yield forecast. Field sampling is considered the more reliable method for that. Sampling has advantages over gather all individuals of a population. For example, the sampling reduces the costs. Also, it speeds up the process of obtaining data, which is of vital importance when information is needed almost at once. Another advantage of sampling is that it can increase accuracy, since counting a large population presents difficulties and can lead to errors. However, if the number of individuals to be sampled is reduced, they are supervised in more detail (Cochran, 1977). In spite of this, sampling is not a recurring theme in PA, so there are few papers published in specialised journals and conferences.

The inadequate planning of the samplings is the Achilles hell of many studies (Kerry et al., 2010). Simple Random Sampling (SRS) is the most used sampling technique in agriculture. The most significant advantage that it presents is its facility of implementation in the field and the subsequent interpretation. One disadvantage of the SRS is that it does not consider the spatial variability of the crop at the time of sampling (Cambardella et al., 1999; Perry et al., 2010). Therefore, SRS is inefficient to estimate parameters with spatial patterns (Webster & Lark, 2013). SRS is included within the sampling group denominated as "Design Based" (Brus & Gruijter, 1997). The same group includes sampling techniques such as stratified random sampling (StRS) or ranked set sampling (RSS), which consider the spatial variability of the crop when sampling. Sampling based on geostatistical techniques (model based) is another option. Geostatistic is based on the premises that nearby points are more similar than distant ones (Oliver et al., 2010). In this Thesis, the possibilities that this type of sampling offers to estimate the crop yield there has not studied in depth. However, geostatistics has been a crucial tool to get necessary information for nourishes the other sampling schemes.

To perform an adequate sampling by considering crop variability reinforces the quality and precision of the subsequent analyses. Moreover, information quality depends on prediction accuracy. For this, to have detailed information which captures orchard attributes and soil variation is very important. At the time to sample, the methods used in agriculture are manual and destructive, which can be a substantial handicap. Usually, collection and process of samples is expensive (Webster & Oliver, 1990). Because of that, the knowledge of the phenological status and consequently management of the plots has historically been based on measures of few points, which only allowed the homogeneous management of orchards. However, PA promotes differentiated plot management and, for it, the spatial variability has to be known in detail. This can be achieved by increasing sampling intensity (Kerry et al., 2010), or by using sampling strategies that consider intrinsic variability. An example would be the schemes based on auxiliary information provided by sensors used in the PA.

One of the options is the use of auxiliary variable maps spatially correlated with the target variable. Digital terrain analysis or remote sensing is options to get these maps. For example, soil depth is a variable correlated with slope. Therefore, soil depth could be sampled considering the slope, which is an auxiliary variable easy to derive from a digital elevation model (DEM). Usually, multiple linear regressions are the way to establish the relationship between both variables (Hengl et al., 2003). This approach is known as "environmental correlation" (McKenzie & Austin, 1993).

Wulfsohn (2010) presented a review on sampling techniques used in PA. One of the mentioned techniques is the Stratified Random Sampling (StRS), which tries to reduce sampling variability by creating homogeneous strata (with less variability within the group or stratum). These homogeneous strata can be delimited from auxiliary information acquired by sensors used in PA. Despite the possible advantages of StRS, this method is not widely used in PA. Wulfsohn (2010) did not present the technique called Ranked Set Sampling (RSS) (McIntyre, 1952), since it was not used in PA until today. RSS, through an iterative process based on the distribution of an auxiliary variable correlated with the target variable, allows obtaining more representative samples (more information about the distribution of the population).

Within this interdisciplinary framework, this research focuses on the study of spatial variability of fruit orchards (peach). For this, geo-referenced information acquired from several sensors was used (Veris3100, Digital Multispectral Camera, LiDAR, and OptRX). As explained above, soil variability analysis is inherently complex. Because of that, the impact of soil movements on soil variability and the effects of these movements on crop development have been less studied than other parameters. Fruit orchards are variable, therefore, sampling to estimate the yield and to know the phenological status should contemplate this variability.

1.2. Aims, hypothesis and objectives

The main aim of this Thesis is to deepen in the utility of auxiliary information based on non-invasive sensors to measure and assess within-field spatial variability in fruit growing plots. Because of this spatial analysis, site-specific fruit management based on plot zoning and on efficient sampling to estimate yield and quality are the two areas in which this Thesis is focused on.

The following hypotheses were contrasted:

- **H1** High concentration of $CaCO_3$ in soils affects variability of ECa.
- **H2** Lands transformations carried out to enlarge fields and favour mechanization have broken soil continuity and induce crop variability.
- **H3** Auxiliary information obtained from non-invasive sensors allows sampling effort and efficiency to be improved in fruit orchards.
- **H4** Auxiliary information can facilitate use of Ranked Set Sampling as new sampling procedure in fruit orchards.

In order to response these hypotheses, the specific objectives of the Thesis are:

- **O1** To analyse the utility of a resistivity sensor, Veris 3100, as a tool to detect and measure the spatial variability of soil properties in soils with high contents of calcium carbonate.
- **O2** To study the effects of land transformations in soil variability and in the development of trees in fruit orchards.
- **O3** To assess the efficiency of stratified sampling schemes based on auxiliary information, such as NDVI and ECa, compared to simple random sampling.
- **O4** To determine the most appropriate auxillary variable to use in Ranked Set Sampling schemes to compare efficiency with simple random sampling in yield estimation in peach orchards.

1.3. Thesis Structure

Figure 3 shows the structure of the Thesis with the different chapters, associated objectives, hypothesis and journals where papers have been published or are in the process of publication. The main objectives of the Thesis are addressed in the focal chapters 3 to 6, with each of them corresponding with a scientific paper. Following the regulation of the University of Lleida (*Article 28 of the Normativa acadèmica de doctorat de la Universitat de Lleida*), which specifies the requirements of the Thesis based on article format, each of the focal chapters can be considered as a self-contained unit. It means that they have an introduction section outlining the specific research context and objectives of the chapter, the full details of the methods used, the results, discussion and conclusions. The present chapter represents a general introduction of the Thesis. Chapter 2 provides an overview of the methods used for the respective focal chapters. Finally, in chapter 7, a comprehensive discussion of the results obtained in focal chapters 3 to 6 is presented, allowing general conclusions to be drawn in relation to each of the objectives of the work. General conclusions of the Thesis are presented in chapter 8.

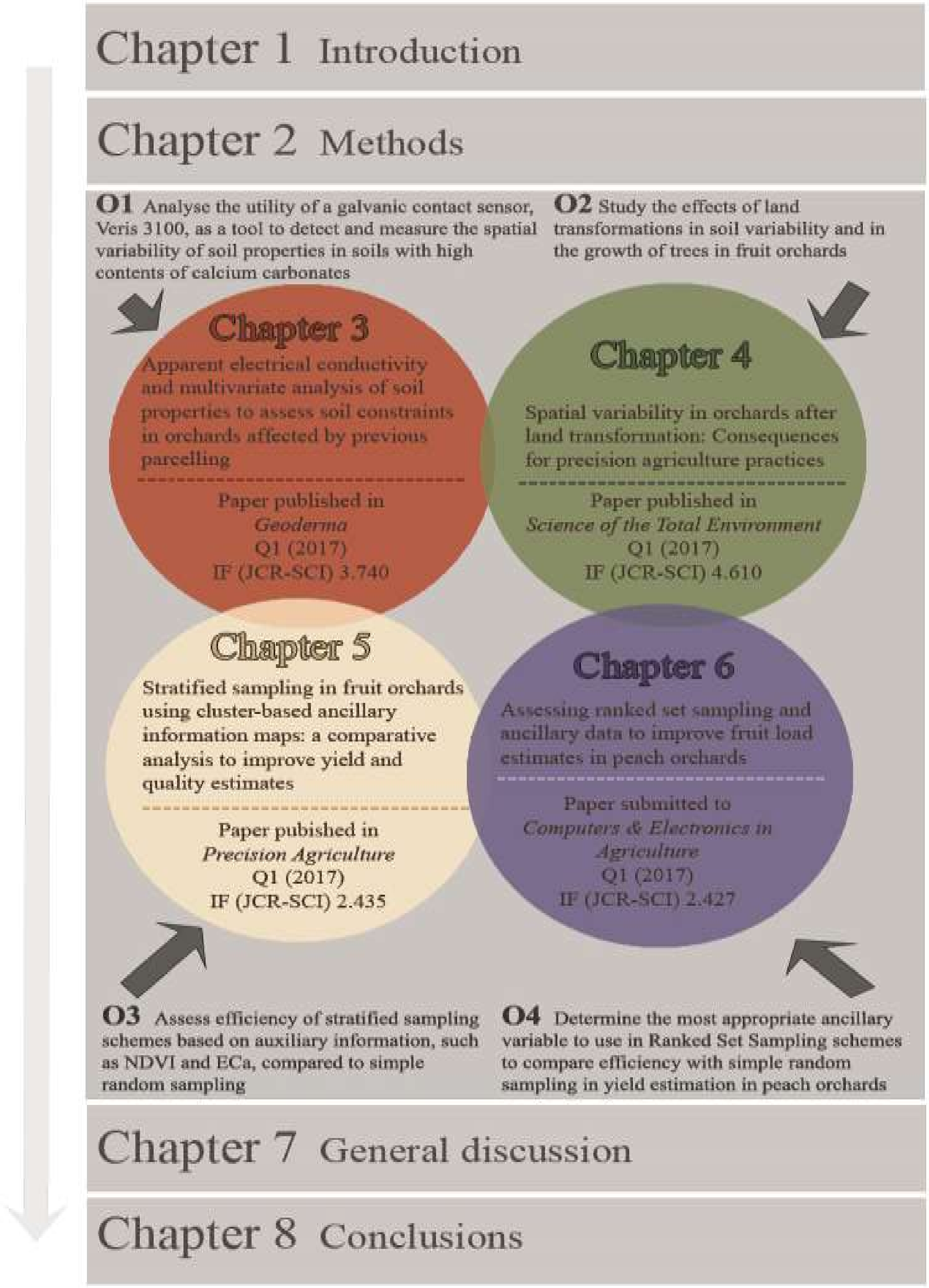


Fig. 3 Structure of the Thesis with the different chapters, associated objectives and journals where papers have been published or have been submitted.

1.4. Thesis timeline

This Thesis was performed within the framework of a 3-years doctoral fellowship funded by the University of Lleida, from February 2016 to February 2019. The timeline of the different studies that comprise the Thesis is shown in Figure 4.

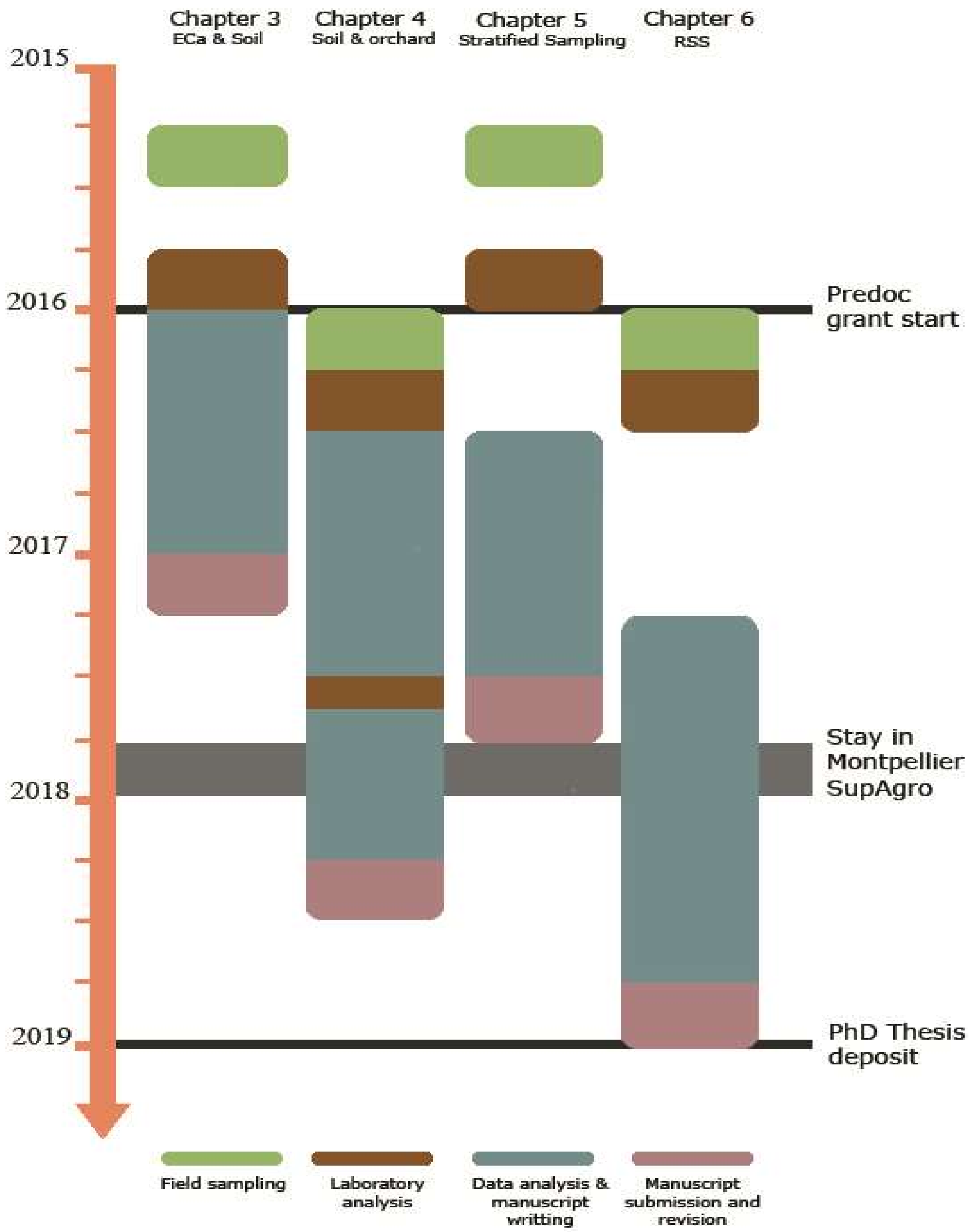


Fig. 4 Timeline of the Thesis where chronologically main temporal tasks are detailed.

References

Aggelopoulou, K., Wulfsohn, D., Fountas, S., Nanos, G., Gemtos, T., Blackmore, S. 2010. Spatial variability of yield and quality in an apple orchard. *Precis. Agric*. 11, 538–556.

Alamo, S., Ramos, M.I., Feito, F.R., Cañas, J.A. 2012. Precision techniques for improving the management of the olive groves of southern Spain. *Span. J. Agric. Res*. 10, 583–595.

Arnó, J., Rosell, J.R., Blanco, R., Ramos, M.C., Martínez-Casasnovas, J.A., 2012. Spatial variability in grape yield and quality influenced by soil and crop nutrition characteristics. *Precis. Agric*. 13, 393-410.

Banco Mundial, Agriculture value. https://datos.bancomundial.org/indicador/_nv.agr.totl.zs (31/10/2018)

Bongiovanni, R. and Lowenberg-DeBoer, J. 2004. Precision agriculture and sustainability. *Precis. Agric*. 5, 359–387

Bonora, E., Norferini, M., Stefanelli, D. Costa, G. 2014. A new simple modeling approach for the early prediction of harvest date and yield in nectarines. *Sci. Hortic*. 172, 1–9.

Bramley, R.G.V., Ouzman, J., Thornton, C. 2011. Selective harvesting is a feasible and profitable strategy even when grape and wine production is geared towards large fermentation volumes. *Aust. J. Grape Wine R*. 17(3), 298–305.

Bramley, R.G.V., Proffitt, A.P.B., Hinze, C.J., Pearse, B., Hamilton, R.P. 2005. Generating benefits from precision viticulture through selective harvesting. In J. V. Stafford (Ed.), Precision Agriculture '05, Proceedings of the 5th European Conference on Precision Agriculture (pp. 891–898). Wageningen: Wageningen Academic Publishers.

Brevik, E.C., Fenton, T.E. 2002. The relative influence of soil water, clay, temperature, and carbonate minerals on soil electrical conductivity readings with an EM-38 along a Mollisol catena in central Iowa. *Soil Surv. Horiz*. 43, 9–13.

Brevik, E.C., Fenton, T.E. 2003. Use of the Geonics EM-38 to delineate soil in a loess over till landscape, southwestern Iowa. *Soil Surv. Ho*riz. 44, 16–24.

Brus, D.J., de Gruijter, J.J. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*. 80, 1–44.

Cambardella, C.A., Karlen, D.L. 1999. Spatial analysis of soil fertility parameters. *Precis. Agric*.1, 5-14.

Cochran, W.G. 1977. Sampling Techniques. New York, NY, USA: John Wiley & Sons, Inc. 428 pp.

Collete, L., Hodgkin, T., Kassan, A., Kennmore, P., Lipper, L., Nolte, C., Stamoulis, K., Steduto, P. 2011. Ahorrar para crecer. *FAO*.1-116.

Corwin, D.L., Lesch, S.M., 2005. Apparent soil electrical conductivity measurements in agriculture. *Comput. Electron. Agric*. 46, 11–43.

Corwin, D.L. & Lesch, S.M. 2003. Application of soil electrical conductivity to precision agriculture: theory, principles and quidelines. *Agron. J*. 95, 455-471.

De Benedetto, D., Castrignanò, A., Rinaldi, M., Ruggieri, S., Santoro, F., Figorito, B., Gualano, S., Diacono, M., Tamborrino, R. 2013. An approach for delineating homogeneous zones by using multi-sensor data. *Geoderma.* 199, 117-127.

De Benedetto, D., Castrignanò, A., Sollitto, D., Modugno, F., Buttafuoco, G., lo Papa, G. 2012. Integrating geophysical and geostatistical techniques to map the spatial variation of clay. *Geoderma*. 171– 172,53–63.

Doolittle, J.A., Brevik, E.C., 2014. The use of electromagnetic induction techniques in soils studies. *Geoderma*. 33–45.

Doraiswamy, P.C., Akhmedov, B., Beard, L., Stern, A., Mueller, R. 2006. Operational predictions of crop yields using MODIS data and products. In B. Baruth, A. Royer, & G. Genovese (Eds.), ISPRS archives XXXVI-8/W48 workshop proceedings: Remote sensing support to crop yield forecast and area estimates (pp. 45–49). Ispra, Italy: Joint Research Centre.

Eurostat Statistics Explained. Data extracted in 2015. Agricultural production – crops.

Francis, D.D., Schepers, J.S. 1997. Selective soil sampling for site-specific nutrient management. In: Stafford JV (Ed), Precision Agriculture ’97, Proceedings of the 1st European Conference on Precision Agriculture. Oxford, UK: BIOS Scientific Publishers, Vol. 1, pp. 119–126.

Griffin,T.W., Lowenberg-DeBoer, J. 2005. Worldwideadoptionandprofitability of precision agriculture. Revista de Politica Agricola 14, 20–38

Hengl, T., Rossiter, D.G., Stein, A. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Aust. J. Soil Res*. 41(8), 1403.

James, I.T., Waine, T.W., Bradley, R.I., Taylor, J.C., Godwin, R.J. 2003. Determination of soil type boundaries using electromagnetic induction scanning techniques. *Biosyst. Eng*. 86 (4), 421-430.

Kerry, R., Oliver, M.A., Frogbrook, Z.L. 2010. Sampling in Precision Agriculture. Sampling in precision agriculture. M.A. Oliver (Ed.), Geostatistical Applications for Precision Agriculture, Springer Science Business Media B.V (2010), pp. 35-63.

Martínez-Casasnovas, J.A., Agelet-Fernández, J., Arnó, J., Ramos, M.C. 2012. Analysis of vineyard differential management zones and relation to vine development, grape maturity and quality. *Span. J. Agric. Res*. 10(2) 326-337.

Martínez-Casasnovas, J.A., Ramos, M.C. 2009. Soil alteration due to erosion, ploughing and levelling of vineyards in north east Spain. *Soil Use Manag*. 25,183–192.

Martínez-Casasnovas, J.A., Ramos, M.C., Espinal-Utgés, S. 2010. Hillslope terracing ef- fects on the spatial variability of plant development as assessed by NDVI in vineyards of the Priorat region (NE Spain). *Environ. Monit. Assess*. 163,379–396.

McIntyre, G.A. 1952. A method for unbiased selective sampling, using ranked sets. Aust. J. *Agric. Res*. 3, 385-390. (Reprinted in *Am. Stat*. 2005. 59(3), 230-232).

McKenzie, N.J., Austin, M.P. 1993 A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma*. 57, 329–355.

Mertens, F.M., Pätzold, S., Welp, G. 2008. Spatial heterogeneity of soil properties and its mapping with apparent electrical conductivity. *J. Plant Nutr. Soil Sci*. 171,146–154.

Miranda, C., Santesteban, L.G., Urrestarazu, J., Loidi, M., Royo, J.B., 2018. Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. *Agriculture*. 8(6), 78.

Morari, F., Castrignano, A., Pagliarin, C. 2009. Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geoelectrical sensors. *Comp. Electron. Agric*. 68, 97–107.

Peralta, N.R., Costa, J.L. 2013. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Can. J. Soil Sci*. 93 205-218.

Perry, E.M., Dezzani, R.J., Seavers, C., Pierce, F.J. 2010. Spatial variation in tree characteristics and yield in a pear orchard. *Prec. Agric*. 11, 42–60.

Seoane, P., Barbier, I., López-Pérez, J.J. 2016. Anuario de estadística Avance 2015. Ministerio de Agricultura, Alimentación y Medio Ambiente.

Webster, R., Lark, R.M. 2013. Field sampling for environmental science and management (p. 192). London and New York: Routledge

Webster, R., Oliver, M.A. 1990. Statistical methods in soil and land resource survey. Spatial information systems. (Oxford University Press: Oxford, UK)

Williams, B.G., & Baker, G.C. 1982. An electromagnetic induction technique for reconnaissance surveys of soil salinity hazards: Australian Journal of Soil Research. 20, 107–118.

Wulfsohn D. 2010. Sampling Techniques for Plants and Soil. Landbauforschung Volkenrode, (Special Issue)

Chapter 2

Methods

1. GENERAL WORKFLOW AND SHORT OVERVIEW OF SENSORS AND TECHNIQUES USED DURING THE THESIS

1.1 Thesis workflow

This chapter shows the general workflow of the research and the main characteristics of the sensors and techniques used. Figures 1 and 2 show the general workflow. Blue boxes represent the necessary steps to achieve the objectives described in chapter 1. Green letters contain the sensors used to measure soil and tree vigour variability. Finally, several sampling techniques and statistical procedures used during the research appear in brown letters. In the following sections, a brief description of the sensors (section 2), in addition to the statistical methods used (section 3), and the sampling techniques assessed in the Thesis (section 4) are provided.

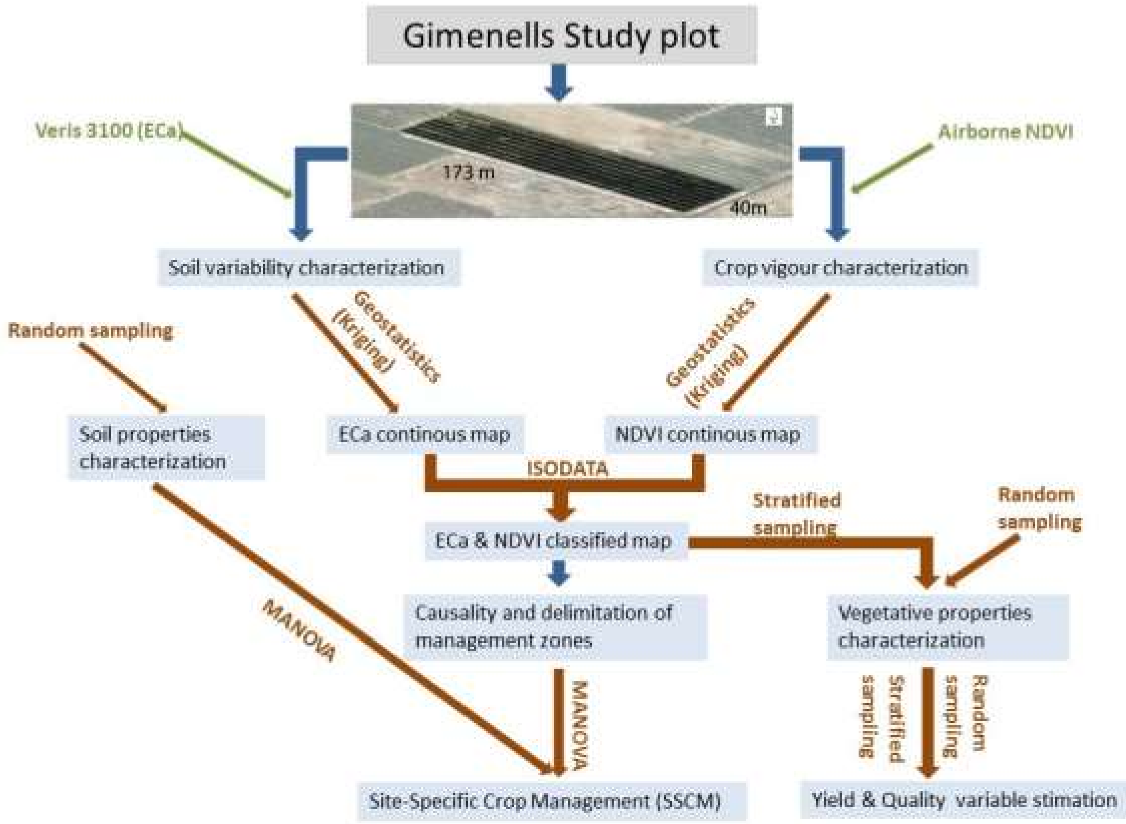


Fig. 1 Schematic illustration of the general workflow of chapters 3 and 5 (Gimenells study plot). The blue boxes indicate significant milestones to achieve the objectives of the chapters. The sensors used in these chapters appear in green letters. Brown letters are indicatives of different types of techniques.

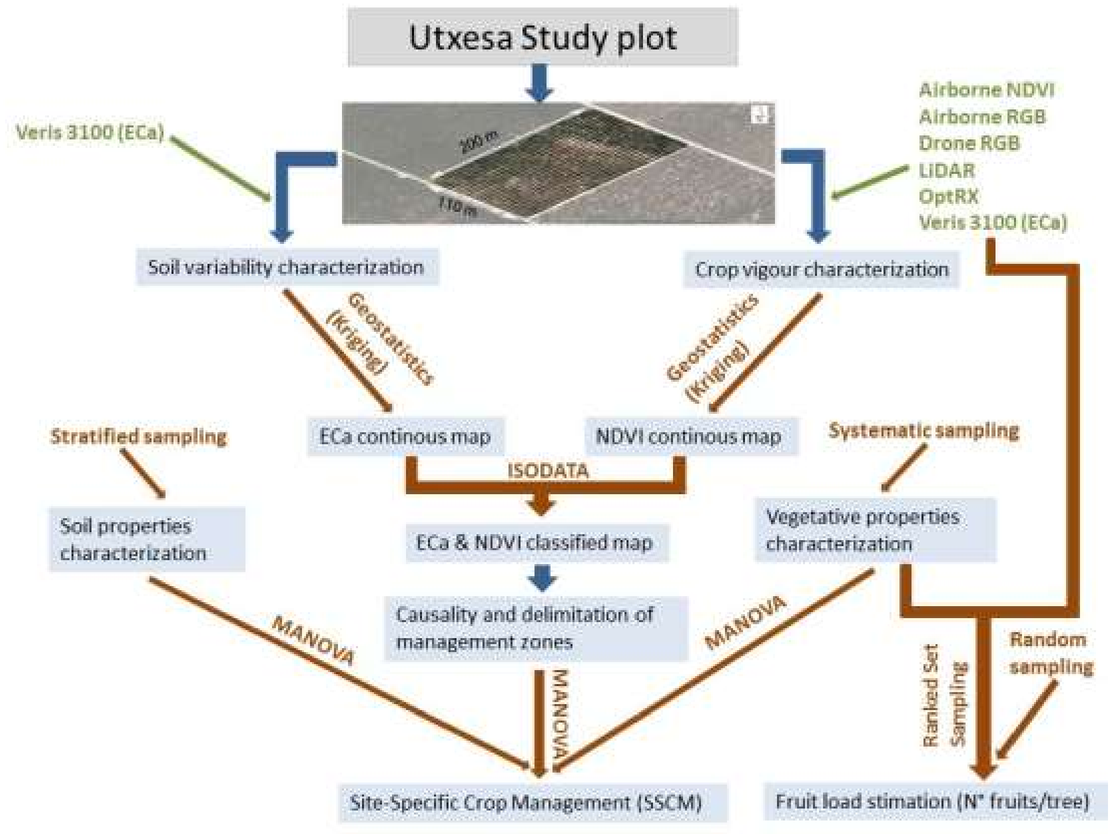


Fig. 2 Schematic illustration of the general workflow of chapters 4 and 6 (Utxesa study plot). In the blue boxes, significant milestones to achieve the objectives of the chapters. The sensor used in these chapters appears in green letters, as well as brown letters are indicatives of different types of techniques.

2. SENSORS USED DURING THE RESEARCH

2.1. On-the-go soil ECa mapping with Veris 3100

Veris 3100 (Veris Technologies, Inc., Salina, KS, USA) is a galvanic contact soil sensor. This sensor is a simple and effective tool to acquire on-the-go information on soil bulk electrical conductivity for subsequent mapping. This type of sensor measures the electrical resistivity of the soil, which is converted into apparent electrical conductivity (ECa), a variable correlated with several soil properties (Corwin & Lesch, 2003). Its advantage lies in the use of two electrical arrays that allow capturing the electrical readings in two different depths of the soil simultaneously. Furthermore, the presence of metals does not interfere the readings. The Veris 3100 equipment consists of six heavy-duty coulter-electrodes, each of one only need to penetrate the soil a few inches. While a pair of electrodes (Figure 3, coulters 2-5) injects electrical current into the soil, the other two pairs measure the voltage drop (Figure 3). Penetration of the electrical current into the soil and, by extension, the volume of soil explored increases as the inter-electrode space increases. The effective depth of electrical arrays is well documented in the scientific literature (Milsom, 1989). In the present case, arrays configuration allowed soil depths of 0–30 cm (shallow ECa reading) and 0–90 cm (deep ECa reading) to be explored (Figure 3). More precise specifications and frequently asked questions (FAQS) of Veris 3100 can be consulted in the web (www.veristech.com/the-sensors/v3100).

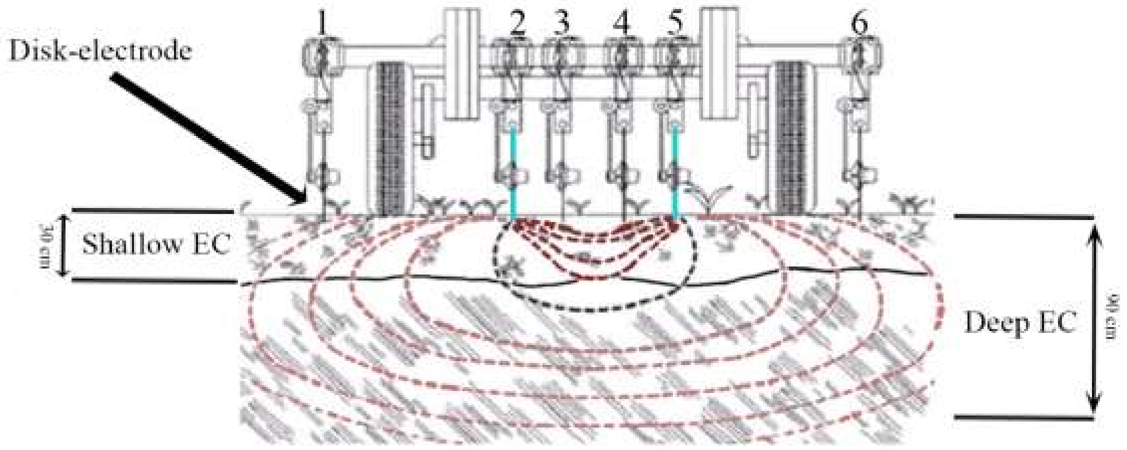


Fig. 3 Operation principle of the contact type ECa sensor. Blue coulters act as transmitting electrodes and the others act as receptors. In red, action range of shallow coulters, in pink, action range of deep coulters.

2.2. Quantifying the crop vigour using aerial and terrestrial platforms and multispectral images

In this Thesis, multispectral data was used to get different vegetation indexes (VIs). These VIs are employed to quantify remotely the vegetative state of fruit trees. A VI is built by a spectral transformation of two or more bands to improve the knowledge about vegetative properties. Combining reflectance values corresponding to different bands of the spectrum, a numeric value per pixel is obtained (Hall et al., 2002). Different VIs can be used allowing temporal and spatial variability to be assessed (Huete et al., 2000). In agriculture, one of the most used band to calculate VIs is the near-infrared (780-2500 nm), since it allows differentiating healthy and vigorous plants from other stressed (Hall et al., 2002).

Throughout the Thesis, two types of sensors were used to capture multi-spectral data. On one hand, a digital multispectral camera (DMSC) (Figure 4, A) mounted on an aeroplane (CESSNA 1725 Sky Hawk) (Figure 4, B). Images captured with this sensor have a spatial resolution of 25 cm^2 per pixel. Each of the pixels captured with the DMSC has 4-band, centred at 450 nm (blue), 550 nm (green), 675 nm (red) and 780 nm (near infrared). Normalized Difference Vegetation Index (NDVI) was obtained through the combination of the red and near-infrared band as proposed by Rouse et al., 1974 (Table 1). This type of aerial images allows vegetative development and vigour of tree canopy to be measured. On the other hand, there is the OptRx terrestrial reflectance sensor (Ag Leader Technology, Ames, IA, USA) (Figure 4, C), that was used to measure the lateral reflectance of the crop. The OptRx can provide, in addition to the NDVI, the NDRE (Normalized Difference Red Edge) proposed by (Gitelson & Mezlyak, 1994) (Table 1).

Table 1 Spectral vegetation indices used in the Thesis

Spectral index	**Reference**
Normalized Difference Vegetation Index (NDVI) $$NDVI = \frac{NIR - Red}{NIR + Red}$$	Rouse et al. 1974
Normalized Difference Red Edge $$NDRE = \frac{NIR - RE}{NIR + RE}$$	Gitelson & Mezlyak, 1994

The OptRx works as an active sensor that projects a light beam to measure and record in real time the reflectance of the light shined on the growing plants (Figure 4, D). NDVI is recommended to measure crops in the early stage of development, while NDRE is more suitable to use for big crops as maize (*Zea mays* L.) or late stages in small crops as wheat (*Triticum sp*.). The OptRx sensor can be coupled with almost all types of terrestrial vehicles. Normally used with nadir viewing angle, reflectance data of the trees have been captured by positioning the sensor to take lateral readings of the canopy along the rows.

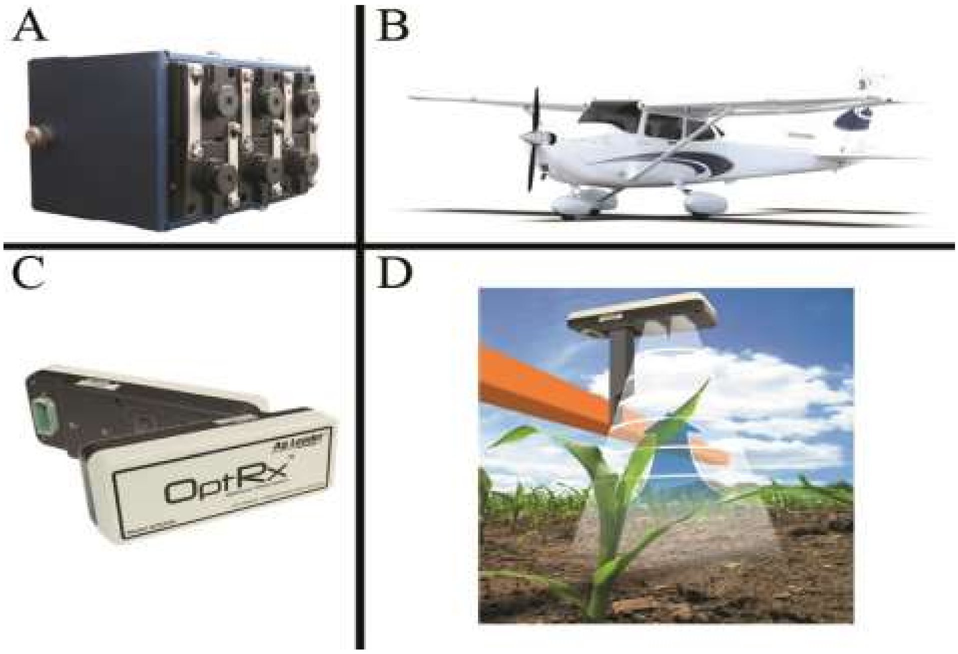


Fig. 4 A) Tetracam Mini MCA-6 multispectral camera. B) CESSNA 1725 Sky Hawk aeroplane. C) OptRx multispectral terrestrial sensor. D) OptRx sensor in operation in arable crops.

2.3. Scanning fruit trees using mobile terrestrial laser scanners (MTLS)

The mobile terrestrial laser scanner (MTLS) used to characterise geometric properties of tree canopy was a UTM30-LX-EW time-of-flight LiDAR (HOKUYO, Osaka, Japan). Apart from the LiDAR (Light Detection and Ranging), the device includes a gimbal (to correct the oscillations that MTLS can have during field operation) and a GNSS sensor. The passive receptor used to georeference the MTLS point cloud was a GNSS 1200+ (Leica Geosystems AG, Heerbrugg, Switzerland) with RTK-GNSS system (a real-time kinematics global navigation satellite system receiving GPS and GLONASS constellation signals). Both sensors were connected to a rugged laptop suitable to work in field conditions. Synchronization between GNSS and LiDAR was carried out using an own design program under Lab View environment (National Instruments, Austin, EE.UU). Concerning the operation of the LiDAR, the sensor allows the distance from the laser emitter to an object to be determined, by measuring the time elapsed since the transmitter emits the light beam until the response is received by the receptor after hitting the target. The sensor performs 40 scans per second (40Hz) with a 30 m scope. On the other hand, the HOKUYO sensor can work with up to 3 returns, but in the Thesis only the first return has been taken into account. This sensor emits a laser beam every 0.25° until completing an angle of 270°, 1081 laser beams every turn, existing a blind angle of 90° (Figure 5). The orientation of this angle is upward to lose as little information as possible. The sensor is mounted on a self-propelled platform (Figure 5). The platform advances at a constant speed of 4 km/h, allowing a distance between consecutive scans of approximately 2.7 cm to be obtained.

The MTLS was used to obtain geometrical properties of the tree canopy (Chapter 6). For functional characterisation of geometric features, it is essential to have georeferenced data in absolute coordinates, such as those provided by the Leica receiver. However, light beam impacts are only defined by polar coordinates (distance and emission angle) taking as reference the emitter (HOKUYO sensor). The combination of both positioning systems is necessary to georeference accurately where each beam has impacted. The result of having each impact point georeferenced is a point cloud, the basic structure to measure geometric properties of trees. Data process to obtain the geometric properties of the trees is explained in depth in chapter 6.

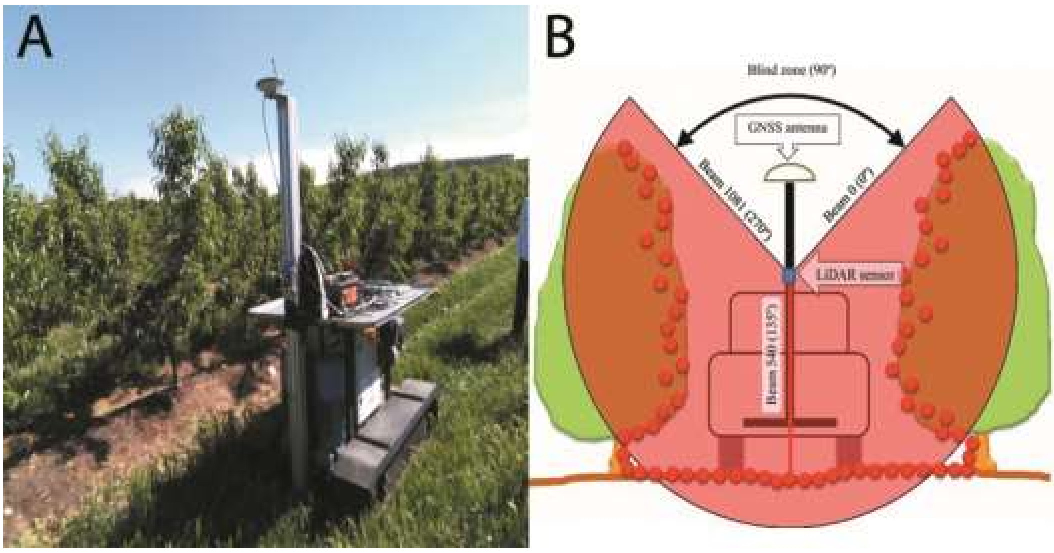


Fig 5 (A) Photography of the MTLS where its main elements are observed: Self-propelled platform, LiDAR and GPS antenna (author: A. Escolà). (B) MTLS diagram showing graphically how the LiDAR sensor works. Figure B has been obtained from Escolà et al. (2017).

3. STATISTICAL ANALYSIS

3.1. Geostatistics to assess spatial variability

3.1.1 The concept

Matheron developed modern geostatistics in the 1960s (Matheron, 1963) by expanding Krige's empirical ideas (Krige, 1951) as, for example, by exploring the concept that neighbouring points to a certain location in space could be used to improve predictions at that point. The theory of regionalised variables provides the appropriate mathematical tools (models) to make this prediction under the assumption of certain properties (Oliver, 2010). A regionalised variable is any spatially correlated attribute on some scale, that is, its value at a particular location depends, in statistical sense, on those of the neighbours, and this spatial dependence decreases as distance increases. This implies that unknown local values can be estimated more accurately from those actual values surrounding them. Interpolating in this way on a grid is the basis for creating a surface map or continuous raster map (Webster & Lark, 2013).

3.1.2 Variogram: the key model in geostatistics

To describe the variation of a regionalized variable in an area or surface, the idea behind the theory is that in locations close to each other, the values will be more similar than if we compare distant positions with each other. This idea is called spatial dependence or, what is the same, it is said that the variable shows spatial autocorrelation. The variogram is the model used in geostatistics to understand how regionalised variables are autocorrelated (Oliver, 2010).

The experimental variogram is calculated as follows:

$$\hat{\gamma}(h)=\frac{1}{2m(h)}\sum_{i=1}^{m(h)}[z(x_i)-z(x_i+h)]^2$$

where z(x_i) and z(x_i+h) are the actual values of Z at places (x_i) and (x_i+h), and m(h) is the number of paired comparisons at lag h. The experimental or sample variogram is obtained by changing h (Oliver, 2010). The quantity $\hat{\gamma}$ (h) is known as the semivariance at lag h. The 'semi' evidently refers to the fact that it is half of a variance (Webster & Oliver, 2007), and corresponds to the expected quadratic difference of the variable between two locations (sites).

Three especially relevant parameters are provided by the variogram once adjusted to the sampled data. The first is the sill (Figure 6). Assuming stationarity of the regionalized variable in each location (constant variance), the sill corresponds graphically to the value of the variogram that stabilizes at a certain distance or lag h. Specifically, this value is the constant variance, σ^2, assumed in the area under study (or variance of the random process). The range is the second parameter of the variogram (Figure 6), and it is directly associated with the sill.

The range is defined as the lag distance where the sill is obtained. This range is also known as the rank correlation, since it is the point where autocorrelation becomes 0 determining the limit of spatial dependence. Locations at greater distance than range are considered spatially independent. Finally, the third parameter of the variogram is the nugget variance (Figure 6). Usually, this parameter refers to measurement errors and/or variation that occur over distances shorter than the sampling interval. Finally, if the pattern of spatial variability does not happen in all directions homogeneously, the variogram will vary with the direction indicating anisotropy in the region under study (Webster & Oliver, 2007).

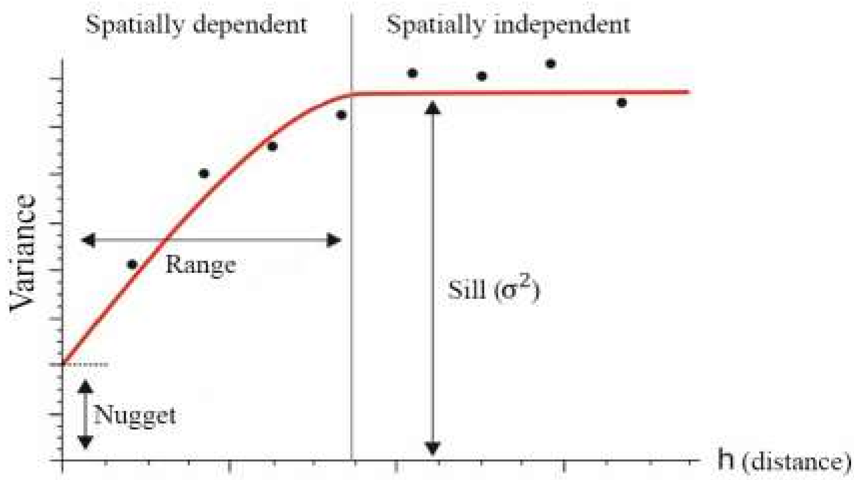


Fig. 6 Example of variogram (dots) fitted to a spherical model (red line) showing its components: nugget, range and sill. In the top, limit of spatial autocorrelation is shown (Figure adapted from Caramés, 2015).

Obviously, the experimental values of the variogram (dots in Figure 6) have to be adjusted to the best possible mathematical model to obtain the parameters mentioned above and minimize errors of subsequent analysis. This adjustment is a controversial step because some practitioners are fitting by eye. One advisable option could be the “weighted least squares” because it takes into account the accuracy of the individual semivariances providing the residual sum of squares a way to select the best fitting function (Webster & Oliver, 2007).

3.1.3 Kriging: the method to interpolate in geostatistics

Finally, a set of techniques allows values of the regionalised variables to be predicted in unsampled locations. Kriging, is the most remarkable technique since it obtains the best predictions (Laslett et al., 1987), combining non-bias and minimum error variance (best linear unbiased predictor, BLUP). Other methods only contribute with predictions while kriging contributes to forecasts, errors and kriging variances, which are a guide of the reliability of the estimates (Oliver, 2010). Many of the Thesis maps have been obtained by applying "ordinary kriging", the method commonly used in geostatistics applied to agriculture.

3.2. Map classification algorithm (ISODATA)

In chapters 3, 4 and 5 of the Thesis, homogeneous management zones were delimited using the ISODATA algorithm (Jensen, 1996). The prefix ISO of ISODATA is an abbreviation of Iterative Self-Organization way of performing classes or clusters within the analyzed maps. ISODATA algorithm is an iterative process to calculate minimum Euclidean distance when allocating each pixel to a cluster. The process begins allocating an arbitrary average value to each cluster (number of clusters are chosen previously because the method is unsupervised). In a second step, the algorithm groups the pixels according to the clusters, taking into account the smallest distance between average values of the clusters and those of the pixels. In a third step, ISODATA recalculates the mean of the new clusters with the pixels grouped in the first iteration. The same process is repeated, reassigning the pixels according to the mean values of the clusters. All this process is repeated until the number of pixels that migrate from one cluster to another will be low, that is, it does not exceed a certain threshold indicating that delimited clusters (zones or areas, since they are maps) are stable. Usually, if the number of clusters increases, a greater number of iterations will be necessary to obtain a stable classification. Figure 7 shows an example of how ISODATA algorithm works. The algorithm starts to allocate pixels from the continuous map (ECa, in this case) to the two previously established clusters. Continuing with this process, stabilisation of the clusters is observed as the number of iterations increases.

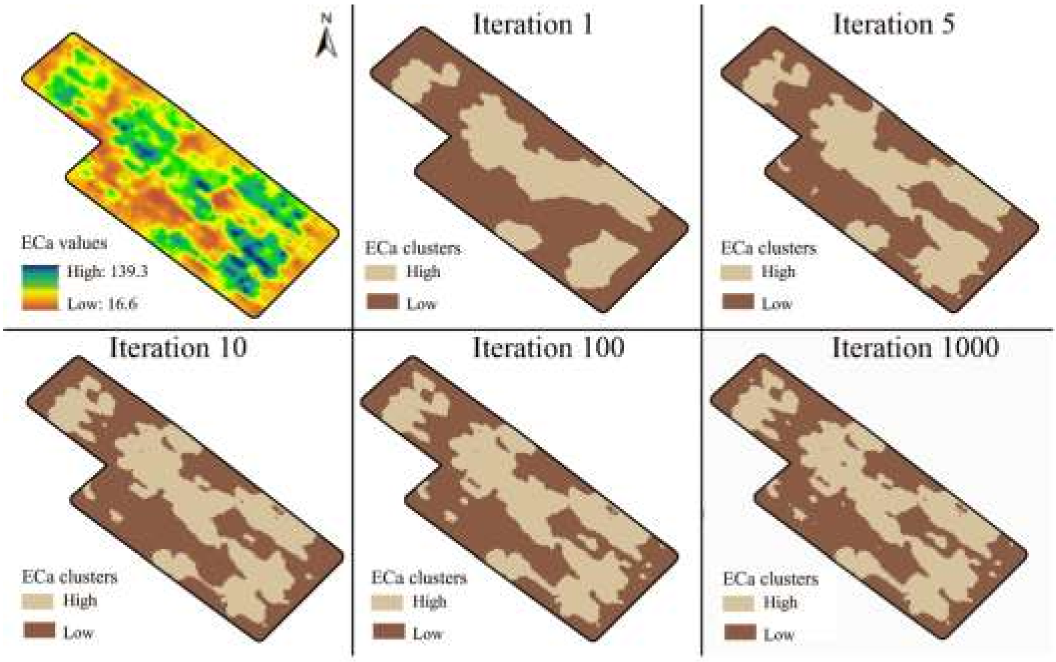


Fig. 7 Iterative process used by the ISODATA algorithm to obtain differentiated management zones.

As additional features of the algorithm, ISODATA allows the minimum number of pixels per zone to be defined. In this way, if this number is not reached, the algorithm will not create a new area (cluster). In addition, a zone is divided when the standard deviation within the original zone exceeds a predefined value, and the number of pixels is at least twice the minimum established above.

3.3. Multivariate analysis of variance (MANOVA)

In statistics, multivariate analysis of variance (MANOVA) is an extension of analysis of variance (ANOVA) to the multivariate case. As a multivariate procedure, it is used when there are two or more dependent variables that can´t be combined easily. The advantages of using MANOVA instead of several ANOVA for each of the dependent variables have been detailed by Warne, (2014). In chapters 3 and 4, differences between management zones delimited according to different levels of the NDVI and ECa auxiliary variables have been measured by this technique. The post-hoc, Descriptive Discriminant Analysis (DDA), is then used to see the contribution, for example, of each soil property to ECa signal variability. A detailed description of DDA after performing a MANOVA can be found in Thomas, (1992).

4. SAMPLING STRATEGIES USED DURING THE THESIS

Next, the sampling methods used in the Thesis are briefly described. Among them, stratified sampling and ranked set sampling have been evaluated as alternative schemes for use in fruit growing. For this reason, stratified sampling and ranked set sampling methods have been described in depth in chapters 5 and 6, respectively.

4.1. Simple Random Sampling (SRS)

This is the simplest sampling method. Every unit in the sample is chosen independently to any other, and all units have the same chance to be selected (Webster & Oliver, 2007). SRS is an unbiased surveying technique. SRS without replacement was the strategy used in the Thesis, but it is theoretically inefficient (imprecise) compared to other methods (Wulfsohn, 2010).

4.2. Systematic Sampling (SS)

In this sampling technique, the first item is taken at random, and the following samples are taken always at the same distance. For instance, if the sampling distance is 100 m and the first individual has been obtained at 60 m from the plot boundary, the following individuals will be taken at 160, 260, etc. This technique often offers an appropriate balance between precision of estimates and time spent to obtain samples. It is easy to implement without mistakes and, in many situations, it is more precise than SRS although biased (Cochran, 1977).

4.3. Stratified Random Sampling (StRS)

Stratified random sampling provides a technique to reduce the variance of the estimates (or improve efficiency). The population is partitioned into no overlapping strata, and each stratum is treated as a separate population for sampling purposes. Subsamples are selected independently at random within each stratum (Wulfsohn, 2010).

4.4. Ranked Set Sampling (RSS)

Ranked set sampling is a sampling technique that does not need to divide the plot into different strata to cover globally the variability of the variable of interest. Instead, the entire distribution of the variable to be sampled is approximated by means of a ranking mechanism. Specifically, RSS obtains the information (items) from different partial distributions (as many as items in the sample) that have been ranked to occupy the complete distribution of the population. To select an item in each range (partial distribution), a ranking mechanism is commonly carried out through the use of auxiliary information. If the variable to be sampled is expensive, it may be reasonable to use an auxiliary variable that is easily measurable and strongly correlated. This makes RSS at least as efficient as SRS.

References

Caramés, C.R. 2015. Assessment of the Spatial Variability of Vegetative Status in Vineyards Using Non-Destructive Sensors. Application of Remote and Proximal Sensing Technologies in Precision Viticulture. *Thesis*.

Cochran, W.G. 1977. Sampling Techniques, 3rd ed. New York: Wiley

Corwin, D.L. & Lesch, S.M. 2003. Application of soil electrical conductivity to precision agriculture: theory, principles and quidelines. *Agron. J*. 95, 455-471

Escolà, A., Martínez-Casasnovas, J.A., Rufat, J., Arnó, J., Arbonés, A., Sebé, F., Pascual, M., Gregorio, E., Rosell-Polo, J.R. 2017. Mobile terrestrial laser scanner applications in precision fruticulture/horticulture and tools to extract information from canopy point clouds. *Prec. Agric*. 18, 111–132.

Gitelson, A., Merzlyak, M.N. 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves *J. Photochem. Photobiol.* 22, 247-252.

Hall, A., Lamb, D.W., Holzapfel, B., Louis, J. 2002. Optical remote sensing applications in viticulture-a review. *Aust. J. Grape Wine R.* 8, 36-47.

Huete, A., Didan, K., Miura., T., Rodriquez., E.P., Gao., X., Ferreria., L.G. 2000. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.

Jensen, J.R. 1996. Introductory Digital Image Processing: Remote Sensing Perspective. 2nd ed. Prentice-Hall, Englewood Cliffs, NJ, USA.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand.

Laslett, G.M., McBratney, A.B., Pahl, P.J., Hutchinson, M.F. 1987. Comparison of several spatial prediction methods for soil pH. *Jour. Soil Sci*. 38, 325–341.

Matheron, G. 1963. Principles of geostatistics. *Econ. Geoly.* 58, 1246–1266.

Milsom, J. 1989. Field Geophysics. Open University Press, Milton Keynes and Halsted Press, John Wiley & Sons, New York.

Oliver, M.A. 2010. Geostatistical Applications for Precision Agriculture, Ed. M.A. Oliver (Springer Netherlands) pp. 1-34.

Rouse Jr., J.W., Haas, R.H., Deering, D.W., Schell, J.A., Harlan, J.C. 1974. Monitoring the Vernal Advancement and Retrogradation (GreenWave Effect) of Natural Vegetation. Greenbelt, MD, USA, NASA/GSFC Type III Final Report. 371p.

Thomas, D.R. 1992. Interpreting discriminant functions. A data analytic approach. *Multivar. Behav. Res.* 27, 335–362.

Veris technologies (05/12/18). www.veristech.com/the-sensors/v3100.

Warne, R.T., 2014. A primer on multivariate analysis of variance (MANOVA) for beha- vioral scientists. *Pract. Assessment Res. Eval.* 19, 17.

Webster, R.,M. Lark. 2013. Field sampling for environmental science and management. (Routledge: London & New York).

Webster, R., Oliver, M.A. 2007. Geostatistics for environmental scientists. Chichester, UK:Wiley

Wulfsohn, D. 2010. Sampling Techniques for Plants and Soil. Landbauforschung Volkenrode, (Special Issue).

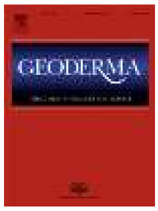
Chapter 3

ECa and multivariate analysis of soil properties

This chapter contains the following published paper in the journal *Geoderma.* JCR-SCI Impact Factor: 3.740. Category: Soil Science; 1st Quartile

Uribeetxebarria, A., Arnó, J., Escolà, A., Martínez-Casasnovas, J.A. 2018. Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. *Geoderma* 319, 185–193.

Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling



Asier Uribeetxebarria, Jaume Arnó, Alex Escolà, José Antonio Martínez-Casasnovas

ABSTRACT

Fruit production is relevant to the European agricultural sector. However, orchards in semi-arid areas of southern Europe may contain soils with constrains for tree development. This is the case ofsoils with high $CaCO_3$ content or limiting layers at variable depth. To assess spatial and in-depth variation of these soil constraints, an apparent electrical conductivity (ECa) survey was conducted in an orchard by using a galvanic contact soil sensor (Veris 3100). Different soil properties were randomly sampled at two depths (topsoil and subsoil) in 20 different sampling points within the plot. ECa raster maps were obtained for shallow (0-30 cm) and deep (0-90 cm) soil profile depths. In addition, an inversion modelling software was used to obtain horizontal ECa slices corresponding to 10 cm thick soil layers from 0-10 cm to 80-90 cm in depth. Concordance analysis of ECa slices allowed the soil profile to be segmented into four homogeneous horizons with different spatial conductivity pattern. Then, a multivariate analysis of variance (MANOVA) was key, i) to better interpret the specific soil properties that mainly contributed to the spatial variation of ECa ($CaCO_3$ and organic mater (OM) contents), and ii) to delimit the soil layer and the specific spatial pattern of ECa that allows potential management areas to be delineated by presenting the same trend in $CaCO_3$ and OM for topsoil and subsoil simultaneously. Moreover, assessing 3D variation of ECa made it possible to identify different soil areas that, linked to previous earthworks to optimize the parcelling of the farm, are the main cause of spatial variability within the orchard.

KEY WORDS

Soil sensing, ECa inversion, MANOVA, Precision fruticulture, Spatial analysis

1. INTRODUCTION

Fruit production and quality are affected to some extent by soil properties given the plant-soil interaction (Pedrera-Parrilla et al., 2014; Unamunzaga et al., 2014; Khan et al., 2016). As soil can vary spatially and at different scales, knowledge of spatial patterns within the plots could help farmers to make better management decisions based on the delimitation of areas with different soil conditions and agronomic needs (Ping et al., 2005; Vitharana et al., 2008; Pedrera-Parrilla et al., 2014; Córdoba et al., 2016). This is particularly relevant in semi-arid fruit growing areas of southern Europe. Soils in these areas are characterized by a high and spatially variable content of carbonates with clear incidence in nutritional deficiencies and chlorosis that affect growth and the normal foliar development. Accordingly, orchards usually show spatial variability in the canopy volume within the plot. In addition, this lack of homogeneity is particularly remarkable in plots that have been affected by successive earthworks over the years to reshape and optimize the parcelling of the farm. Fruit growers are therefore especially interested in locating and delimiting areas within the orchards that can be a major constraint for management (Fulton et al., 2011).

Soil sensors for mapping the apparent soil electrical conductivity (ECa in mS/m) are increasingly used to understand and evaluate how soil varies spatially (Corwin and Lesch, 2003; Abdu et al., 2008; Fulton et al., 2011) to delineate ECa-based management zones (Moral et al., 2010; Peralta and Costa, 2013). At present, it begins to be applied as a key sensing system in the framework of precision fruticulture (Käthner and Zude-Sasse, 2015). As ECa varies on a similar spatial scale as many soil physico-chemical properties (Sudduth et al., 2003; Carroll and Oliver, 2005), these soil monitoring systems have been widely accepted. Specifically, good correlations with soil salinity, soil water content and soil texture have been widely documented (Corwin & Lesch, 2005; Heil & Schmidhalter, 2012). Even, other soil properties affecting conductivity may be the organic C (Sudduth et al., 2003; Martinez et al., 2009), the cation exchange capacity (Sudduth et al., 2005) and the $CaCO_3$ content (Kühn et al., 2009). However, despite these good predictive characteristics, there are few studies that refer the use of such sensors in horticulture and, more specifically, in fruit orchards located in Mediterranean latitudes. One reason could be the small size of many fruit orchards. This induces farmers to think that tree plantations are rather homogeneous, and spatial variability is not enough to justify investing in this technology. By contrast, Käthner and Zude-Sasse (2015) show that even in small orchards there may be differences in soil properties that relate to tree growth and fruit size. Two soil sensing systems are commonly used in agriculture (Corwin and Lesch, 2005). In both cases (galvanic contact with the soil and electromagnetic induction), sensors measure the ECa on a soil volume basis including both topsoil and subsoil. This is very interesting since soil influences fruit trees at least to the depth covered by the roots, and ECa measurements should cover the same depth. Depending on the system, soil sensors provide with several electrical signals corresponding to several explored depths. When two signals are provided, they are known as shallow and deep ECa, and may correspond to the topsoil and whole profile depending on the sensor range. Farmers can get maps of both signals to evaluate the spatial variation of ECa, and indirectly the spatial pattern of soil related properties. Moreover, by overlapping maps they can also assess whether the soil is uniform or varies in depth. The problem occurs when the interest is to determine exact depths at which changes in the soil profile are produced (e.g. petrocalcic horizons) using such averaging procedures that encompass all or part of the soil profile (Heege, 2013).

Mapping the thickness or depth to a contrasting textural layer using ECa has been also referenced in several studies to detect clay horizons (Doolittle et al., 1994; Saey et al., 2009), or estimate topsoil depth explored by roots (Khan et al., 2016). Depth estimates may be based on empirical equations (using a single ECa signal that integrates the relative contribution of soil at each depth) or by combining data from multiple ECa sensors in both two- and three-layer models (Sudduth et al., 2010, 2013). Electromagnetic conductivity imaging (EMCI) of soil is another option that has been recently investigated (Triantafilis et al., 2013). Combining conductivity data and an inversion modelling software, a two-dimensional model of the ECa can be generated to assess soil variation (i.e. horizons) at discrete depth intervals (Triantafilis and Monteiro Santos, 2013). Researchers can take advantage of this additional information regarding the signal variation probably caused by layers of different thickness and composition. In short, soil properties sampled at varying depths may be better interpreted if a model indicating the variation of ECa with soil depth is available.

The main objective of the present research was to analyse the capacity of a galvanic contact soil sensor (Veris 3100) to be used as a diagnostic tool in fruit growing areas with high calcium carbonate content, and plots affected by previous parcelling works. Special attention was devoted to assess the spatial variability of physico-chemical soil properties to properly define differential management zones within an orchard. For that, we focused our research on i) evaluating the sensing system and its signal mapping, ii) inverting the ECa signal to obtain electrical imaging of ECa variation with soil profile, and iii) applying not conventional statistical methods i.e. multivariate analysis of variance (MANOVA) for a better interpretation of ECa and soil data.

2. MATERIAL & METHODS

2.1. Study area

The study was carried out at the IRTA Experimental Station (Lon. 0.392017, Lat. 41.654413, Datum WGS84), located in Gimenells, 24 km west from Lleida (Catalonia, Spain). The research was focused on a 0.65 ha plot that was planted in 2011 with peach trees (*Prunus persica* L. Stokes var. platycarpa) according to a 5 x 2.80 m plantation pattern (Fig. 1). Soil was classified as Petrocalcic Calcixerept (Soil Survey Staff, 2014), and it is a well-drained soil without salinity problems. The climate, typical of semi-arid areas of the Mediterranean region, is characterized by strong seasonal temperature variations (cold winters and hot summers), and an annual precipitation that is usually below 400 mm, although with significant interannual variability.

However, the most important feature of the plot was the presence of a petrocalcic horizon at a variable depth from 40 cm to 80 cm. This spatial variation in depth could be explained by the successive earthworks made in recent years in order to improve or adapt the parcelling of the farm. Probably, the petrocalcic layer was broken over time due to soil tillage and now appears even at shallow depths in certain areas. In fact, the history of transformation and land uses of this plot has been relatively complex as shown in Figure 1. Since 1946, when the Experimental Station was created, the plot has been cultivated with different crops and was modified in shape and size in several occasions (at least, the plot undergone a minimum of four major transformations in recent years, Fig. 1).

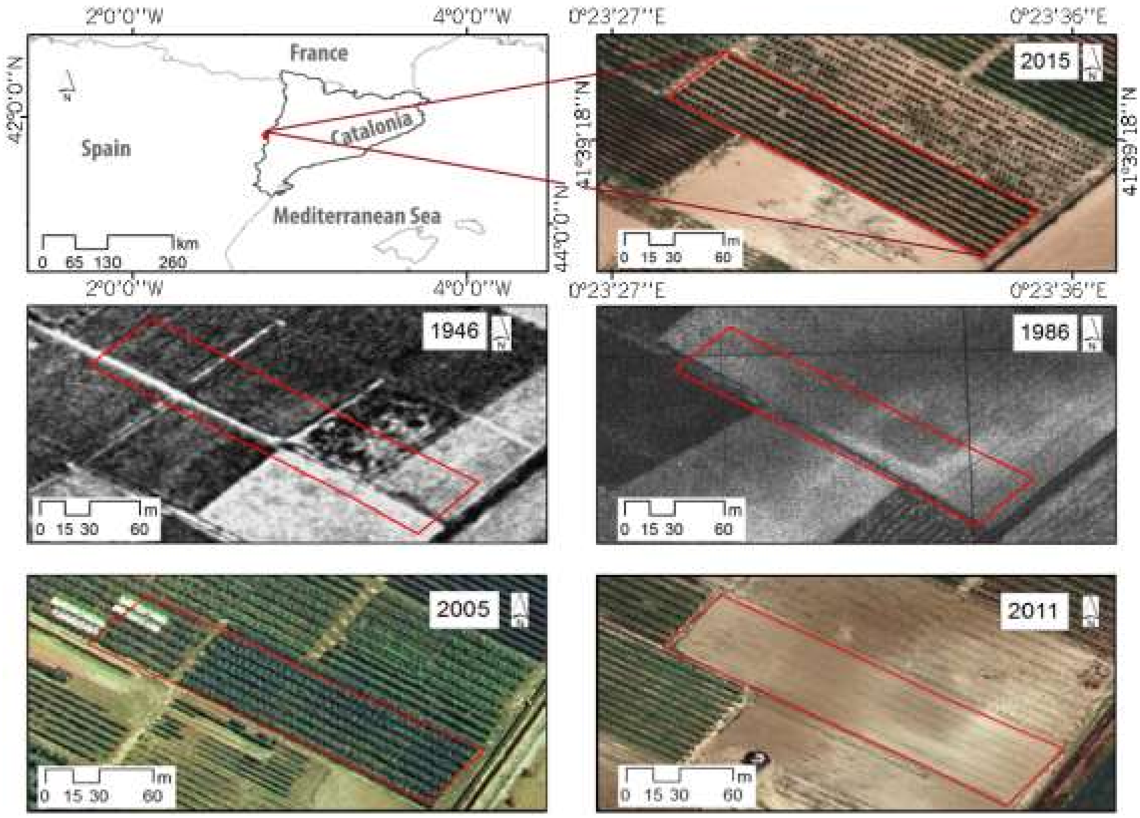


Fig. 1 Location of the study area and recent orthophoto (2015) of the experimental peach orchard (top). Other pictures: ancient orthophotos of the same area corresponding to different years since 1946.

2.2. Soil sampling

A simple random soil sampling was carried out in 20 different points within the plot (Fig. 2). Soil was sampled on March 15th, 2015. Samples were collected with the aid of a manual auger-hole at three different depths (0-30, 30-60, 60-90 cm). It is necessary to clarify that only in 4 of these sampling points it was possible to take a sample of the deepest layer, since the soil was shallow at most of the sampled sites. Sample locations were also georeferenced with submetric precision using a Trimble GPS Geo XH receiver and SBAS differential correction based on EGNOS. Soil samples were air-dried and sieved, and different physicochemical properties were analysed in laboratory according to the standard procedures. Specifically, data were obtained on the following properties: calcium carbonate content ($CaCO_3$), cation exchange capacity (CEC), electrical conductivity in a 1:5 soil-water solution ($EC_{1:5}$), organic matter (OM), pH measured in a 1:2.5 soil-water ratio, soil texture, total nitrogen (TN) in soil, and water holding capacity (WHC).

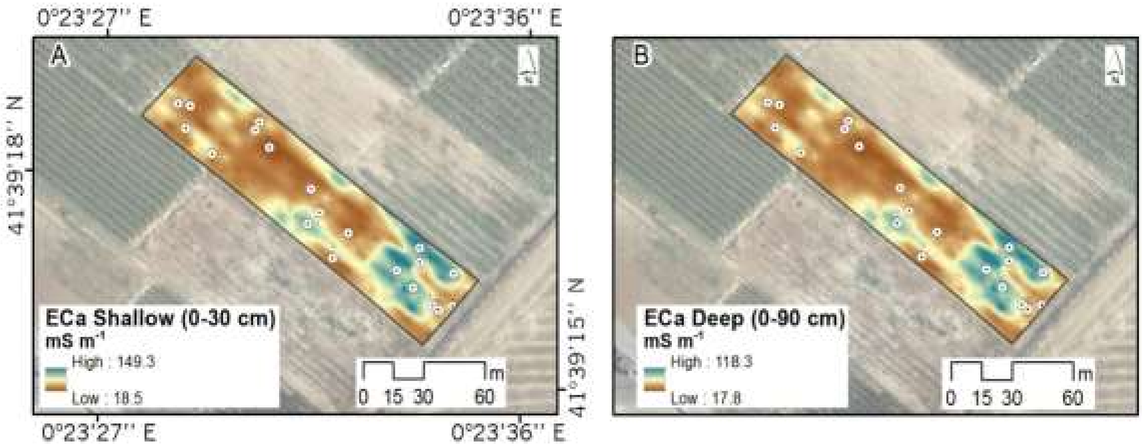


Fig. 2 ECa maps obtained by spatial interpolation and random soil sampling points within the plot. A) shallow ECa (0-30 cm). B) deep ECa (0-90 cm).

In addition to manual soil sampling, an ECa survey was conducted by using the Soil EC Surveyor Veris 3100 (Veris Technologies, Inc., Salina, KS, USA). The Veris 3100 implement is a simple and effective tool to acquire on-the-go information on soil bulk electrical conductivity for subsequent mapping. Its advantage lies in using two electrical arrays that allows ECa readings to be obtained at two different soil depths simultaneously and free of metal interference. Equipped with six heavy-duty coulter-electrodes, a pair of electrodes injects electrical current into the soil while the other two pairs measure the voltage drop. The penetration of the electrical current into the soil and, by extension, the volume of soil explored increases as the inter-electrode spacing increases. In our case, the array configuration allowed 0-30 cm (shallow ECa lecture) and 0-90 cm (deep ECa lecture) soil depths to be explored.

The ECa survey was carried out on March 23rd, 2015. For that, the Veris 3100 system was pulled by a 4-wheel drive vehicle passing along all the alleyways of the peach orchard. As tree rows were spaced 5 m, parallel ECa measurements were spaced this same distance. On the other hand, the soil sensor was connected to a Trimble AgGPS332 receiver for georeferencing purposes, and SBAS differential correction based on EGNOS was used. Regarding the spatial sampling resolution, data were recorded every second providing a total of 644 georeferenced ECa readings within the orchard (990 sampling points per hectare).

2.3. Apparent electrical conductivity maps and quasi-3D inversion modelling

Both ECa values (shallow and deep) were mapped by ordinary kriging. Maps were obtained after checking the normality of the acquired data and having removed extreme outliers from the analysis. Regarding the latter, ECa values lower than $Q_1 - 3 \times IQR$ or greater than $Q_3 + 3 \times IQR$ were not considered in the spatial interpolation (Q_1 and Q_3 were the first and third quartiles, and IQR was the interquartile range of the distribution). ArcMap 10.4.1 Geostatistical Analyst (Environmental Systems Research Institute, Redlands, CA, USA) was then used to finally interpolate shallow and deep ECa values by kriging on a 1 m grid. Geometric anisotropy along peach rows was taken into account when adjusting the experimental variograms (exponential model for the shallow

ECa, and stable model for the deep ECa). A strong spatial variation was also found on both maps (Fig. 2) (Cambardella et al., 1994).

In addition to raster ECa maps (0-30 cm and 0-90 cm depth profile), inversion modelling was applied in order to estimate ECa values at different specific soil depths. The software invVERIS 1.1 (EMTOMO Lda, Lisbon, Portugal) was used for this purpose. Specifically, invVERIS enables the inversion of ECa data acquired by galvanic contact soil sensors such as Veris 3100. The inversion procedure consists in a 1-dimensional laterally constrained technique to evaluate the ECa for a given soil transects (Quasi-2D inversion). The technique of signal inversion is based on a nonlinear, smoothness-constrained algorithm described by Monteiro Santos et al. (2011) and Monteiro Santos (2004). As the Veris 3100 sensor was used passing along all the alleyways of the orchard (different transects), the possibility of inversion in each of these profiles makes it possible to obtain a soil layer model from the set of 1D models distributed according to the locations of the ECa measurement sites. The program finally allows horizontal slices (maps) of soil layers of the same thickness to be displayed at selected depths and covering the whole area of the plot (quasi-3D inversion modelling). In our case, we chose to model and visualize 9 layers of ECa of 10 cm in thickness from 0-10 cm to 80-90 cm in depth.

2.4. Data analysis

2.4.1. Clustering and map comparison

A cluster analysis was performed to classify ECa maps. Once the shallow and deep ECa maps were created, each map was clustered into two classes (low and high EC_a) using the Iterative Self-Organizing Data Analysis Technique (ISODATA) implemented in ArcGIS 10.4.1 (IsoCluster function). The procedure is based on an iterative algorithm that begins assigning an arbitrary mean to each class. Pixels are then successively reassigned based on minimizing the Euclidean distance of each pixel to the mean value of the class. In each iteration, class means are recalculated and pixels are reallocated until the last iteration is reached, or the number of pixels that change from one class to another does not exceed a certain threshold (Guastaferro et al., 2010). Classified conductivity maps were then used to assess whether the soil was significantly different depending on the ECa in each area. Multivariate analysis of variance (MANOVA) of the sampled soil properties according to conductivity classes (high and low) was used to evaluate this effect.

This same procedure was repeated for the horizontal slices (maps) resulting from the quasi-3D inversion modelling of ECa values. However, to avoid redundant analysis, maps from the 9 inverted ECa layers were first compared with each other using the Map Comparison Kit (MCK) software (Visser & de Nijs, 2006). The degree of similarity between maps was quantified by the Kappa coefficient (Cohen, 1960) and, as a result of the comparison, the nine layers previously established were finally grouped into four different homogeneous horizons.

2.4.2. Multivariate analysis of variance (MANOVA)

Separate analysis of each sampled soil property according to different levels of ECa (ANOVA) may lead to misleading and inconsistent results. In fact, ECa reflects the combined effect of soil properties as a whole, and delimitation of areas within the plot based on ECa maps should be checked from a multivariate approach. To detect the specific soil properties that mainly contributed to the spatial variation of EC_a, a multivariate analysis of variance (MANOVA) was performed. The methodis slightly more complex and scarcerly used in soil science (Taylor and Whelan, 2011). However, it has proved to be an effective technique to delineate differential management zones in precision agriculture (Ping et al., 2005). In our case, the effectof ECa was then evaluatedby performinga MANOVAusingsoil sampled propertiesas dependentvariablesand classesofEC_a (high and low) as the factorunder analysis.

The problem arises whena significant result must be interpreted, since there is no unanimity as to the most appropriate post hoc procedures to be used (Warne, 2014). In this research, a descriptive discriminant analysis (DDA) was used to interpret significant MANOVAs (Thomas, 1992). DDA is a statistical procedure that, in our case, provided a linear combination of the soil properties (discriminant function) that managed to separate the two classes of ECa in a meaningful way. Standardized coefficients of the discriminant function and structure coefficients were used for interpretation. Standardized discriminant function coefficients (SDFCs) were indicative ofthe contributionof each soil variable tothe discriminant function, whereas thestructure coefficients (SCs) were the correlations between each observed variable and the discriminant function scores. The most important soil variables affecting the differential ECa were finally identified through the so-calledparallel discriminant ratio coefficients (parallel DRCs) by multiplying SDFCs by the corresponding SCs. So parallel DRCs were used to assess non-redundant soil variables contributing to discriminate two types of soil in terms of ECa.

3. RESULTS & DISCUSION

3.1. Soil characterization

Table 1 shows the main descriptive statistics for the soil properties. Only soil properties analysed at both depths (0-30 cm and 30-60 cm) were included in the analysis (total nitrogen was excluded for this reason). Soils in the study plot were found to have an average depth of about 60 cm, basically limited by the petrocalcic horizon. As the standard deviation was 18 cm, soil depth showed a considerable spatial variability within the plot (CV of 30 %).Other soil properties that showed spatial variability were the electrical conductivity at the two sampling depths and, with much lower incidence, the water holding capacity at the deepest layer. Regarding the latter, average WHC did not vary significantly between soil and subsoil, and a rather low value of 64.23 mm was obtained as an average for the whole soil profile. Carbonates also varied spatially (CV of 20 %), but the most significant was the high value of the carbonates content in the soil (27 % at the top layer and 33 % at the bottom layer). Probably, the observed enrichment of $CaCO_3$ in the second layer (6 % higher) could be explained by the near presence of the petrocalcic horizon and its breakage over the years by tillage operations. Derived from this, a moderately basic pH was expected in the soil. Finally, the organic matter was lower in the subsoil but at the expense of a strong spatial variation. The soil could be considered as well-drained, not saline, and with a loam soil texture. No other consideration was noteworthy.

Table 1 Soil properties for two sampling depths (N=20 sampling points)

Soil property	Mean	Standard deviation	CV (%)	Minimum	Maximum
Soil depth* (cm)	59.75	17.91	29.99	42.0	90.0
Sampling depth 0- 30 cm					
pH	7.92	0.10	1.26	7.90	8.10
$EC_{1:5}$ (dS/m)	0.93	0.51	54.83	0.23	1.84
$CaCO_3$ (%)	26.95	5.61	20.81	16.27	35.32
CEC (meq/100 g)	13.97	1.02	7.30	11.50	15.90
OM (%)	2.70	0.64	23.70	0.89	3.93
Sand (%)	42.45	3.66	8.62	37.60	54.40
Silt (%)	30.38	3.38	11.12	18.30	35.60
Clay (%)	27.18	1.99	7.32	23.80	31.50
WHC (%)	10.95	0.94	8.58	9.00	13.00
Sampling depth 30- 60 cm					
pH	7.57	0.12	1.58	7.30	7.80
$EC_{1:5}$ (dS/m)	1.92	0.96	50.00	0.50	3.46
$CaCO_3$ (%)	33.04	6.00	18.15	19.29	41.71
CEC (meq/100 g)	11.61	1.47	12.66	8.96	14.10
OM (%)	1.22	0.62	50.81	0.16	3.24
Sand (%)	42.12	4.50	10.68	35.80	53.10
Silt (%)	31.22	6.22	19.92	14.20	39.90
Clay (%)	26.21	3.15	12.01	19.40	31.10
WHC (%)	10.55	1.63	15.44	8.00	14.00

*Soil depth refers to the depth needed to reach the petrocalcic horizon.

3.2. Soil horizons delimited by ECa patterns at different depths

Figure 3 shows the interpolated maps of ECa (shallow and deep) and the corresponding maps where ECa was classified into two classes (high ECa and low ECa). Comparing the shallow and deep ECa maps (Fig. 3), one realizes that the pattern of spatial variation is quite similar. In theory, this was indicative of a uniform soil in depth. However, classified maps are not so similar (Fig. 3), occupying the high conductivity class a larger area (59 % of the plot area) when the plot was classified based on the deep signal compared to 38 % for the case of shallow signal.

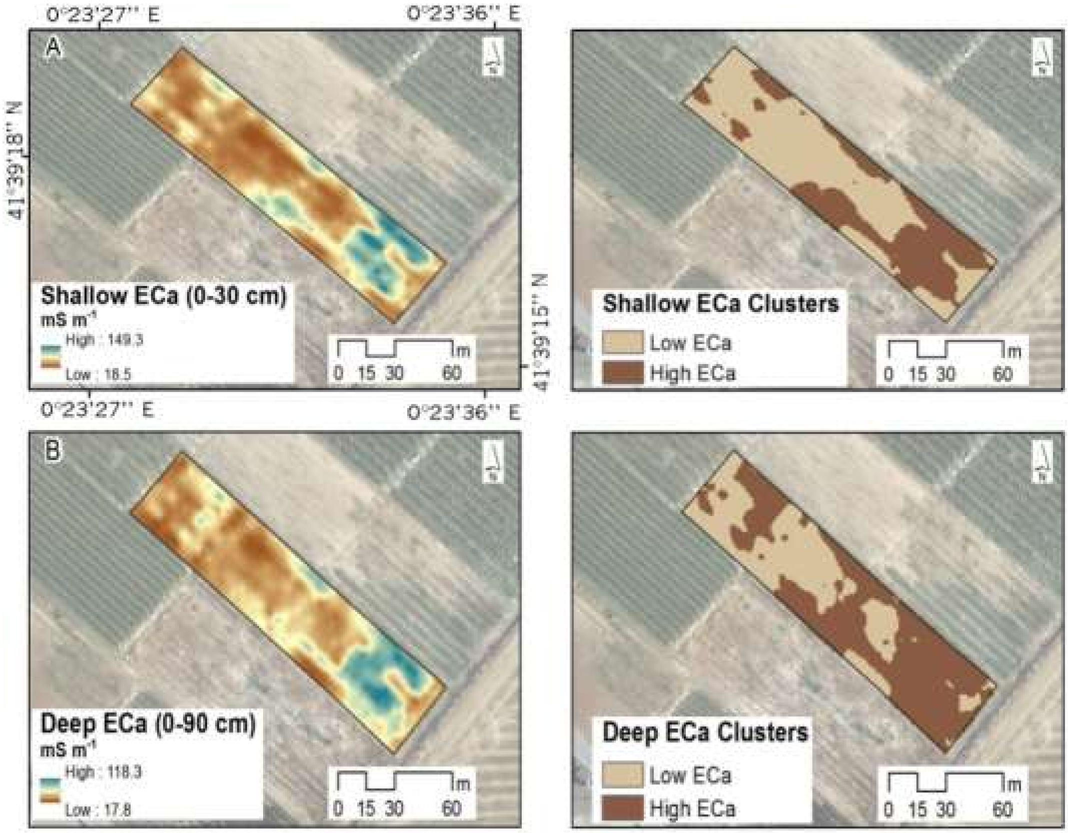


Fig. 3 A) Shallow interpolated ECa map (left), and shallow clustered map with low and high ECa classes (right). B) Deep interpolated ECa map (left), and deep clustered map with low and high ECa classes (right).

To assess in more detail the variation of ECa within depth, inversion modelling software invVERIS 1.1 was used to obtain electrical conductivity maps (ECM) or horizontal slices every 10 cm in depth. In this respect, nine different maps were obtained corresponding to depths from 0-10 cm to 80-90 cm (Fig. 4). Maps corresponding to the topsoil layers (0-30 cm in depth) showed higher ECa values and greater spatial variability than the deeper layers (CV of 45 % for layer 0-10 cm was reduced to CV of 15 % for layer 80-90 cm). The attenuation of the conductivity signal was therefore evident, hindering differentiation of soils although there was considerable spatial variation in certain soil properties at greater depth (Table 1). This same result was observed by Sudduth et al. (2005).

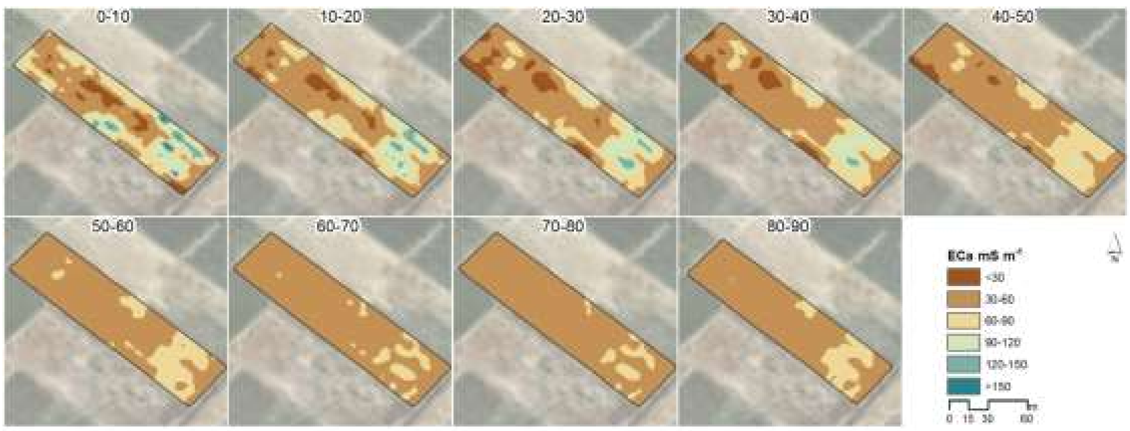


Fig. 4 Horizontal slices 10 cm thick at different depths obtained by inversion of the ECa values with invVERIS 1.1.

Concordance analysis between ECa maps allowed the horizontal layers of 10 cm to be grouped according to 4 soil horizons that were homogeneous but different from each other in both signal intensity and spatial pattern. Only layers with high spatial agreement were grouped (Kappa coefficient greater than 0.6, data not shown; Landis and Koch, 1977), resulting in a soil profile that could be segmented into (i) a first horizon occupying the first 10 cm (0 to 10 cm), (ii) a second horizon of the same thickness, from 10 to 20 cm), (iii) a third horizon with a greater thickness to a depth of 50 cm (20 to 50 cm), and (iv) a deeper and homogeneous layer up to 90 cm (50-90 cm). The representative conductivity map of each horizon was classified into two ECa classes (high and low) following the same procedure as for the original maps (Fig. 5). Multivariate analysis of variance (MANOVA) of soil properties was then performed for each of the identified soil horizons (Table 2).

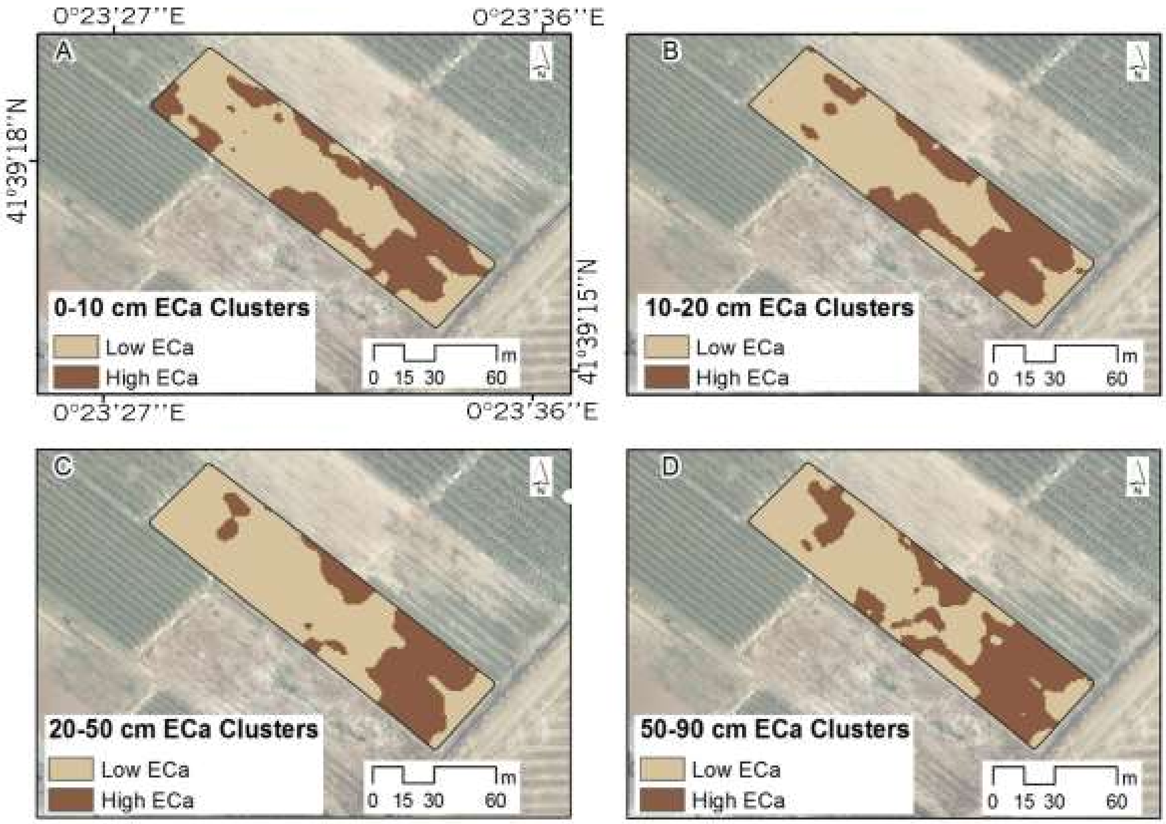


Fig. 5 Clustered maps (high and low ECa) for the four soil horizons delimited by concordance analysis.

3.3. Soil properties influencing the spatial and in-depth variation of the Eca

A series of multivariate analysis of variance (MANOVAs) were performed to determine specific soil properties mainly linked to the spatial variation of ECa measured with the Veris 3100 sensor. Results are shown in Table 2. For topsoil properties (0-30 cm), pH, $CaCO_3$ and organic matter (OM) were the properties that contributed most to the spatial variation of the ECa. This result was somehow unexpected since, besides carbonates, organic matter appeared as a soil property that influenced the ECa signal. Descriptive discriminant analysis (DDA) highlighted the importance of OM through the so-called parallel discriminant ratio coefficient (parallel DRC), that indicates the relative contribution of each soil property in the canonical (discriminant) function. Even more, the influence of OM on the ECa was evident for both the shallow values and for the discretized values for the first three soil horizons (Table 2). Among the latter, discriminant function of soil properties corresponding to the 10-20 cm soil layer was able to better differentiate low and high electrical conductivity. However, good discrimination between low and high signals was also achieved by using the soil layer conductivity corresponding to a depth of 20-50 cm. This soil classification was especially interesting given the strong contribution of both $CaCO_3$ and OM in the corresponding discriminant function (Table 2). On the other hand, contribution of water holding capacity (WHC) did not seem to significantly influence ECa, and the original idea that water holding capacity could be behind the spatial variation of ECa was no longer supported.

A similar trend occurred for the subsoil (30-60 cm) and, again, $CaCO_3$ and OM were the properties that contributed most to explain the variability of ECa. Discriminant function for the deeper layer (50-90 cm) provided the best differentiation between low and high electrical conductivity. But, as with topsoil, it was necessary to classify the ECa for a boundary horizon between topsoil and subsoil (20-50 cm) to detect such differences with almost exclusive contribution of $CaCO_3$ and OM (Table 2). Another possibility would be to focus the deep soil management on the carbonates content and, in this case, areas could be delimited using the deep ECa. The relationship between soil variables and ECa coincided with that reported by other authors. Thus, there was an increase in electric conductivity with increasing carbonates content (Kühn et al., 2009), while the effect of organic matter was just the opposite (Moral et al., 2010).

Spatial and in-depth variation of $CaCO_3$ and OM made it possible to propose a site-specific management within the plot based on applying chelates and organic amendments in a differentiated way. Two basic issues need to be addressed. Should the plot be managed based on the differences between topsoil and subsoil, or is it more advisable to consider the entire profile globally? And, faced with the delineation of potential management zones, should areas be defined using the shallow ECa, the deep ECa, or the discretized ECa for a particular soil layer? MANOVA provided very interesting information to assist in such decision making process (Table 2). First, differential management should primarily focus on the $CaCO_3$ spatial distribution because this property clearly influenced the ECa for the entire soil profile. The petrocalcic horizon would probably be behind this spatial variation as a result of previous parcelling and earthworks in recent years. Secondly, the delimitation of areas of low and high conductivity by respectively matching low and high $CaCO_3$ contents for both topsoil and subsoil would be ideal for differential management. The soil layer covering a depth between 20 and 50 cm has shown a spatial pattern of electrical conductivity that meets this requirement. OM was also important, and its spatial variation in the topsoil also seems to be linked to the variation in the subsoil in view of discriminant functions obtained for the soil layer from 20 to 50 cm depth (Table 2).

Table 2 Descriptive discriminant analysis (DDA) of soil properties affecting ECa for different soil depths

ECa classified map	**Discriminant analysis (DDA)**	***Soil properties sampled in the topsoil (0-30 cm)***							
		pH	**$EC_{1:5}$**	**$CaCO_3$**	**OM**	**CEC**	**Clay**	**Sand**	**WHC**
Shallow ECa	SDFC	1.02	0.26	-1.04	1.19	-0.32	0.45	1.08	-0.03
	SC	0.37	-0.23	-0.36	0.23	0.06	-0.09	0.08	-0.20
	Parallel DRC	**0.38**	-0.06	**0.37**	**0.27**	-0.02	-0.04	0.09	0.01
Depth 0-10 cm	SDFC	1.21	1.85	1.26	-1.60	0.35	-0.01	0.27	0.19
	SC	-0.35	0.50	0.22	-0.14	0.05	0.12	-0.13	0.16
	Parallel DRC	-0.43	**0.93**	**0.27**	**0.22**	0.02	0.00	-0.04	0.03
Depth 10-20 cm	SDFC	0.57	-0.66	-1.42	1.68	-0.28	0.74	1.60	-0.11
	SC	0.27	-0.22	-0.25	0.15	0.03	-0.07	0.09	-0.15
	Parallel DRC	0.15	0.15	**0.36**	**0.25**	-0.01	-0.05	0.14	0.02
Depth 20-50 cm	SDFC	1.21	1.36	-1.14	1.83	-1.77	-	-0.21	0.13
	SC	0.27	-0.12	-0.46	0.26	0.08	-	0.06	-0.19
	Parallel DRC	**0.33**	-0.16	**0.53**	**0.48**	-0.13	-	-0.01	-0.02
ECa classified map	Discriminant analysis (DDA)	*Soil properties sampled in the subsoil (30-60 cm)*							
		pH	$EC_{1:5}$	$CaCO_3$	OM	CEC	Clay	Sand	WHC
Deep ECa	SDFC	0.97	0.35	1.56	-0.02	-	-	-1.42	0.69
	SC	-0.26	0.35	0.42	-0.09	-	-	-0.21	0.26
	Parallel DRC	-0.26	0.12	**0.65**	0.00	-	-	**0.30**	0.18
Depth 20-50 cm	SDFC	1.19	-	2.20	-0.87	-	-	-	-
	SC	0.04	-	0.26	-0.45	-	-	-	-
	Parallel DRC	0.04	-	**0.58**	**0.39**	-	-	-	-
Depth 50-90 cm	SDFC	-	0.48	1.04	-0.48	0.66	-1.36	-0.75	0.58
	SC	-	0.37	0.38	-0.38	0.02	-0.05	-0.07	0.18
	Parallel DRC	-	**0.18**	**0.40**	**0.18**	0.01	0.07	0.06	0.10

Hyphens indicate variables that were removed to obtain significant discriminant functions. Parallel DRC in bold indicates soil properties with greater contribution in the discriminant function from MANOVA. SDFC: standardized discriminant function coefficient; SC: structure coefficient; parallel DRC: parallel discriminant ratio coefficient. $EC_{1:5}$: electrical conductivity in a 1:5 soil-water solution (dS/m); $CaCO_3$ (%); OM: organic matter content (%); CEC: cation exchange capacity (meq/100 g); Clay (%); Sand (%); WHC: water holding capacity (%).

Probably, the boundary condition between topsoil and subsoil of this intermediate layer allowed to use it as representative of the whole soil profile. Contrary to the post hoc interpretation of MANOVAs, separate ANOVAs for each of the most relevant soil properties led to somewhat different results (Table 3). Specifically, spatial pattern of ECa for this horizon (20 cm to 50 cm) would only be justified to delimit potential zones of topsoil management. However, this same ECa pattern can be applied to the subsoil according to the MANOVA results, highlighting the need for a multivariate approach when deciding on potential management areas.

Table 3 Relevant soil properties with significant differences according to ECa classes

ECa classified map	ECa classe	pH	$EC_{1:5}$	$CaCO_3$	OM
Soil properties sampled (0-30 cm)					
Shallow ECa	Low	7.98	-	25.09	-
	High	7.84	-	31.73	-
Depth 0-10 cm	Low	7.98	0.69	-	-
	High	7.83	1.52	-	-
Depth 10-20 cm	Low	7.98	0.70	25.09	-
	High	7.83	1.40	31.11	-
Depth 20-50 cm	Low	7.95	-	24.01	2.78
	High	7.85	-	31.87	2.22
Soil properties sampled (30-60 cm)					
Deep ECa	Low	-	-	29.28	-
	High	-	-	35.47	-
Depth 20-50 cm	Low	-	-	-	-
	High	-	-	-	-
Depth 50-90 cm	Low	-	1.29	29.05	1.57
	High	-	2.26	35.09	0.93

Hyphens indicate variables that did not show significant differences ($p<0.05$) in the corresponding ANOVAs. $EC_{1:5}$: electrical conductivity in a 1:5 soil-water solution (dS/m); $CaCO_3$ (%); OM: organic matter content (%). CEC, clay, sand and WHC are not shown in the table, because they were not significantly different in any of the soil layers.

3.4. Spatial pattern of the ECa as a result of previous parcelling: consequences for management

As previously mentioned, the plot under study had been subjected to different parcelling processes in recent decades. Figure 1 shows the evolution from 1946. By overlapping the ECa map derived from the Veris 3100 sensor readings (deep signal) (Fig. 6), it is interesting to observe how it clearly reproduces the edges and divisions of previous plots. Depending on the use and the parent material, the soil of the present plot is the result of all these transformations affecting productivity and causing the current spatial variability. From this point of view, the use of combined information from soil sensors and historical orthophotos becomes an interesting tool for better soil interpretation and better diagnosis of management actions to be performed.

As showed in Fig. 6, the two transverse subdivisions (paths) in 1946, one of which was still in place in 2005, are relatively marked as areas of highest conductivity in the ECa cluster map for the reference horizon corresponding to 20-50 cm depth (Fig. 5). On the other hand, a more compact area with also high ECa values appears as a consequence of incorporating a piece from another different plot in 1986. This area has remained different from the rest of the plot until today (perfectly marked on the cluster map), and corresponds to the area where problems commonly due to high carbonates contents are done. Our management proposal for this plot could then be established making use of the classified map for the aforementioned reference horizon that matches the joint differential needs for topsoil and subsoil. Therefore, in areas with potential chlorosis and soil fertility problems (high ECa), the farmer could implement a fertilization plan by adding chelates and organic fertilizers to correct these nutritional imbalances more optimally.

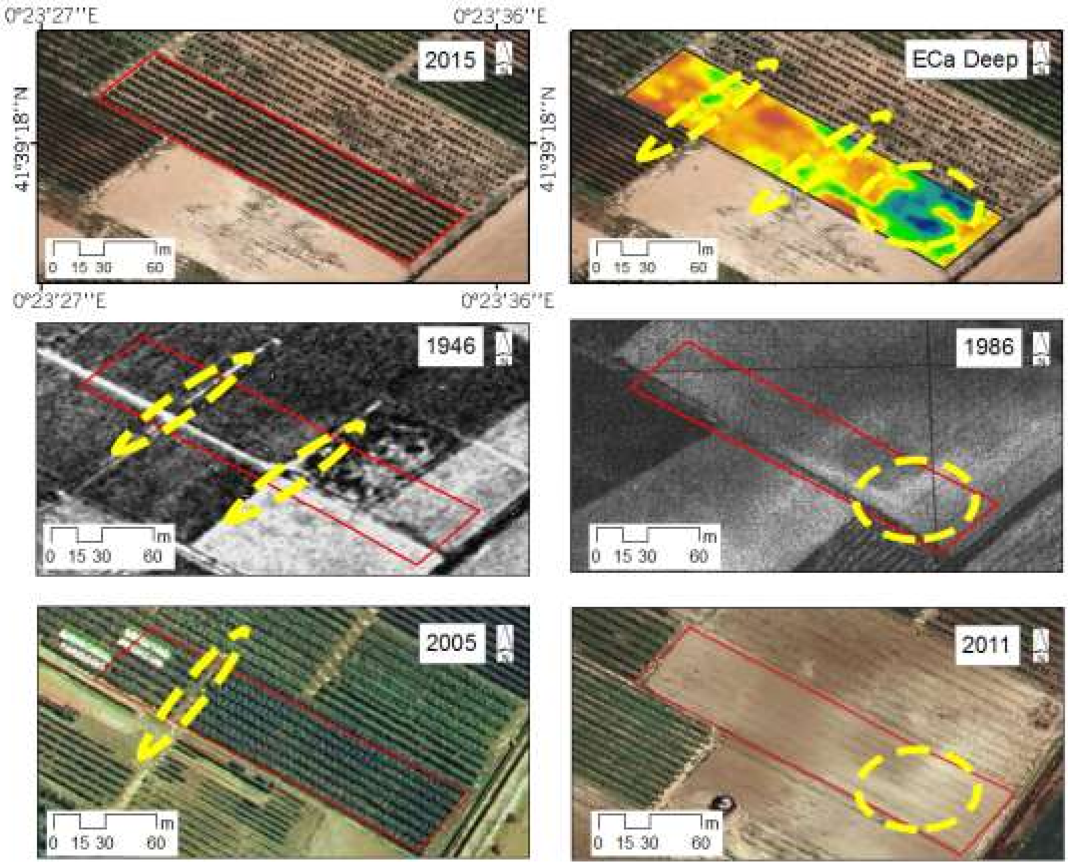


Fig. 6 Evolution of historical parcelling until 2015 and current design. The overlapping of the deep ECa map shows where the transformations occurred.

4. CONCLUSIONS

The acquisition and mapping of apparent electrical conductivity (ECa) allowed areas with potential chlorosis problems to be delimited in the study plot. These areas were characterized by high $CaCO_3$ content due to the presence of a petrocalcic horizon at variable depth. On the other hand, parcelling carried out over the years has been revealed as a key factor in understanding the soil spatial variability allowing ECa measures to be estimated in depth at discrete intervals makes it possible to divide the soil profile into homogeneous horizons by comparing classified maps of ECa at different depths. Then, multivariate analysis of variance (MANOVA) based on the previous maps offers interesting outputs in agronomy because, i) the overall relationship between ECa and soil properties is better interpreted, and ii) potential management zones can be validated knowing in detail the specific causes behind the variation of ECa, in our case, both $CaCO_3$ and organic matter contents.

Acknowledgements

This work was funded by the Spanish Ministry of Science and Innovation through the project AgVANCE (AGL2013-48297-C2-2-R). We are also grateful to the University of Lleida for funding PhD scholarship for the first author.

References

Abdu, H., Robinson, D.A., Seyfried, M., Jones, S.B., 2008. Geophysical imaging of watershed su bsurface patterns and prediction of soil texture and water holding capacity. *Water Resour. Res*., 44(4).

Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., Konopa, A.E. 1994., Field-scale variability of soil properties in central Iowa. *Soils Sci. Soc. Am. J*. 58, 1501-1511.

Carroll, Z.L., Oliver, M.A., 2005. Exploring the spatial relations between soil physical properties and apparent electrical conductivity. *Geoderma*. 128, 354-374.

Cohen, J. 1960., A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20 (1), 37–46.

Córdoba, M.A., Bruno, C.I., Costa, J.L., Peralta, N.R., Balzarini, M.G., 2016. Protocol for multivariate homogeneous zone delineation in precision agriculture. *Biosys. Eng*. 143, 95-107.

Corwin, D.L., Lesch, S.M. 2003., Application of Soil Electrical Conductivity to Precision Agriculture: Theory, Principles, and Guidelines. *Agron. J*., 95(3), 455-471.

Corwin, D.L., Lesch, S.M. 2005., Apparent soil electrical conductivity measurements in agriculture. *Comput. Electron. Agric.,* 46, 11-43.

Doolittle, J.A., Sudduth, K.A., Kitchen, N.R., Indorante, S.J. 1994.., Estimating depths to claypans using electromagnetic induction methods. *J. Soil Water Conserv*. 49(6), 572-575.

Fulton, A., Schwankl, L., Lynn, K., Lampinen, B., Edstrom, J., Prichard, T. 2011. Using EM and VERIS technology to assess land suitability for orchard and vineyard development. *Irrigation Sci*., 29, 497-512.

Guastaferro, F., Castrignanò, A., De Benedetto, D., Sollitto, D., Troccoli, A., Cafarelli, B., 2010. A comparison of different algorithms for the delineation of management zones. *Precis. Agric*. 11, 600-620.

Heege, H.J. (ed.)., 2013. Precision in Crop Farming: Site Specific Concepts and Sensing Methods: Applications and Results.

Heil, K., Schmidhalter, U., 2012. Characterisation of soil texture variability using the apparent soil electrical conductivity at a highly variable site. *Comput Geosci-UK*. 39, 98-110.

Käthner, J., Zude-Sasse, M., 2015. Interaction of 3D soil electrical conductivity and generative growth in *Prunus domestica* L. *Eur. J. Hortic. Sci*. 80(5), 231-239.

Khan, F.S., Zaman, Q.U., Chang, Y.K., Farooque, A.A., Schumann, A.W., Madani, A., 2016. Estimation of the rootzone depth above a gravel layer (in wild blueberry fields) using electromagnetic induction method. *Precis. Agric.,* 17, 155-167.

Kühn, J., Brenning, A., Wehrhan, M., Koszinski, S., Sommer, M., 2009. Interpretation of electrical conductivity patterns by soil properties and geological maps for precision agriculture. *Precis. Agric.,* 10, 490-507.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*. 33 (1), 159–174.

Martinez, G., Vanderlinden, K., Ordónez, R., Muriel, J.L., 2009. Can Apparent Electrical Conductivity Improve the Spatial Characterization of Soil Organic Carbon? *Vadose Zone J.* 8, 586-593.

Monteiro Santos, F.A., 2004. 1-D laterally constrained inversion of EM34 profiling data. *J. Appl. Geophys*. 56, 123-134.

Monteiro Santos, F.A., Triantafilis, J., Bruzgulis, K., 2011. A spatially constrained 1D inversion algorithm for quasi-3D conductivity imaging: Application to DUALEM-421 data collected in a reverine plain. *Geophysics*. 76(2), B43-B55.

Moral, F.J., Terrón, J.M., Marques da Silva, J.R., 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil Tillage Res*. 106, 335–343.

Pedrera-Parrilla, A., Martínez, G., Espejo-Pérez, A.J., Gómez, J.A, Giráldez, J.V., Vanderlinden, K., 2014. Mapping impaired olive tree development using electromagnetic induction surveys. *Plant. Soil.* 384, 381-400.

Peralta, N.R., Costa, J.L., 2013. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Comput. Electron. Agric.* 99, 218-226.

Ping, J.L., Green, C.J., Bronson, K.F., Zartman, R.E., Dobermann, A., 2005. Delineating potential management zones for cotton based on yields and soil properties. *Soil Science*. 170(5), 371-385.

Saey, T., Simpson, D., Vermeersch, H., Cockx, L., Van Meirvenne, M., 2009. Comparing the EM38DD and DUALEM-21S Sensors for Depth-to-Clay Mapping. *Soil Sci. Soc. Am. J.* 73, 7-12.

Soil Survey Staff., 2014. Keys to Soil Taxonomy, 12th ed. United States Department of Agriculture.

Sudduth, K.A., Kitchen, N.R., Bollero, G.A., Bullock, D.G., Wiebold, W.J., 2003. Comparison of Electromagnetic Induction and Direct Sensing of Soil Electrical Conductivity. *Agron. J.* 95, 472-482.

Sudduth, K.A., Kitchen, N.R., Myers, D.B., Drummond, S.T., 2010. Mapping Depth to Argillic Soil Horizons Using Apparent Electrical Conductivity. *J. Environ. Eng. Geoph*. 15(3), 135-146.

Sudduth, K.A., Kitchen, N.R., Wiebold, W.J., Batchelor, W.D., Bollero, G.A., Bullock, D.G., Clay, D.E, Palm, H.L., Pierce, F.J., Schuler, R.T., Thelen, K.D. 2005., Relating apparent electrical conductivity to soil properties across the north-central USA. *Comput. Electron Agric*. 46, 263-283.

Sudduth, K.A., Myers, D.B., Kitchen, N.R., Drummond, S.T., 2013. Modeling soil electrical conductivity-depth relationships with data from proximal and penetrating EC_a sensors. *Geoderma*. 199, 12-21.

Taylor, J.A., Whelan, B.M., 2011. Selection of ancillary data to derive production management units in sweet corn (*Zea Mays* var. rugosa) using MANOVA and an information criterion. *Precis. Agric*. 12, 519-533.

Thomas, D.R., 1992. Interpreting Discriminant Functions. A Data Analytic Approach. *Multivariate Behav. Res.* 27(3), 335-362.

Triantafilis, J., Monteiro Santos, F.A. 2013. Electromagnetic conductivity imaging (EMCI) of soil using a DUALEM-421 and inversion modelling software (EM4Soil). *Geoderma*. 211-212, 28-38.

Triantafilis, J., Terhune IV, C.H., Monteiro Santos, F.A., 2013. An inversion approach to generate electromagnetic conductivity images from signal data. *Environ. Modell. Softw*. 43, 88-95.

Unamunzaga, O., Besga, G., Castellón, A., Usón, M.A., Chéry, P., Gallejones, P., Aizpurua, A., 2014. Spatial and vertical analysis of soil properties in a Mediterranean vineyard soil. *Soil Use Manage*. 30, 285-296.

Visser, H., de Nijs, T., 2006. The Map Comparison Kit. Environmental Modelling & Software.21, 346-358.

Vitharana, U.W.A., Van Meirvenne, M., Simpson, D., Cockx, L., De Baerdemaeker, J., 2008. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma*. 143, 206-215.

Warne, R.T., 2014. A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists. Practical Assessment. *Res. Eval*. 19(17).

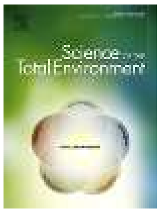
Chapter 4

Spatial variability in orchards after land transformation

This chapter contains the following published paper in the journal *Science of the Total Environment.* JCR-SCI Impact Factor: 4.610. Category: Environmental science; 1st Quartile

Uribeetxebarria, A., Daniele, E., Escolà, A., Arnó, J., Martínez-Casasnovas, J.A. 2018. Spatial variability in orchards after land transformation: Consequences for precision agriculture practices. *Sci. Total Environ*. 635, 343–352.

Spatial variability in orchards after land transformation: consequences for precision agriculture practices



Asier Uribeetxebarria, Elisa Daniele, Alexandre Escolà, Jaume Arnó, José Antonio Martínez-Casasnovas

ABSTRACT

The change from traditional to a more mechanized and technical agriculture has involved, in many cases, land transformations. This has supposed alteration of landforms and soils, with significant consequences. The effects of induced soil variability and the subsequent implications in site-specific crop management have not been sufficiently studied. The present work investigated the application of a resistivity soil sensor (Veris 3100), to map the apparent electrical conductivity (ECa), and detailed multispectral airborne images to analyse soil and crop spatial variability to assist in site-specific orchard management. The study was carried out in a peach orchard (*Prunus persica* (L.) Stokes), in an area transformed in the 1980 decade to change from rainfed arable crops to irrigated orchards. A total of 40 soil samples at two depths (0-30 cm and 30-60 cm) were analysed and compared to ECa and the normalised difference vegetation index (NDVI). Two types of statistical analysis were performed between ECa or NDVI classes with soil properties: a linear correlation analysis and multivariate analysis of variance (MANOVA). The results showed that the land transformation altered the spatial distribution and continuity of soil properties. Although a relationship between ECa and peach tree vigour could be expected, it was not found, even in the case of trees planted in soils with salts content above the tolerance threshold. Two types of management zones were proposed: a) zones delineated according to ECa classes to leach salts in the high ECa zones, and b) zones delineated according to NDVI classes to regulate tree vigour and yield. These strategies respond to the alteration of the original soil functions due to the land transformation carried out in previous years.

KEY WORDS

Land use change, Apparent electrical conductivity, Vegetation index, Potential management zones, MANOVA

1. INTRODUCTION

Since the mid-twentieth century, and particularly since the 1980s-90s, traditional agriculture is undergoing a change to a more modern, mechanized and technical agriculture. In many cases, these changes have involved land transformation, with land use alteration and intensification (Ritcher, 1984). This has been the case of cash crop development by the market-oriented agriculture. It is a global phenomenon that has promoted the expansion of hazelnut, rubber, fruit, and tea in developing tropical and subtropical countries (Xiao et al., 2015); citrus in Brazil (Moraes et al., 2017); or vineyards, almonds, olive and fruit trees in the Mediterranean Europe (García-Ruiz, 2010; Martínez-Casasnovas et al., 2010a), among others.

This intensification of agriculture has supposed the alteration of landforms and soils, with significant ecological consequences (Xiao et al., 2015). Some works have reported specific examples of those effects. For example, local hydrology (Yi et al., 2014), soil profile dismantlement (Laudicina et al., 2016; Öztekin, 2013), acceleration of soil erosion (García-Ruiz, 2010; Ramos and Martínez-Casasnovas, 2010; Xiao et al., 2015), fragmentation of traditional landscapes (de Oliveira et al., 2017), increase of CO2 emissions (Carlson et al., 2013), elimination of traditional soil conservation measures and increase of soil spatial variability (Laudicina et al., 2016; Martínez-Casasnovas and Ramos, 2009; Su et al., 2016; Xiao et al., 2015). Another major problem is the effect of topsoil removing on plant growth. Reduced growth may occur on the fill areas (Martínez-Casasnovas et al., 2010b), although the exposure of subsoil in the cuts is usually a more serious problem (Öztekin, 2013). Moreover, many of these land transformations have been supported by subsidies, as happens in the Mediterranean Europe, where many orchards planted in the last decades were also supported by the European Agricultural Policy in response to market demand (Cots-Folch et al., 2009; Nainggolan et al., 2012).

Although there have been attempts to document the process of cash crops expansion, the effects of induced soil variability due to land transformations and the subsequent implications in crop management have not been sufficiently investigated. However, this is of particular interest to fruit growers since, due to the soil-plant interaction, fruit trees development and their potential production are affected by the spatial variability of soil properties (Khan et al., 2016; Panda et al., 2010; Pedrera-Parrilla et al., 2016). Then, changes produced by land transformations can become a main constraint to consider when planning orchard management operations (Fulton et al., 2011). On the other hand, field size should not be considered as a limitation for precision agriculture applications in fruticulture since, even in small orchards, there may be differences in soil properties affecting tree growth and fruit quality (Käthner & Zude-Sasse, 2015; Zude-Sasse et al., 2016).

Soil information is often not available at a spatial resolution intrinsically needed for precision agriculture or other site-specific soil uses and management purposes (Mertens et al., 2008); and specifically is not available after land transformations. One approach to obtain detailed spatially distributed soil data is the non-invasive measurement of the apparent electrical conductivity (ECa). Soil sensors for on-the-go ECa mapping are increasingly used for this purpose (Corwin and Lesch, 2003; Fulton et al., 2011; Mertens et al., 2008), and to delineate management zones according to the concept of precision agriculture (Moral et al., 2010; Peralta & Costa, 2013). In orchards, ECa has been used for the analysis of soil variability, and some researchers have found correlations between ECa, generative tree growth, fruit development and fruit size (Käthner and Zude-Sasse, 2015; Zude-Sasse et al., 2016). In this respect, it was pointed out that fruit development and soil ECa were well correlated. However, quality parameters, although very variable, are spatially less stable and may be poorly related to the ECa as indicated Aggelopoulou et al. (2010) in apple tree plantations. Regarding the interpretation of the ECa signal, some authors have highlighted the difficulty to determine the soil properties that most affect the variability of ECa in a particular field (Uribeetxebarria et al., 2018). Because of that, they proposed the use of multivariate analysis of variance (MANOVA) to better interpret which soil properties are behind the variation of the electrical conductivity signal. This was particularly useful in orchards affected by previous parcelling (Uribeetxebarria et al., 2018).

Additionally, site-specific management zones (SSMZ) can also be delimited based on remote sensing data. In this respect, the most frequently used vegetation index is the Normalised Difference Vegetation Index (NDVI) (Rouse et al., 1974), which is feasible in low-chlorophyll fruits and canopy imaging. NDVI is correlated to plant vigour and has strong interaction with yield and sometimes quality parameters (Zude-Sasse et al., 2016). Different authors have used spectral indices to estimate orchard variables. For example, Peña-Barragán et al. (2004) developed a methodology to determine tree cover in olive groves using aerial images and different spectral indexes. González-Dugo et al. (2013), using high resolution airborne thermal imagery, assessed the heterogeneity in water status in commercial orchards (almond, apricot, peach, lemon and orange), as a prerequisite for precision irrigation management. Other authors (Noori and Panda, 2016) also studied the relationship between vegetation indexes (SAVI, NDVI and Vegetative Vigour Index) with field environmental data including soil and tree structure attributes in an olive orchard, suggesting that these relationships would help in Site Specific Crop Management (SSCM) of orchards. Other works used airborne hyperspectral imagery for predicting yield in citrus crops (Ye et al., 2009), or more specifically, to quantify fluorescence emission in a commercial citrus orchard as well (Zarco-Tejada et al., 2016). In the latter, the objective was to track photosynthesis at different phenological and stress stages throughout the season to suggest its operational use in precision agriculture.

Despite these findings and advances, there are not many examples of practical application of SSCM in commercial fruit orchards (Noori & Panda, 2016). However, emerging research knowledge in this field demonstrates clear advantages of precision agriculture tools in fruit production management. In this respect, different authors suggest the combination of ECa with spectral vegetation indices to help in the delineation of SSMZ (De Benedetto et al., 2013; Ortega-Blu & Molina-Roco, 2016; Panda et al., 2010). This approach allows identifying homogenous sub-field areas related to the intrinsic properties of soil and, above all, differentiated crop response. This is because ECa, by itself, may not be a good estimator of the most commonly measured soil properties and, under irrigation and fertigation conditions, the vegetation status may be more affected by water and nutrient management than by soil properties (De Benedetto et al., 2013).

ECa and/or spectral vegetation indices have been mainly applied in field crops and vineyard (Priori et al., 2013), but fewer studies refer to their use in fruit orchards, and even less in Mediterranean latitudes. One important reason could be the small sized orchards usually have there. Nevertheless, and as pointed out by Käthner & Zude-Sasse (2015) and Arnó et al. (2017), even in small orchards there may be differences in soil properties affecting tree growth and fruit quality.

As showed above, precision agriculture applications in tree crops are rather limited in the literature (Aggelopoulou et al., 2011). Moreover, as suggested by Öztekin (2013), after land transformations, some site-specific management practices should be taken into account to regain productivity and improve homogeneity in soil properties. However, to the best of our knowledge, there are no works addressing the implications of land transformations in precision agriculture. In this context, the aim of the present work was to investigate the effects of land transformations in soil variability and crop development in a peach orchard (*Prunus persica* (L.) Stokes) located in Lleida (Catalonia, NE Spain). The area suffered land transformations in the 1980 decade to enlarge fields and changed from rainfed crops to irrigated orchards. The hypothesis was that land transformations aimed to enlarge fields, instead to create more homogeneous areas, alter the spatial distribution and continuity of soil properties, with consequences in crop vigour. This opens an opportunity to precision agriculture techniques to assist in site-specific orchard management.

2. MATERIALS AND METHODS

2.1. Study area

The research was carried out in a 2.24 ha commercial peach tree orchard located 20 km south from the city of Lleida (Catalonia, NE Spain) (Lat 41.477157°, Long 0.509500° WGS84) (Fig.1). It was planted in 2012 with white peach (Prunus persica (L.) Stokes, var. Patty), which is early harvested. The training system was the so-called “Catalan” vase or vessel shape, with a plantation pattern of 5x2 m. Peach trees were fertirrigated by means of a drip irrigation system. The system consisted of a unique irrigation sector, with one drip tubing per tree row and two emitters of 2 l h-1 per tree.

The elevation ranges from 156 – 167 m a.s.l. The slope is gentle to moderate, with an average of 5.3 % and a direction almost parallel to the tree rows (Fig. 1). The current morphology is the result of land clearing and levelling carried out during the 1980 decade. Previously, the relief was composed of low hills, with terraces protected with stone walls. Land transformation was carried out with heavy machinery (bulldozers). First, stones were removed from the old terraces and then terrain was smoothed. The upper more fertile soil layer was not specifically preserved. Marls and calcareous rocks belonging to the subsoil were put on the surface.

Soils of the area were classified as Typic Xerorthent, coarse-silty, mixed (calcareous), thermic (Soil Survey Staff, 2014). They have a typical sequence of horizons Ap-Bw-C (lutites), with the latter usually presenting a moderate salt content.

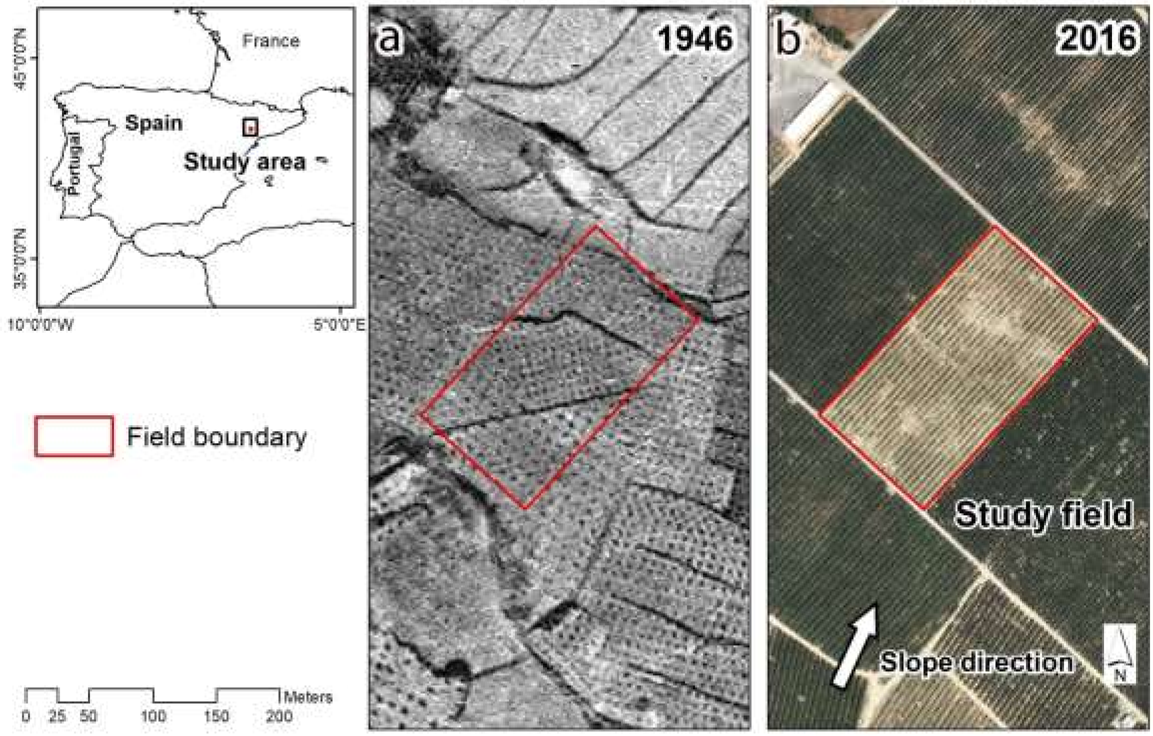


Fig. 1 Location of the study area and comparison of land uses and crop systems before and after the land.

2.2. Apparent electrical conductivity survey

An ECa survey was conducted on March 1st, 2016. The survey was carried out with a Veris 3100 sensor (Veris Technologies Inc. Salina, Kansas, USA). Veris 3100 uses two ECa arrays to measure the 0-30 cm (shallow ECa) and 0-90 cm (deep ECa) soil depths simultaneously. Data was georeferenced by means of a Trimble AgGPS332 receiver with EGNOS differential correction in geographic coordinates WGS84 (EPSG 4326). ECa values above or below ±2.5 standard deviations (SD) were considered outliers and were removed from the original data file according to the criteria of Taylor et al. (2007). The final ECa data set consisted of 1668 points with shallow and deep readings. For interpretation and comparison purposes, ECa values were standardized at the reference temperature of 25 °C. In order to do that, a polynomial function was used as proposed by Sheets and Hendrickx (Ma et al., 2011). The adjusted ECa values were then renamed to EC25 and expressed in dS m-1 at 25 °C. These data were interpolated on a 1-m grid by means of ordinary kriging using the exponential semivariogram model. In addition, anisotropy was considered. This was because semivariance values presented a clear directional distribution in the NW to SE direction, perpendicular to the tree rows. For this purpose, ArcGIS Geostatistical Analyst 10.4 (ESRI, Redlands, California, USA) was used.

2.3. Soil sampling

Before soil sampling, an unsupervised classification of the shallow and deep EC25 maps in 5 different classes was performed (Fig. 2) in order to stratify samples to obtain more structured information about the soil of the plot. For that, the ISODATA algorithm implemented in the Image Analyst of ArcGIS 10.4 was applied. The ISODATA is a k-means algorithm that uses minimum Euclidean distance to assign a cluster to each candidate pixel in an iterative process (Jensen, 1996), removing redundant clusters or clusters to which not enough pixels are assigned. Table 1 shows the average and standard deviation of the shallow and deep EC25 in the 5 classes that showed increasing electrical conductivity values.

Table 1. Unsupervised classes based on the shallow and deep EC_{25}[1] and basic statistics (average and standard deviation) for each class.

Class	Shallow EC_{25} dS m^{-1} at 25 °C	Deep EC_{25} dS m^{-1} at 25 °C
1	0.74±0.20	0.69±0.18
2	1.21±0.18	1.02±0.17
3	1.57±0.15	1.32±0.16
4	1.93±0.14	1.58±0.17
5	2.39±0.19	1.86±0.19

[1] EC_{25}: apparent electrical conductivity values (ECa) standardized at the reference temperature of 25 °C.

In each EC25 class, eight sampling points were randomly distributed, considering a minimum distance of 30 m between sampling points. This minimum distance corresponded to the range of the exponential semivariogram to interpolate ECa data from the Veris 3100 sensor. A total of 40 points were sampled in the plot (Fig. 2). Soils were sampled with an auger up to 90 cm or up to the limiting layer depth. This limiting layer corresponded to Tertiary lutites. The samples were taken in the space between the tree rows, between the central marks of the Veris 3100 coulters. The following properties were analysed for the 0-30 cm and 30-60 cm layers: pH, electrical conductivity 1:5 soil:water extract (EC1:5), equivalent calcium carbonate (CaCO3), cationic exchange capacity (CEC), particle-size (texture), water holding capacity (WHC) at -33 and -1500 kPa (field capacity and wilting point, respectively), and organic matter content (Org M) (the latter only at the 0-30 cm layer).

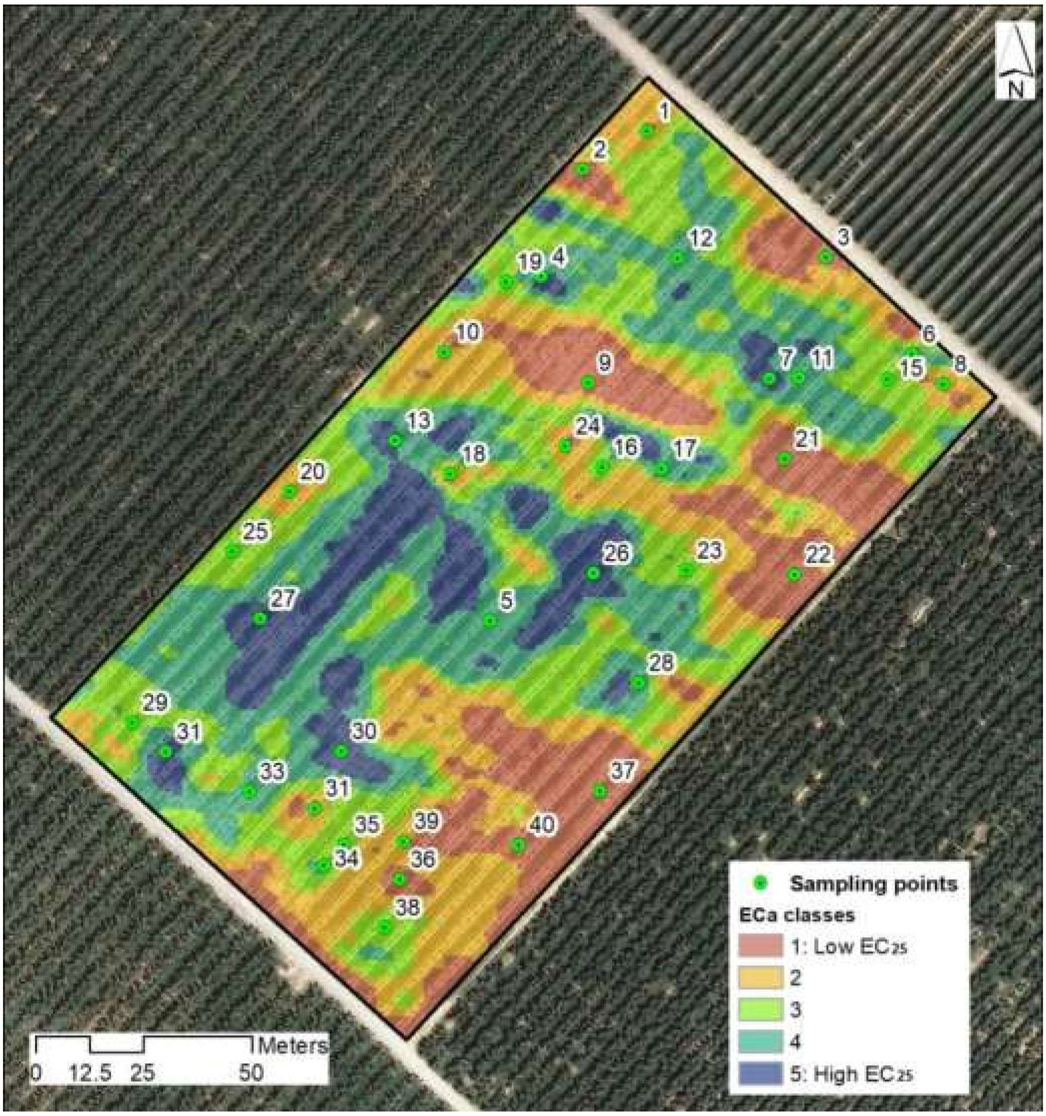


Fig. 2 Location of the soil sampling points stratified according to the spatial variability of EC25 data organized in 5 classes. See Table 1 for class description. (EC25: apparent electrical conductivity (ECa) standardized at the reference temperature of 25 °C).

2.4. Multispectral data acquisition and vegetation index

A 4-band multispectral image was acquired on May 16th, 2016 (approximately one month before harvest). For that, a Digital Multi-Spectral Camera (DMSC) (Specterra Services-Australia) mounted on a CESSNA 172S SKYHAWK airplane operated by RS Servicios de Teledetección (Lleida, Spain) was used. The DMSC captured four spectral bands 20 nm width, centred at 450 nm (blue), 550 nm (green), 675 nm (red) and 780 nm (near infrared). The spatial resolution of the image was 0.25 m. The image was pre-processed by the provider's software to compensate for miss-registration due to lens distortion, less than 0.2 pixels, and for scene brightness due to the bi-directional reflectance distribution function (Wallace et al., 2008). Absolute radiometric calibration was not carried out since the purpose of the study was not a multi-temporal analysis of the tree vigour.

The near infrared and red bands were used to calculate the Normalised Difference Vegetation Index (NDVI) (Rouse et al., 1974) according to Equation 1.

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (1)$$

where NIR is the near infrared band (780 nm) and Red the red band (675 nm) of the multispectral image.

Only the pixels including canopy vegetation of peach trees (NDVI > 0.4) were mapped (Fig. 3). These pixels were then used to define the tree canopy cover. This was done by converting the NDVI mask to a polygon layer and segmenting joined canopies into individual polygons. In this way, each tree in the plot was identified as an individual object. The polygons were used to calculate per tree NDVI zonal statistics (min, max, mean and standard deviation). These basic statistics were merged to the tree canopy layer. Finally, the polygons were converted to points using ArcGIS 10.4 and were stored in a point layer. The tool was forced to locate the centroids inside the polygons (Fig. 3). With the trees represented by their centroids, ordinary kriging with an exponential semivariogram was performed to interpolate a surface with the NDVI-per-tree continuous spatial distribution.

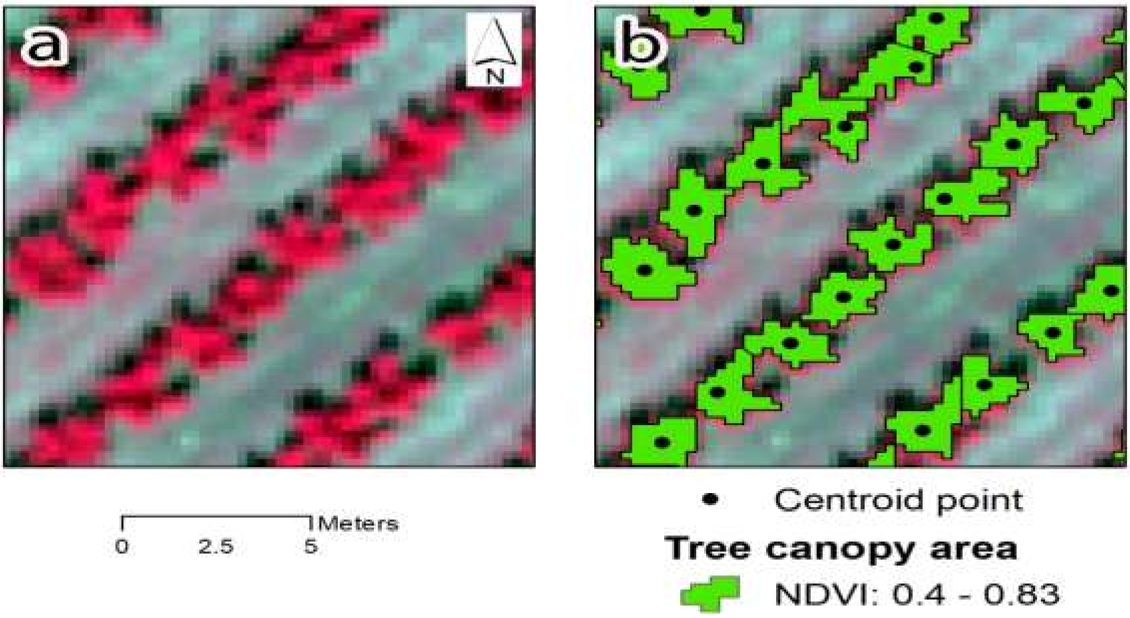


Fig. 3 (**a**) False colour composite RGB (NIR-Red-Green) of the airborne multispectral image acquired on May 16th, 2016 (pixel size 0.25 m). (**b**) Example of the tree canopy area derived from the classification of the NDVI above 0.4 and centroids of the tree canopy areas (polygons).

2.5. Statistical analysis

A linear correlation analysis (Pearson test) was carried out between individual soil properties, EC25 and NDVI values at the sampling points. As shown later, some unexpected differences were found in the relationships between the EC25 and NDVI with soil properties, probably because soil samples were taken in the alleways and not on the tree rows where localized irrigation can influence certain soil properties. For this reason, new 20 points situated upon the wet bulb and located 2.5 m from the previously sampling points were sampled and analysed (0-30 cm). To compare the means of the soil variables according to the location (inside or outside the wet bulb), a series of multiple t-tests were performed, adjusting the usual significance level of 0.05 with Bonferroni correction (Faraway, 2014). The software JMP Pro 12 (SAS Institute Inc.) was used for this purpose.

Different types of potential management zones were delineated according to the shallow and deep EC25 and NDVI surface data applying unsupervised classifications by means of the ISODATA algorithm implemented in the Image Analyst of ArcGIS 10.4. For each parameter, 2 classes (high and low values) were created. As proposed by Uribeetxebarria et al. (2018), the difference between high and low classes were determined by applying in each case a multivariate analysis of variance (MANOVA) of soil properties. This method was preferred instead of a separate analysis of variance (ANOVA) for each soil property to avoid misleading and inconsistent results. In fact, ECa and NDVI values can be considered as the result of the combined effect of soil properties as a whole (Corwin and Lesch, 2003), and delimitation of areas within the plot based on ECa or NDVI maps should be checked from a multivariate approach.

To interpret the results of the MANOVAs, a descriptive discriminant analysis (DDA) was used (Uribeetxebarria et al., 2018). As a result of the procedure, linear combinations of soil properties (discriminant functions) were provided, which managed to separate the two classes of EC25 (shallow or deep) or NDVI in a meaningful way. Standardized coefficients of each discriminant function (SDFCs) and structure coefficients (SCs) were used for interpretation. The first, SDFCs, were indicative of the contribution of each soil variable to the discriminant function, whereas the SCs were the correlations between each observed variable and the discriminant function scores. The most important soil variables were finally identified through the parallel discriminant ratio coefficients (parallel DRCs), by multiplying SDFCs by the SCs. Then, parallel DRCs were used to identify non-redundant soil variables contributing to discriminate two types of soil in terms of EC25 or NDVI.

3. RESULTS

3.1. Soil properties

The soils of the study area were characterised by a basic pH of 8.2±0.2 - 8.3±0.2 and an average EC1:5 between 1.6±0.8 and 1.8±0.7 dS m-1 at 25 °C (Table 2). In addition, soils had a high content of calcium carbonate (33.3±6.3 - 37.4±9.5 %). The average CEC was low to moderate (10.3±2.2 - 9.4±2.7 meq 100g-1). The organic matter content of the first layer was also low to moderate (2.2±0.7 %) and the AWHC of both layers was very similar (9.8±1.5 - 10.5±1.1 %). Taking into account an average bulk density of 1400 kg m-3, the available water holding capacity (AWHC) in the average soil depth (61.1 cm) would be 86.8±11.1 mm, indicating a low AWHC for a xeric soil moisture regime and the necessity to irrigate the fruit trees. Regarding the texture, the most frequent textural classes were loam, clay loam or silty clay loam, which do not represent particular limitations for crop development.

Table 2. Basic statistics of soil properties for the 0-30 cm and 30-60 cm depth layers (average values and standard deviation).

Soil property	0-30 cm	30-60 cm
$pH_{1:2.5}$	8.2±0.2	8.3±0.2
$EC_{1:5}$ (dS m^{-1})	1.6±0.8	1.8±0.7
$CaCO_3$ (%)	33.3±6.3	37.4±9.5
CEC (meq $100g^{-1}$)	10.3±2.2	9.4±2.7
Org M (%)	2.2±0.7	-
WHC -33kPa (%)	22.7±2.8	23.2±2.4
WHC -1500kPa (%)	12.9±2.3	12.8±2.0
AWHC (%)	9.8±1.5	10.5±1.1
Clay (%)	23.9±5.8	25.6±4.6
Silt (%)	38.8±8.9	42.2±8.4
Sand (%)	34.8±10.6	30.7±12.1
Soil depth (cm)(1)	61.1±21.0	-

(1) The soil depth refers to the depth of the whole soil profile. The maximum measured soil depth was 90 cm, which was the maximum reached with the auger hole used for soil sampling

3.2. EC25 and vegetation index: spatial pattern and comparison with former landforms

The soil volume explored by the Veris 3100 ECa surveyor presented average values of 1.54±0.52 dS m-1 (0-30 cm) and 1.28±0.40 dS m-1 (0-90 cm) at 25 °C.

As described in section 2.1 (study area), land levelling works were carried out prior to planting the fruit trees. Stone-wall terraces were removed in order to enlarge fields. Figure 4 shows the comparison between the location of the old stone-wall terraces and the apparent electrical conductivity surface (shallow and deep readings). Lower EC25 values appeared in the northern part of the plot, where trees had also low vigour values, and in the southern part of the plot following the pattern of the terraces. Between the terraces there were higher EC25 values, probably due to the existence of soils with higher clay content (23.4 %) and less sand (34.7 %) than in the southern part of the plot (21.7 % clay and 43.5 % sand), as result of the land levelling works.

Regarding NDVI, average per tree values ranged from 0.40 to 0.75. Two main zones could be distinguished: one with higher NDVI values, in the northern part of the plot, and another with lower NDVI values, in the south (Fig. 4).

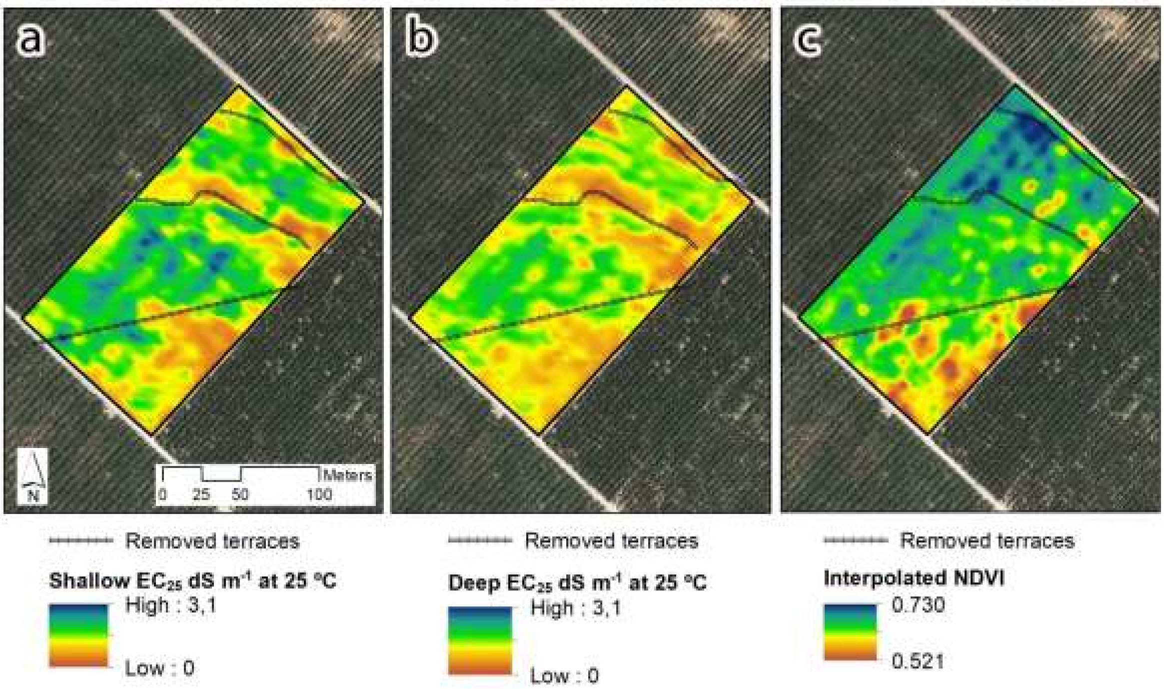


Fig.4. Comparison between the locations of the removed stone-wall terraces, the apparent electrical conductivity ((a) shallow and (b) deep EC_{25} dS m^{-1} at 25 °C) and (c) the interpolated NDVI in the study plot.

3.3. Relationship between soil properties, EC25 and vegetation index

Table 3 shows the correlation coefficients between the soil properties of the two analysed layers (0-30 cm and 30-60 cm), the EC25 (shallow and deep) and the NDVI. As expected, significant positive correlations were found between both measures of EC25 and EC1:5 (0.547 and 0.575, p-value < 0.01). Regarding the availability of water, only the shallow EC25 showed a positive correlation with the WHC at -1500 kPa (p-value < 0.05). On the other hand, soil depth presented a positive correlation (p-value < 0.01) with both EC25 readings.

Differently from the relationships with the EC25, the NDVI was not related to properties such as shallow or deep EC25, nor with EC1:5, water holding capacity or soil depth (Table 3). Only textural fractions coarser than clay were correlated. In the case of sand, the relationship was negative, probably indicating that at higher sand content the trees were less vigorous. This was an expected relationship since higher sand contents indicate less soil fertility. Regarding other properties, only the CEC at 30-60 cm showed a positive relationship with the NDVI. Although premature to conclude, this correlation would be expected as a sign of better soil fertility conditions in these locations.

Table 3 Correlation coefficients between soil properties (0-30 cm and 30-60 cm), shallow and deep EC_{25} and NDVI (N=40).

	Shallow EC_{25} with **0-30 cm** **soil samples**	**Deep EC_{25}** with **30-60 cm soil samples**	**NDVI** with **0-30 cm** **soil samples**	**NDVI** with **30-60 cm** **soil samples**
Shallow EC_{25}	-	0.910**	0.002	-
Deep EC_{25}	0.910**	-	0.156	-
$pH_{1:2.5}$	-0.193	0.360*	0.110	-0.045
$EC_{1:5}$ (dS m^{-1})	0.547**	0.575**	0.075	0.164
$CaCO_3$ (%)	-0.086	0.037	0.119	0.062
CEC (meq $100g^{-1}$)	0.136	0.313	0.284	0.446**
Org M (%)	0.129	-	0.097	-
WHC -33kPa (%)	0.269	0.250	0.167	0.084
WHC -1500kPa (%)	0.337*	0.217	0.241	0.260
Clay (%)	0.207	0.252	0.048	0.253
Silt (%)	0.123	0.194	0.523**	0.194
Sand (%)	-0.197	-0.196	-0.365*	-0.233
Soil depth (cm)	0.439**	0.487**	0.077	-

*p-value < 0.05; ** p-value < 0.01

To check the differences found before in the relationships between EC25 and NDVI with soil properties, Table 4 shows the results of the comparison of some soil properties (pH1:2.5, EC1:5, organic matter and CaCO3) at 20 locations, inside and outside the wet bulb. The significant differences between samples inside and outside the bulb would explain the different expected relationship between soil properties and the ECa, the latter measured outside the bulb; and between these same soil properties and the NDVI, the latter mainly conditioned by the drip irrigation system.

Table 4 Comparison of some relevant soil properties inside and outside the wet bulb (0-30 cm). Results of *t*-test adjusted by the Bonferroni correction (N=20).

Soil property	Inside wet bulb	Outside wet bulb	t-test p-value
$pH_{1:2.5}$	7.96±0.1	8.23±0.24	<0.01
$EC_{1:5}$ (dS m^{-1})	0.97±0.47	1.43±0.54	<0.01
Org M (%)	3.01±0.47	2.15±0.66	<0.01
$CaCO_3$ (%)	29.24±6.23	31.32±6.32	0.301

3.4. Zonal analysis between soil properties, EC25 and vegetation index

In addition to the previous analysis, different multivariate analyses of variance (MANOVAs) were performed to determine specific soil properties mainly linked to the spatial variation of EC25 and NDVI classes. Results are presented in Table 5. Regarding the shallow EC25 classes, properties such as EC1:5, soil depth and clay were the ones that contributed most to the discriminant function explaining the spatial variation. The importance of those properties is highlighted by the parallel discriminant ratio coefficients (parallel DRC), which indicates the relative contribution of each soil property in the canonical function. Similarly, in the case of the deep EC25 classes, the soil depth and EC1:5 were also key properties in the variation of the ECa in addition to silt as a textural class. These results were similar to those obtained previously with the linear correlation analysis (Table 3). However, the use of MANOVAs made it possible to notice that, as expected, the use of ECa sensors is a good tool to indirectly observe the spatial variation of soil texture. On the other hand, the possible influence of the water content was now unnoticed. This should not be a major problem in the case of irrigation. From the point of view of land transformation, the ECa signal allowed to know the spatial variability of properties that are more stable in time, such as depth and soil texture, were now detected.

Regarding the NDVI classes, contribution to the discriminant function was found only for silt and sand contents in the top layer; and with the water holding capacity at -1500 kPa and silt content in the 30-60 cm layer (Table 5). The expected contribution of soil depth to differentiate NDVI classes was not found.

Table 5 Results of the descriptive discriminant analysis (DDA) of soil properties affecting EC_{25} and NDVI at two different soil depths (0-30 cm and 30-60 cm).

		Soil properties (0-30 cm)										
		Depth	pH	$EC_{1:5}$	$CaCO_3$	Org M	CEC	Clay	Silt	Sand	WHC -33kPa	WHC -1500kPa
Shallow EC_{25}	SDFC	0.64	-0.42	0.79	0.21	-	-0.16	0.86	-0.02	0.12	-	-0.24
	SC	0.51	-0.33	0.51	-0.11	-	0.133	0.32	0.13	-0.20	-	0.37
	Parallel DRC	0.33*	0.14	0.41*	-0.02	-	-0.02	0.28*	0.00	-0.03	-	-0.09
NDVI	SDFC	-0.18	-	0.10	0.37	0.32	-0.36	-	0.99	-0.26	-0.31	0.19
	SC	-0.02	-	0.03	0.21	0.36	0.57	-	0.86	-0.72	0.44	0.43
	Parallel DRC	0.00	-	0	0.08	0.12	-0.21	-	0.86*	0.19*	-0.13	0.08
		Soil properties (30-60 cm)										
		Depth	pH	$EC_{1:5}$	$CaCO_3$	Org M	CEC	Clay	Silt	Sand	WHC -33kPa	WHC -1500kPa
Deep EC_{25}	SDFC	1.03	0.02	1.32	0.33	-	0.32	0.66	1.57	1.41	0.20	-1.30
	SC	0.48	0.22	0.33	-0.01	-	0.09	0.10	0.12	-0.11	0.09	0.07
	Parallel DRC	0.50*	0.00	0.44*	0.00		0.03	0.07	0.19*	-0.16	0.02	-0.10
NDVI	SDFC	-	-0.87	0.54	0.74	-	-	-	1.29	1.52	-2.39	2.62
	SC	-	-0.16	0.14	0.16	-	-	-	0.13	-0.15	-0.04	0.23
	Parallel DRC	-	0.14	0.08	0.13	-	-	-	0.17*	-0.24	0.10	0.62*

SDFC: Standardized discriminant function coefficient; SC: structure coefficient; Parallel DRC: parallel discriminant ratio coefficient; Hyphens indicate variables that were removed to obtain significant discriminant functions. * Indicates properties with a greater contribution to the discriminant function from MANOVA.

As previously noted, soil properties inside and outside the wet bulb vary significantly. Because of that, and although the spatial distribution of the ECa and the NDVI seem to follow similar patterns (Fig. 4), it is difficult to discern whether the NDVI varies linked to the variation of soil properties as a whole or, as the results of Table 5 seem to suggest, there is some local influence given the irrigation system used. In other words, it can be thought that there could be something altering the relationship between soil properties, EC25 and NDVI. To find the reason of this alteration, a combination of the EC25 and NDVI high and low classes was done and the coincidences and differences were mapped. The results can be observed in Fig. 5, which shows the four combinations of low and high EC25 and NDVI classes, with the coincidences high-high and low-low showed in green colours. The major inconsistencies between NDVI and EC25 were located upon the old terraces (in red and blue colours).

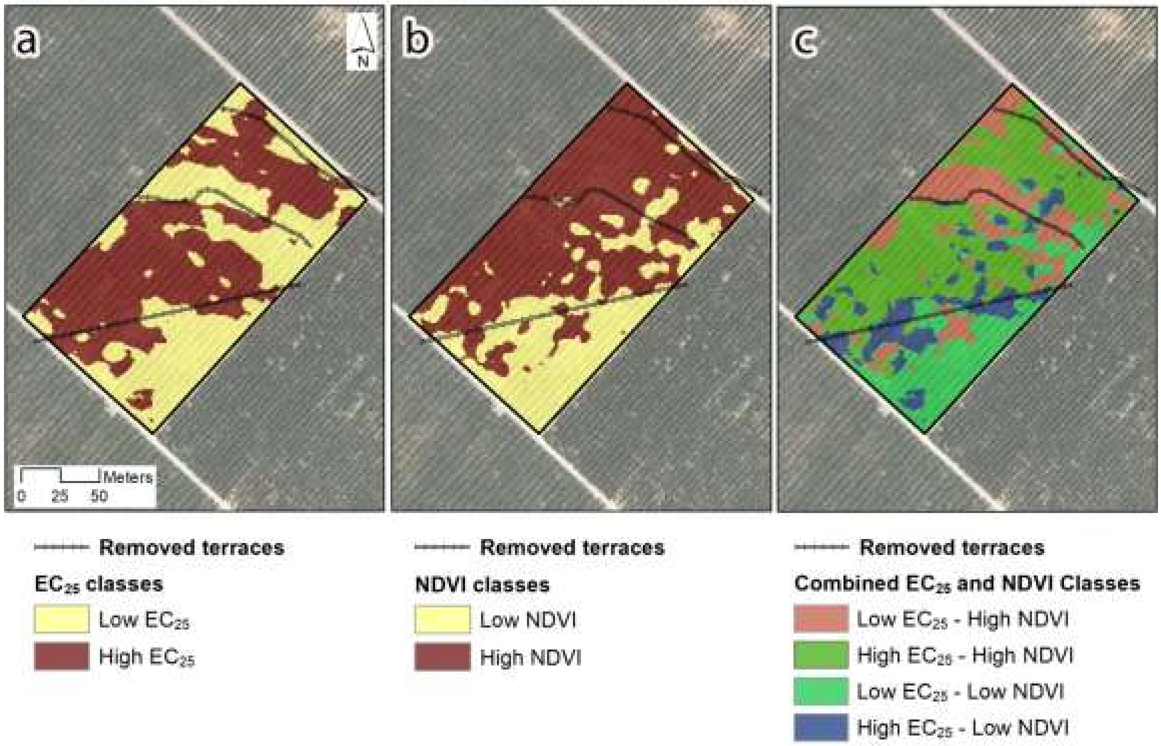


Fig. 5 Comparison between the location of old stone-wall terraces and (a) EC_{25} classes, (b) NDVI classes and (c) combined classes of EC_{25} and NDVI.

4. DISCUSSION

The soils of the study area were characterised by average EC1:5 values at 25 °C between 1.6±0.8 dS m-1 (0-30 cm) and 1.8±0.7 dS m-1 (30-60 cm) (Table 2), although there were maximum values of 3.58 dS m-1. These mean values (< 2 dS m-1 at 25 °C) were indicative of non-saline soils or slightly saline soils (2-4 dS m-1 at 25 °C) (Rhoades et al., 1999). However, this classification refers to salinity in saturated extracts of soil, but not to the 1:5 extract analysed in this work. Therefore, this interpretation may not be conclusive, although induces to think that peach trees development could be influenced by salinity in some parts of the plot, since peach trees are sensitive to salts, with a threshold value around 1.7 dS m-1 at 25 °C. According to Tanji and Kielen (2002) and Stassen and Wooldridge (2011), yields begin to diminish at 1.5 dS m-1, and at 2.7 dS m-1 yields may be reduced by 50 % (Maas & Grattan, 1999 in Stassen and Wooldridge, 2011). The main effects of high salinity are reduction of water uptake and the onset of sodium and, particularly, chloride toxicities. These toxicities cause leaf burn, reduced vigour, stunted growth and low yields. Moreover, the slope of the reference regression line between electrical conductivity and yield is -21 % (Tanji and Kielen, 2002), which indicates a fast yield decrease as salinity increases above the threshold.

Regarding the values of EC25 (shallow and deep), average values of 1.54±0.52 dS m-1 at 25 °C (0-30 cm) and 1.28±0.40 dS m-1 at 25 °C (0-90 cm), indicate that deep readings were lower on average, which is opposite to the EC1:5 measured in the soil samples. Nevertheless, both types of measures are not totally comparable, since deep measurements made with the ECa surveyor integrate the reading from 0 to 90 or 100 cm (Sudduth et al., 2005), and the results of the soil samples were specifically from 30 to 60 cm. The spatial pattern of both EC25 readings in Fig. 4 presents these differences in the average values, but shows the consistency and continuity of the signal in the two layers. However, per tree NDVI pattern presented some relevant differences with respect EC25 variability. NDVI showed a more continuous and gradual distribution from the southern part of the plot to the northern part. Unlike the case of the EC25 spatial variability, the NDVI did not show a significant discontinuity where the old terraces were located (Fig. 4). This could reveal a specific behaviour of the fruit trees, in terms of development, independent of the soil properties that determined the electrical conductivity readings by the resistivity sensor.

The analysis of the relationship between EC25 and NDVI (Table 3) revealed a lack of it. In addition, the results of the MANOVAs (Table 5) showed that the soil properties that contributed to the canonical discriminant functions of the EC25 and NDVI were not the same. In this case, the cause for this different behaviour between variables could be the influence of the fertigation system (drip irrigation), which maintains the root area free of salts, or with certain levels that are tolerated by the peach trees. This is in line with the findings of De Benedetto et al. (2013), who stated that under irrigation conditions vegetation might be more affected by water management than by soil properties. The results of Table 4 confirmed the hypothesis of the differences in relevant properties measured inside and outside the wet bulb, which could be responsible for the differences found in the relationships between the EC25 and the NDVI with soil properties. Except for the $CaCO_3$ content, which was very high in both locations (inside and outside the bulb), the rest of the properties showed significant differences (p-value <0.01). It is worth noting the differences in the EC1:5, which were significantly lower within the wet bulb. This means that peach trees were maintained with a tolerable salt content thanks to the drip irrigation system. In other words, tree vigour would not be affected by the salts content detected outside the wet bulbs by the ECa surveyor. The same reasoning could be applied to the lack of relationship between NDVI and soil depth. One could expect the higher the soil depth, the higher vigour trees, but this was not the case in the present study orchard, because the water and nutrients were supplied by fertigation and soil depth (61.1±21.0 cm) is not a constraint for tree development.

As mentioned above, the lack of relationship between EC25 and NDVI would be particularly affecting the areas where differences in EC25 and NDVI classes occur (high-low and low-high), which were found where the old terraces were located in the past, before land transformation (Fig. 5). The removal of the terraces and the levelling influenced the ECa, since subsoil original materials (Tertiary marls with a variable content of salts) were put on the top layer, breaking the continuity of the original soils. This is what occurred in zones with high EC25 and low NDVI (Fig. 5 and supplementary material). However, in the low EC25 and high NDVI zones the subsoil material put on surface were calcareous gravels, which provide better drainage conditions than the marls and lower salts content. In these zones, the local supply of water and fertilizers through the drip irrigation system would be providing the conditions for high tree vigour. Nevertheless, it could influence that there was not a good correspondence of EC25 and NDVI zones, as shown in the results.

As regards the relationship between the soil properties and EC25 as derived from the readings of the Veris 3100 ECa surveyor, the MANOVA offers an added value with respect to either the linear relationship or the ANOVA (Uribeetxebarria et al., 2018). A separate analysis of each sampled soil property with respect to EC25 or EC25 classes (ANOVA) may lead to misleading and inconsistent results. In fact, ECa reflects the combined effect of soil properties as a whole, and the delimitation of areas within the plot based on ECa maps should be checked from a multivariate approach. In the present case, the results of the MANOVA are in line with the theoretical basis for the relationship between ECa and soil properties developed by Rhoades et al. (1999). The parallel DRC values of the discriminant functions for the shallow and deep EC25 (Table 5) show the contribution of EC1:5, indicating that soil salinity is governing a significant part of the ECa readings. In addition, soil depth, clay and silt contents are also significantly contributing properties, in agreement with results reported in previous studies (Sudduth et al., 2005, Pedrera-Parrilla et al., 2016). In this case, the MANOVA was fundamental to identify clay as a relevant contributing property to the measured ECa, in comparison to the simple linear correlation analysis, in which clay was not correlated to EC25. On the other hand, discriminant functions for the NDVI showed that textural classes silt and sand were behind the variation of the NDVI in the top layer, while the WHC at -1500kPa and silt did the same in the second layer.

The results confirmed the different behaviour of both types of variables (EC25 and NDVI) and suggest several possibilities of differential management in the orchard. In this respect, and although different authors have suggested the combination of ECa with spectral vegetation indices to help in the delineation of SSMZ (Panda et al., 2010; De Benedetto et al., 2013; Ortega-Blu and Molina-Roco, 2016), in the present case, and because of the lack of relationship between EC25 and NDVI, we propose two strategies. One of them would be delineating SSMZ according to the combined EC25 classes, which would mainly serve to increase the irrigation doses in the high EC25 zones to reduce the salts content in the root zone and to enlarge the dimensions of the wet bulb. At present, this recommendation would not be easy to implement because the irrigation system consists of only one sector, since it was designed without having into account the soil spatial variability. Nevertheless, it would be possible to actuate by increasing the number of emitters per tree in those zones of high EC25 zones.

The second strategy would be delineating SSMZ according to NDVI classes, which would serve as a reference to regulate the tree vigour and yield through different managements actions such as pruning, application of growth regulators or fruit thinning.

5. CONCLUSIONS

The present work is a contribution to the application of precision agriculture (PA) techniques in fruticulture (precision fruticulture, PF), which are not so extensively used as in arable crops. Specifically, PA and PF can help in establishing optimized management actions in those orchards where land transformations have occurred.

The results of soil sampling and ECa survey showed that land transformation carried out in the 1980 decade to enlarge fields could have altered the spatial distribution and continuity of soil properties. In this respect, although a relationship between apparent electrical conductivity and peach tree vigour could be expected, it was not found, even in the case of trees planted in soils with salts content above the tolerance threshold. This could be due to the drip irrigation system used in the orchard, which keeps the trees free of high salt contents in the root-explored region.

Adopting PA and PF strategies may be appropriate to manage the orchard according to SSMZ. In the present case, two management zones delineation strategies were proposed depending on the final objective of the action: a) zones delineated according to the combined EC25 classes, mainly addressed to leach salts in the high EC25 zone, and b) zones delineated according NDVI classes to regulate tree vigour and yield. These strategies respond to the alteration of the original soil functions due to the land transformation carried out in previous years.

Acknowledgments

This work was funded by the Spanish Ministry of Economy and Competitiveness through the AgVANCE Project (AGL2013-48297-C2-2-R). We are also grateful to Frutas Hermanos Espax SL for the possibility to carry out the research in its farm and to the University of Lleida for funding the PhD scholarship of the first author.

References

Aggelopoulou, K., Wulfsohn, D., Fountas, S., Nanos, G., Gemtos, T., Blackmore, S., 2010. Spatial variability of yield and quality in an apple orchard. *Precis. Agric*. 11, 538–556.

Aggelopoulou, K., Fountas, S., Pateras, D., Nanos, G., Gemtos, T., 2011. Soil spatial variability and site-specific fertilization maps in an apple orchard. *Precis. Agric.* 12, 118–129.

Arnó, J., Martínez-Casasnovas, J.A., Uribeetxebarria, A., Escolà, A., Rosell-Polo, J.R., 2017. Comparing efficiency of different sampling schemes to estimate yield and quality parameters in fruit orchards. Advances in Animal Biosciences: *Precision Agriculture* (ECPA). 8:2, 471–476.

Carlson, K.M., Curran, L.M., Asner, G.P., Pittman, A.M., Trigg, S.N., Adeney, J.M., 2013. Carbon emissions from forest conversion by Kalimantan oil palm plantations. *Nat. Clim. Change 3*, 283–287.

Corwin, D.L., Lesch, S.M., 2003. Application of Soil Electrical Conductivity to Precision Agriculture: Theory, Principles, and Guidelines. *Agron. J.* 95, 455-471.

Cots-Folch, R., Martínez-Casasnovas, J.A., Ramos, M.C., 2009. Agricultural trajectories in a mediterranean mountain region (Priorat, Ne Spain) as a consequence of vineyard conversion plans. *Land Degrad. Dev.* 20, 1-13.

De Benedetto, D., Castrignano, A., Diacono, M., Rinaldi, M., Ruggieri, S., Tamborrino, R., 2013. Field partition by proximal and remote sensing data fusion. *Biosyst. Eng.* 114, 372–383.

De Oliveira, S.N., De Carvalho Júnior, O.A., Gomes, R.A.T., Guimarães, R.F., McManus, C.M., 2017. Landscape-fragmentation change due to recent agricultural expansion in the Brazilian Savanna, Western Bahia, Brazil. *Reg. Environ. Change.*17, 411-423.

Faraway, J.J., 2014. Linear models with R. 2nd edition. London: Chapman and Hall/CRC, 286 pages.

Fulton, A., Schwankl, L., Lynn, K., Lampinen, B., Edstrom, J., Prichard, T., 2011. Using EM and VERIS technology to assess land suitability for orchard and vineyard development. *Irrigation Sci.* 29, 497-512.

García-Ruiz, J.M., 2010. The effects of land uses on soil erosion in Spain: A review. *Catena.* 81, 1-11.

González-Dugo, M.V., Zarco-Tejada, P.J., Nicolás, E., Nortes, P.A., Alarcón, J.J., Intrigliolo, D.S., Fereres, E., 2013. Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precis. Agric.* 14, 660–678.

Jensen, J.R., 1996. Introductory digital image processing: remote sensing perspective, 2nd ed., Prentice-Hall: Englewood Cliffs, NJ, USA.

Käthner, J., Zude-Sasse, M., 2015. Interaction of 3D soil electrical conductivity and generative growth in *Prunus domestica* L. *Eur. J. Hortic. Sci*. 80, 231–239.

Khan, F.S., Zaman, Q.U., Chang, Y.K., Farooque, A.A., Schumann, A.W., Madani, A., 2016. Estimation of the root zone depth above a gravel layer (in wild blueberry fields) using electromagnetic induction method. *Precis. Agric*. 17, 155-167.

Laudicina, V.A., Palazzolo, E., Piotrowska-Długosz, A., Badalucco, L., 2016. Soil profile dismantlement by land levelling and deep tillage damages soil functioning but not quality. *Appl. Soil. Ecol.* 107, 298-306.

Ma, R., McBratney, A., Whelan, B., Minasny, B., Short, M. 2011. Comparing temperature correction models for soil electrical conductivity measurement. *Precis. Agric.* 12, 55–66.

Martínez-Casasnovas, J.A., Ramos, M.C., 2009. Soil alteration due to erosion, ploughing and levelling of vineyards in north east Spain. *Soil Use Manage.* 25, 183-192.

Martínez-Casasnovas, J.A., Ramos, M.C., Cost-Folch, R., 2010a. Influence of the EU CAP on terrain morphology and vineyard cultivation in the Priorat region of NE Spain. *Land Use Policy.27*, 11-21.

Martínez-Casasnovas, J.A., Ramos, M.C., Espinal-Utgés, S., 2010b. Hillslope terracing effects on the spatial variability of plant development as assessed by NDVI in vineyards of the Priorat region (NE Spain). *Environ. Monit. Assess.* 163, 379–396.

Mertens, F.M., Pätzold, S., Welp, G., 2008. Spatial heterogeneity of soil properties and its mapping with apparent electrical conductivity. *J. Plant Nutr. Soil Sci.* 171, 146-154.

Moraes, M.C.P.D., Mello, K.D., Toppa, R.H., 2017. Protected areas and agricultural expansion: Biodiversity conservation versus economic growth in the Southeast of Brazil. *J. Environ. Manage.* 188, 73-84.

Moral, F.J., Terrón, J.M., Marques da Silva, J.R.M., 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil Till. Res.* 106**,** 335–343.

Nainggolan, D., de Vente, J., Boix-Fayos, C., Termansen, M., Hubacek, K., Reed, M.S., 2012. Afforestation, agricultural abandonment and intensification: Competing trajectories in semi-arid Mediterranean agro-ecosystems. *Agr. Ecosyst. Environ.* 159**,** 90–104.

Noori, O., Panda, S.S., 2016. Site-specific management of common olive: Remote sensing, geospatial, and advanced image processing applications. *Comput. Electron. Agric*. 127, 680-689.

Ortega-Blu, R., Molina-Roco, M., 2016. Evaluation of vegetation indices and apparent soil electrical conductivity for site-specific vineyard management in Chile. *Precis. Agric.* 17, 434–450. DOI 10.1007/s11119-016-9429-x

Öztekin, T., 2013. Short-term effects of land leveling on irrigation-related some soil properties in a clay loam soil*. Sci. World J.* 1-16.

Panda, S.S., Hoogenboom, G., Paz, J.O., 2010. Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: a review. *Remote Sens.* 2, 1973-7997.

Pedrera-Parrilla, A., Van De Vijver, E., Van Meirvenne, M., Espejo-Pérez, A.J., Giráldez, J.V., Vanderlinden, K., 2016. Apparent electrical conductivity measurements in an olive orchard under wet and dry soil conditions: significance for clay and soil water content mapping. *Precis. Agric.* 17, 531-545.

Peralta, N.R., Costa, J.R., 2013. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Comput. Electron. Agric.* 99, 218-226.

Peña-Barragán, J.M., Jurado-Expósito, M., López-Granados, F., Atenciano, S., Sánchez-de la Orden, M., García-Ferrer, A., García-Torres, L., 2004. Assessing land-use in olive groves from aerial photographs. *Agr. Ecosyst. Environ.* 103, 117-122.

Priori, S., Martini, E., Andrenelli, M.C., Magini, S., Agnelli, A.E., Bucelli, P., Biagi, M., Pellegrini, S., Costantini, E.A.C., 2013. Improving Wine Quality through Harvest Zoning and Combined Use of Remote and Soil Proximal Sensing. *Soil Sci. Soc. Am. J.* 77, 1338-1348.

Ramos M.C., Martínez-Casasnovas, J.A., 2010. Effects of field reorganisation on the spatial variability of runoff and erosion rates in vineyards of Northeastern Spain. *Land Degrad. Dev.* 21, 1-12.

Rhoades, J.D., Chanduvi, F., Lesch, S., 1999. Soil salinity assessment: methods and interpretation of electrical conductivity measurements. FAO Irrigation and Drainage Paper 57. FAO: Rome, Italy, 150 pp, ISSN 0254-5284

Richter, H.G., 1984. Land use and land transformation. *Geo. Journal.* 8, 67-74.

Rouse, J.W. Jr, Haas, R.H., Deering, D.W., Schell, J.A., Harlan, J.C., 1974. Monitoring the Vernal Advancement and Retrogradation (GreenWave Effect) of Natural Vegetation, NASA/GSFC Type III Final Report: Greenbelt, MD, USA, 371p.

Stassen, P., Wooldridge, J., 2011. Peach rootstocks for soil with moderately high pH and salinity. *SA Fruit J*. 34-38.

Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed., United States Department of Agriculture: Washington DC, USA.

Su, S., Zhou, X., Wan, C., Li, Y., Kong, W., 2016. Land use changes to cash crop plantations: crop types, multilevel determinants and policy implications. *Land Use Policy* 50, 379-389.

Sudduth, K.A., Kitchen, N.R., Wiebold, W.J., Batchelor, W.D., Bollero, G.A., Bullock, D.G., Clay, D.E., Palm, H.L., Pierce, F.J., Schuler, T.T., Thelen, K.D., 2005. Relating apparent electrical conductivity to soil properties across the north-central USA. *Comput. Electron. Agric.* 46, 263–283.

Tanji, K.K., Kielen, N.C., 2002. Agricultural drainage water management in arid and semi-arid areas. FAO Irrigation and Drainage paper 61. FAO: Rome, Italy, 188 pp, ISSN 0254-5284.

Taylor, J.A., McBratney, A.B., Whelan, B.M., 2007. Establishing management classes for broadacre grain production. *Agron. J.* 99, 1366–1376.

Uribeetxebarria, A., Arnó, J., Escolà, A., Martínez-Casasnovas, J.A., 2018. Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. *Geoderma.* 319, 185-193.

Wallace, J.F., Canci, M., Wu, X., Baddeley, A., 2008 Monitoring native vegetation on an urban groundwater supply mound using Airborne Digital Imagery. *J. Spat. Sci.* 53, 63–73.

Xiao, R., Su, S., Mai, G., Zhang, Z., Yang, C., 2015. Quantifying determinants of cash crop expansion and their relative effects using logistic regression modeling and variance partitioning. *Int. J. Appl. Earth Obs. Geoinf.*34, 258-263.

Ye, X., Sakai, K., Sasao, A., Asada, S., 2009. Estimation of citrus yield from canopy spectral features determined by airborne hyperspectral imagery. *Int. J. Remote Sens.* 30, 4621-4642.

Yi, Z., Cannon, C.H., Chen, J., Ye, C., Swetnam, R.D., 2014. Developing indicators ofeconomic value and biodiversity loss for rubber plantations in Xishuangbanna, southwest China: a case study from Menglun township. *Ecol. Indic.* 36, 788–797.

Zarco-Tejada, P.J., González-Dugo, M.V., Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. *Remote Sens. Environ.* 179, 89–103.

Zude-Sasse, M., Fountas, S., Gemtos, T.A., Abu-Khalaf, N., 2016. Applications of precision agriculture in horticultural crops. *Eur. J. Hortic. Sci.* 81, 78–90.

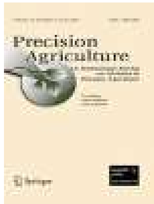
Chapter 5

Stratified sampling in fruit a orchard

This chapter contains the following published paper in the journal *Precision Agriculture.*
JCR-SCI Impact Factor: 2.435. Category: Agriculture, multidisciplinary; 1st Quartile

Uribeetxebarria, A., Martínez-Casasnovas, J.A., Escolà, A., Rosell-Polo, J.R., Arnó, J. 2019. Stratified sampling in fruit orchards using cluster-based ancillary information maps: a comparative analysis to improve yield and quality estimates. *Precis. Agric.* 20, 179-192.

Stratified sampling in fruit orchards using cluster-based ancillary information maps: a comparative analysis to improve yield and quality estimates



Asier Uribeetxebarria, José A. Martínez-Casasnovas, Alexandre Escolà, Joan R. Rosell-Polo, Jaume Arnó

ABSTRACT

Estimation of yield or other fruit quality parameter is of great interest to farmers, technicians and agricultural cooperatives to decide on management actions just before harvesting and, in any case, to anticipate and plan harvesting operations. Making accurate and reliable estimates often requires systematic sampling that, when covering the whole plot, can result in the use of a large number of samples and a significant effort in time and cost for fruit growers. Faced with this whole area sampling strategy, simple random sampling (SRS) using reduced sample sizes is currently a widely used technique despite the less precise estimates that it provides. In this work, different stratified sampling schemes have been tested to estimate yield (kg/tree), fruit firmness (kg/cm^2) and the refractometric index (ºBaumé) in a peach orchard located in Gimenells (Lleida, Catalonia, Spain). In contrast to SRS, the use of ancillary information (NDVI and apparent electrical conductivity, ECa) allowed sampling units or trees to be stratified according to two or three classes (strata) within the plot. The classes or homogeneous stratification zones were delimited by cluster analysis using, either separately or in combination, a multispectral airborne image (NDVI) and a ECa survey map acquired by means of a soil resistivity sensor (Veris 3100). Sampling schemes were then compared in terms of efficiency. In general, stratified sampling showed better results than SRS. Regarding yield estimates, stratified sampling according to two strata of NDVI allowed the sample size to be reduced by 17 % compared to the SRS for the same precision. On the other hand, quality parameters may require different stratification strategies concerning the number of strata to be used. While ºBaumé was better-estimated using also stratified samples based on two strata of NDVI, fruit firmness showed better results when stratifying by three classes or strata of NDVI. In any case, neither the ECa nor the combined use of NDVI + ECa has improved sampling efficiency when used as ancillary maps for stratification.

KEY WORDS

Sampling efficiency, Fruit yield and quality, Peach, NDVI, Apparent electrical conductivity

1. INTRODUCTION

Sampling to estimate yield and/or fruit quality at harvest time is of great interest in fruit growing. However, reliable prediction of these parameters is not easy, especially when systematic sampling is usually replaced by a less complex simple random sampling (SRS) to reduce time and cost. In other occasions, random sampling raises doubts to both growers and advisors about how many trees should be sampled and, above all, which specific ones should be sampled within a plot. Facing this situation, there is a need to develop new and more precise methods with acceptable costs to guide fruit growers during field sampling. SRS is a widely used design, because it is relatively simple to implement by random selection of sampling units (trees) within the plot. However, SRS is inefficient when estimating parameters that show spatial autocorrelation within the plots (Webster & Lark, 2013). Taylor et al. (2005) and Kazmierski et al. (2011) showed that vineyards are spatially variable and that grape yield usually follows well-defined and consistent spatial patterns over time. This same situation can be expected in fruit orchards and, for this reason, sampling methods that take into account the different areas within the plot with different expected yield values would be preferable to optimally locate sampling trees to obtain better yield estimates.

On the other hand, fruit growers can hire service companies that provide crop vigour and/or soil apparent electrical conductivity (ECa) maps obtained with suitable sensors (proximal and remote sensing). Normalized difference vegetation index (NDVI) derived from airborne images were used by Meyers and Vanden Heuvel, (2014) to optimize sampling protocols in vineyard and reduce sample sizes. Applying a heuristic optimization algorithm (Tabu Search Algorithm) to NDVI images, specific samples to conform the spatial distribution of NDVI within the plot can be established (Meyers and Vanden Heuvel, 2014). As NDVI is related to vine vigour, the method is a way for distributing sampling units by covering the areas of different vigour to capture vineyard canopy variability within the plot. This idea is also behind the method proposed by Carrillo et al. (2016) to improve grape yield estimates. The authors concluded with the need to consider a two-step sampling method combining NDVI-based samples with random vine samples to predict specific components of the productive potential in a vineyard. Regarding apparent electrical conductivity (ECa), there are several studies that address the use of ECa classified maps for site-specific management practices (Moral et al., 2010; Peralta & Costa, 2013). The suitability of this information in fruit-growing sampling is a pending issue, although soil characteristics are expected to impact yield and/or quality parameters.

There are few studies on sampling in fruit orchards. Monestiez et al. (1990) proposed a geostatistical approach to assess spatial dependence between fruits to choose the most appropriate sampling designs inside the tree structure. Multilevel systematic sampling can also be an interesting option to estimate the number of fruits for yield forecasts (Wulfsohn et al., 2012), obtaining error coefficients of only 10 %. More recently, sampling stratification using NDVI-based aerial images allowed different areas to be better delimited for sampling in nectarine orchards (Miranda et al., 2015), with a significant reduction in sample size (20-35 %) compared to random sampling (Miranda et al., 2018). As is known, SRS can produce local clusters of trees and leave unrepresented areas within a plot (Webster & Lark, 2013). Alternatively, farmers can consider using NDVI or ECa data to stratify samples, assuming that yield and quality parameters in orchards often present spatial autocorrelation and, what is more important, possible spatial cross-correlation with ancillary variables supplied by proximal and remote sensors of increasingly common use in agriculture. Cross-correlogram is a powerful tool to test the spatial correlation between two variables, and checking this spatial correlation may be the key factor before stratifying the samples.

The aim of this study was to investigate how the use of ancillary data (NDVI and ECa) in stratified sampling schemes can improve sampling efficiency compared to a SRS of equal size for the whole of a plot. Efforts in time and cost could be reduced with this new sampling strategy by optimizing sample sizes through the application of technological advances in the framework of precision agriculture. Sampling in orchards is then proposed as a design-based sampling strategy, making use of classical sampling theory (that is, assuming normality and independence of observations). This may be a limitation in plots with spatial autocorrelation. However, the use of geostatistical methods is beyond the scope of this paper.

2. MATERIALS AND METHODS

2.1. Study plot

The research was conducted in a peach orchard (*Prunus persica* cv. 'Platycarpa') located at the IRTA Experimental Station (41° 39' 19" N, 0° 23' 36" E, ETRS89) in Gimenells (Lleida, Catalonia, Spain). The plot covered an area of 0.65 ha, and was planted in 2011 according to a 5 x 2.80 m pattern (Fig. 1). Soil was classified as Petrocalcic Calcixerept (Soil Survey Staff, 2014), and it was a well-drained soil without salinity problems. The presence of a petrocalcic horizon at a variable depth (0.4-0.8 m) and high $CaCO_3$ content were the main soil limiting factors. The horizon may be at shallow depth due to successive earth movements and tillage operations that, over time and since 1946, have contributed to modify in shape and size of the parcelling in the farm. The climate is typical of hot semi-arid areas, with strong seasonal temperature variations (cold winters and hot summers). Annual precipitation is frequently below 400 mm, and basically distributed from September to May.

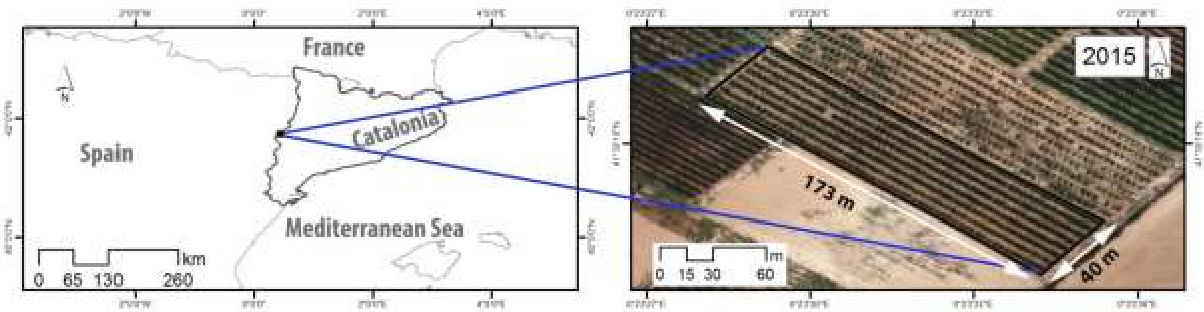


Fig. 1 Location of the study area (left), and orthophoto of the peach orchard plot in 2015 (right).

2.2. Sample size and stratification

Three production and quality variables were sampled within the plot: yield (kg/tree), fruit firmness (kg/cm^2) and refractometric index (ºBaumé). To determine the sample size, an aerial multi-spectral image was taken on June 9th, 2015. The image resolution was 0.25 m/pixel. Once the canopies were individually delimited on the basis of this image (ESRI® ArcMap^{TM} 10.4.1) to obtain a map of georeferenced trees within the plot (statistical population), a weighted average value of NDVI according to the area of the canopy was assigned to each tree. These tree-averaged NDVI values were then used as base data for determining the sample size for a SRS without replacement using Eq. 1:

$$n = \frac{\zeta^2_{\alpha/2} CV^2}{E^2_R} \quad (1)$$

Where n is the sample size (number of trees) assuming sample independence, $\zeta_{\alpha/2}$ (1.96) is the value of the standard normal variate for a 95 % confidence (α = 0.05), CV is the Coefficient of Variation (17.5 % in the present case), and E_R is the relative error assumed (10 %). The result of Eq. 1 was 12 sampling trees that were first randomly distributed within the plot (samplingscheme A, Fig. 2). Apart from being a usual index for detecting spatial variability in tree crops (Kazmierski et al., 2011), the use of NDVI for this approach was justified because previous successful applications in fruit sampling were known (Miranda et al., 2015, 2018).

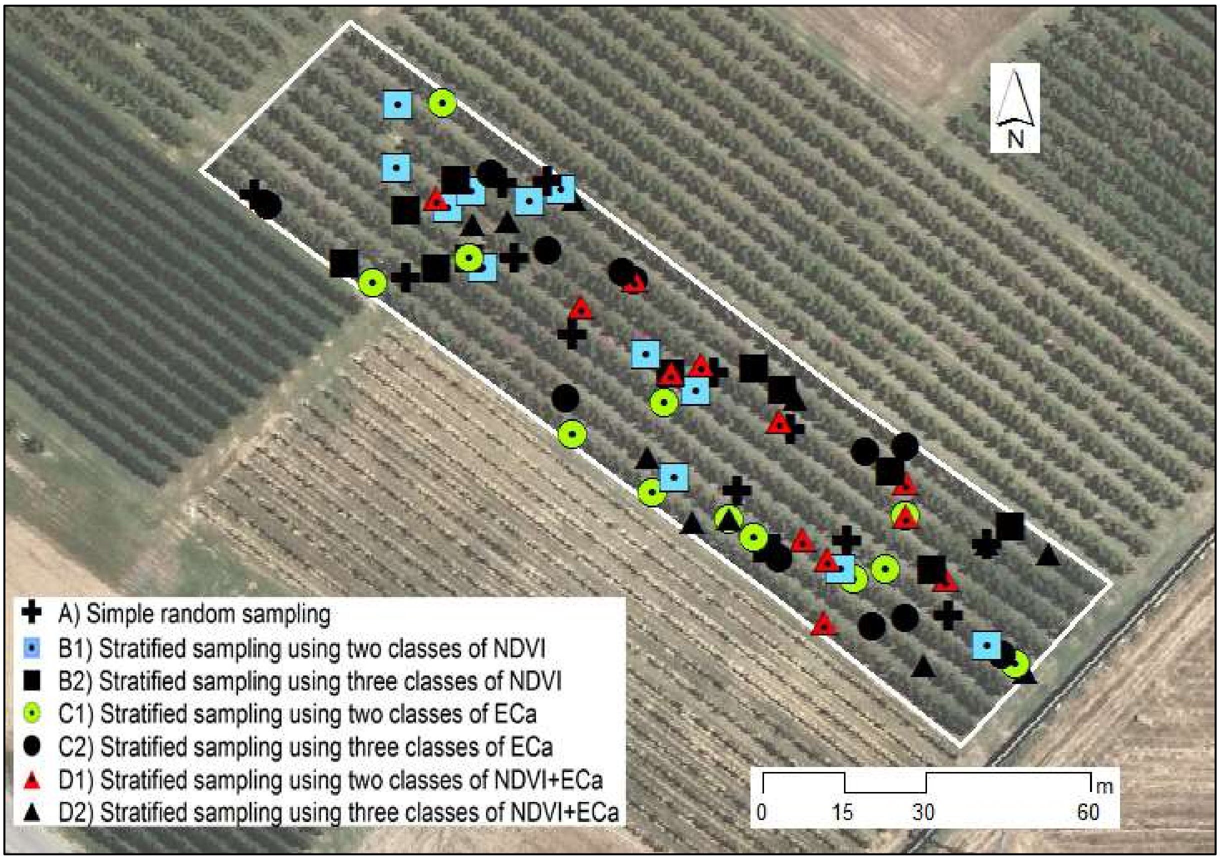


Fig. 2 Sampling units (trees) corresponding to seven different sampling schemes.

Additional schemes were tested in which new sampling trees (twelve in each case) were first obtained by stratified random sampling according to two and three classes of NDVI. Specifically, NDVI classified maps were built by clustering interpolated NDVI values (NDVI raster map) using the unsupervised classification algorithm ISODATA (Jensen, 1996). The process on which this algorithm is based is well known. Assigning an arbitrary mean to each class, pixels were then successively reassigned minimizing the Euclidean distance from each pixel to the mean value of the class. Each iteration, class means were recalculated and pixels were reallocated until the last iteration is reached, or the number of pixels that change from one class to another does not exceed a certain threshold (Guastaferro et al., 2010). The same strategy (stratified sampling based on clustered maps) was repeated using the information provided by a Veris 3100 ECa surveyor. As a widely used sensor for soil characterization (Sudduth et al., 2005), the information provided may be very useful in sampling given the soil-tree interaction. This sensor measured the ECa at two soil depths: shallow (0-0.3 m) and deep (0-0.9 m). Both ECa value layers were interpolated by ordinary kriging, and ECa classes were established based on the cluster analysis of the two maps (shallow and deep) simultaneously. Finally, the same procedure was repeated again by taking all three ancillary layers (NDVI, shallow ECa and deep ECa). In short, seven sampling schemes (including scheme A) were compared to each other based on a total number of 84 sampled trees (7x12) within the plot (Fig. 2). Figure 3 shows five of the proposed sampling schemes, (i) SRS (scheme A), (ii) stratified sampling based on two classes of NDVI (scheme B1), (iii) stratified sampling based on three classes of NDVI (scheme B2), (iv) stratified sampling based on two classes of ECa (scheme C1), and (v) stratified sampling based on three classes of ECa (scheme C2). Schemes that use both information layers (schemes D1 and D2) are not shown. In each case, sampling trees within each stratum were randomly sampled without replacement.

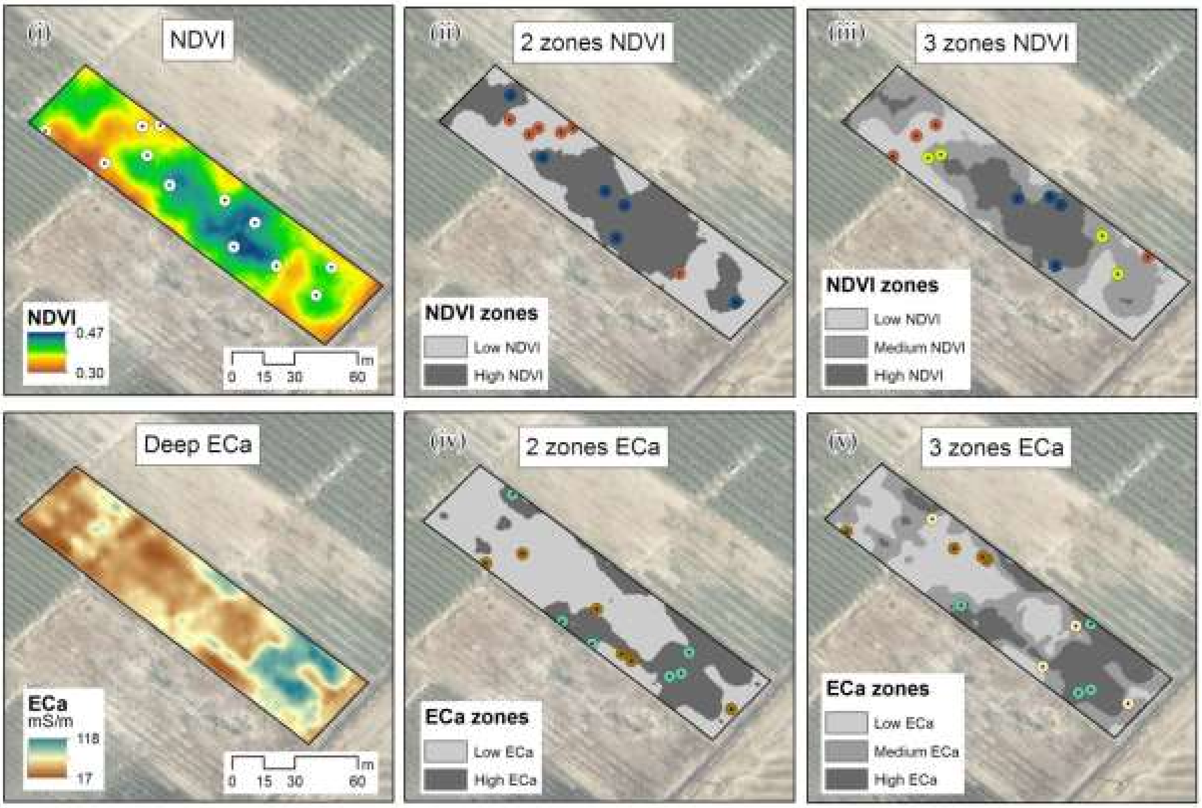


Fig. 3 Sampling schemes: (i) simple random sampling, (ii) stratified sampling by NDVI (two strata), (iii) stratified sampling by NDVI (three strata), (iv) stratified sampling by ECa (two strata), (v) stratified sampling by ECa (three strata).

2.3. Estimation in stratified sampling schemes

In a SRS approach, the sample mean ($\overline{z}_{SRS}$) has proven to be an unbiased estimator of the population mean (μ), with a variance that can be estimated by $v(\overline{z}_{SRS}) = \frac{s^2}{n} \times (1-f) = \frac{s^2}{n} \times \left(1 - \frac{n}{N}\right)$ where s^2 is the sample variance, and $1 - f = 1 - \frac{n}{N}$ is the finite population correction or fpc, where *n* is the sample size and *N* the size of the population (459 trees in the plot under study). As the interest was to work with small size samples, confidence limits for the mean can be formulated as $\overline{z}_{SRS} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{1-f}$, where $\overline{z}_{SRS}$ is the sample mean, $\frac{s}{\sqrt{n}} \sqrt{1-f}$ is the standard error of the mean, and $t_{\alpha/2}$ is the Student's *t* value corresponding to *n*-1 degrees of freedom for a 95 % confidence.

In order to sample more efficiently, other sampling schemes were used by stratifying the 12 sampling trees according to two strata (6 trees per stratum) or three strata (4 trees per stratum). As a reminder, strata corresponded to the classes obtained after classification of the plot according to NDVI, ECa or both auxiliary data layers. The different stratifications produced classes that were not equal in area (therefore, with different number of trees per stratum), and so the plot mean (μ) was then estimated for *K* classes (strata) within the plot using a weighted average as suggested by Cochran, (1977), and more recently by Webster & Lark,(2013)in what is called regional classification techniques:

$$\overline{z}_{StRS} = \sum_{k=1}^{K} w_k \times \overline{z}_k \tag{2}$$

where $\overline{z}_k$ is the sample mean of the *k*th class, and w_k allowed the number of individuals (trees) of the *k*th class to be weighted using Eq. 3,

$$w_k = \frac{N_k}{N} \tag{3}$$

where N_k is the number of trees within stratum *k*, and *N* is the total number within the plot.

As in SRS, confidence limits were obtained using the standard error of the mean, in this case, the square root of the estimated variance (Cochran, 1977):

$$v(\overline{z}_{StRS}) = \sum_{k=1}^{K} \frac{w_k^2 s_k^2}{n_k} \times (1 - f_k) \tag{4}$$

where s_k^2 is the within-class sample variance of the *k*th stratum, n_k is the sampling trees within the stratum (6 or 4), and $1 - f_k$ is the fpc for the *k*th stratum calculated as $1 - f_k = 1 - \frac{n_k}{N_k}$. Finally, the value $t_{\alpha/2}$was adjusted for each stratified sampling scheme according to an effective number of degrees of freedom as established by Cochran (1977) in these cases.

The above confidence intervals were obtained assuming normality of observations. Since this hypothesis was not tested (for example, using Shapiro-Wilk test), additional intervals were calculated by applying a bootstrap estimation with the aim of contrasting the results. Bootstrap is a method of resampling to obtain approximately the precision of an estimator without hypothesizing about its distribution. Thus, for each set of 12 sampling trees corresponding to the different sampling schemes (which are now the statistical population), sampling is done with replacement until obtaining 1000 sample arrangements each of equal size 12. By averaging the 12 values in each new sample, it is known that 1-α level confidence intervals can be obtained from the distribution of the 1000 calculated mean values through the use of the percentile method (Efron, 1982). Specifically, confidence limits were established excluding the $1000\,\alpha/2$ values located at the extreme positions of the distribution (α = 0.05). In all cases, sample arrangement generation was performed by programming in R software, version 3.3.2.

2.4. Sampling efficiency

The most interesting sampling scheme is that which provides, on average, the least mean squared error (*MSE*). Since the seven sample means were unbiased estimators of the plot mean, *MSE* can be used as a measure of accuracy. Coinciding *MSE* with the variance (Eq. 5), efficiency to estimate the plot or population mean (μ) can be established as the inverse of the estimated variance of the sample mean.

$$MSE(\overline{z}) = [bias(\overline{z})]^2 + v(\overline{z}) = v(\overline{z}) \tag{5}$$

To compare any of the stratified sampling schemes ($\overline{z}_{StRS}$) with respect to the simple random sampling design ($\overline{z}_{SRS}$), the relative efficiency (*RE*) was obtained as shown in Eq. 6:

$$RE = \frac{Efficiency(\overline{z}_{StRS})}{Efficiency(\overline{z}_{SRS})} = \frac{v(\overline{z}_{SRS})}{v(\overline{z}_{StRS})} \tag{6}$$

where $v(\overline{z}_{StRS})$ was in each case the variance of the stratified sample mean, and $v(\overline{z}_{SRS})$or variance of a random sample mean of the same size (taken as reference) was best estimated by applying the method suggested by Cochran (1977). Specifically, given the results of a stratified random sample, an unbiased estimator of the variance of the mean for a simple random sample from the same population is (Eq. 7),

$$v(\overline{z}_{SRS}) = \frac{(N-n)}{n(N-1)}\left[\frac{1}{N}\sum_k^K \frac{N_k}{n_k}\sum_i^{n_k} z_{ki}^2 - \overline{z}_{StRS}^2 + v(\overline{z}_{StRS})\right] \tag{7}$$

where z_{ki} were the values sampled at trees within stratum *k*. The other parameters are those stated in previous paragraphs. By averaging the six previously calculated variances (one for each stratified sample) with the variance previously obtained for scheme A (SRS), the resulting variance, $v(\overline{z}_{SRS})$, was the one used in the calculation of the *RE*. The reason for using Eq. 7 was the use of non-proportional allocation of sampling trees, that is, the same number of sampling units (6 or 4) was assigned regardless of the size (number of trees) of each stratum.

Both the *MSE* and the *RE* were the statistics that served for the comparison of the different sampling schemes and, above all, for the verification of the possible gain due to stratification. Knowing the *RE* allowed the necessary sample size for the same precision to be compared between sampling schemes. Low *MSE* valuesand values of *RE* greater than 1 are those sought for stratified sampling schemes.

2.5. An estimate of the population mean

Considering the spatial distribution of the 84 sampling trees resulting from the seven sampling schemes (7x12) (Fig. 2), it is important to emphasize that only 5 % of the plot area resulted in a weak sampling density, i. e. with sampling units separated from each other by a distance larger than 9.78 m (range of the NDVI exponential variogram, not shown). So, the sampled information contained in these 84 trees was finally considered to estimate the mean of the plot as accurately as possible by calculating a weighted average of the means of the samples. The most accurate linear combination of the seven independent sample means was obtained by assigning proportionally greater weighting to the more precise (Eq. 8),

$$\overline{z}_w = \sum a_i \times \overline{z}_i \quad (8)$$

where $\overline{z}_w$ is the weighted average for the plot, $\overline{z}_i$ are the sample means calculated for each of the seven sampling schemes, anda_i are the relative weights calculated using the inverse of the variance of the sample means (Eq. 9):

$$a_i = \frac{1/v(\overline{z}_i)}{\sum 1/v(\overline{z}_i)} \quad (9)$$

2.6. Goodness of stratification

Finally, and as already mentioned, stratified sampling schemes were based on a previous classification of the plot. A more accurate and efficient estimation of the mean was linked to the ability of the NDVI and/or ECa auxiliary layers to discriminate different mean values between classes, while the values within the classes have lower intra-class variability compared to the total variability of the plot. A parameter that served to judge the goodness of these classifications was the relative variance ($r_V = s_W^2/s_T^2$), where s_W^2 was the pooled or average within-class variance, and s_T^2 was the total variance in the sample (Webster & Lark, 2013). Used in the form of its complement (Eq. 10),

$$1 - (s_W^2/s_T^2) = 1 - r_V \quad (10)$$

it allowed values close to 1 to be obtained for those more effective sampling schemes. Values close to 0 or even negative corresponded to non-effective stratifications.

2.7. Spatial cross-correlation

To check stratified sampling results using ancillary variables, bivariate Moran's coefficient was also calculated to assess the spatial cross-correlation between ancillary information layers (NDVI and ECa) and the sampled yield and quality variables (GeoDa 1.12 software, Anselin et al., 2010). Hypothetically, the most efficient stratified sampling schemes would be those with significant spatial correlation with the variables to be sampled. Having verified significant spatial autocorrelation for the three variables of interest (Moran's *I* coefficient on the total of 84 sampled trees, data not shown), assessing spatial cross-correlation between ancillary variables and sampled variables could report information (even if a posteriori) on what ancillary information was most convenient in each case. However, it must be said that the use of geostatistical methods was beyond the scope of this paper. So, classical sampling theory was prevalent to assess stratified methods in this work under what is called design-based sampling strategies (Brus & de Gruijter, 1997).

3. RESULTS AND DISCUSSION

Table 1 shows the mean squared error (*MSE*) and the relative efficiency (*RE*) for the different sampling schemes tested. For each of the variables (yield, fruit firmness and refractometric index), confidence intervals (CIs) for the population mean (μ) are also shown. Two types of confidence intervals were built for each sampling scheme as a result of using, i) the standard error of the corresponding sample mean (parametric approach) or ii) the non-parametric bootstrap approach. In the same Table 1, the weighted average of the plot $\bar{z}_w$for each field variable is added next to the sample means. By completing this table of results, each sampling scheme is valued according to the goodness of stratification using the value 1 minus the relative variance.

Concerning the confidence intervals for the mean, bootstrap CIs were always slightly narrower compared to CIs based on the normality of the sample means. This may be due to the asymptotic approximation of the bootstrap method and, in any case, could prove the non-normality of the distributions. However, and for comparison purposes, relative efficiency (*RE,* Table 1) based on the estimated variances of the sample means (Eq.6) was the statistic taken as a reference instead of the CIs. As general results, stratified sampling seemed to improve efficiency (*RE*) compared to SRS, mostly for the quality variables (fruit firmness and refractometric index). The improvement in yield estimation efficiency using stratified sampling was lower than for quality variables, and it was only evident in very particular cases of stratification. To aid interpretation, a more detailed analysis of the results in Table 1 is addressed in the following sections.

3.1. Sampling to estimate yield

Compared to the other sampling schemes, stratified sampling based on two classes of NDVI (scheme B1) was the one that showed the best results in estimating yield, with an expected average error $\left(\sqrt{MSE}\right)$ of 2.71 kg/tree (Table1). Surprisingly, when stratifying the sample in three NDVI classes (scheme B2), the method failed to improve the efficiency or precision compared to SRS. This result could be explained by the poor effectiveness of the stratification (negative value of $1 - r_V$). In fact, negative values of the goodness of the stratification have always been obtained in those inefficient schemes with *RE* less than 1.

Concerning the use of ECa as ancillary information, stratifying the sample according to three classes (strata) of soil conductivity (scheme C2) has also shown better efficiency results than SRS. However, the error (*MSE*) and relative efficiency (*RE*) are not as good as in scheme B1 (stratification according to two classes of NDVI). Again and unexpectedly, the stratified sampling has shown better efficiency than SRS despite the poor result of the goodness of the stratification (positive but very low value of $1 - r_V$, and very far from the optimal values close to 1).

Table 1 Efficiency parameters for the sampling schemes tested.

Sampling scheme	Mean ($\overline{z}$)	$(MSE)^{1/2}$	CI_L	CI_U	CI_{LB}	CI_{UB}	*RE*	$1 - r_V$
Yield (kg/tree)								
Weighted average of the plot	24.49							
A	26.36	2.32	21.26	31.47	22.41	30.73		0.00
B1	24.33	2.71	17.93	30.73	19.65	30.09	1.20	0.04
B2	24.29	3.65	15.37	33.21	18.28	30.58	0.66	-0.10
C1	23.57	3.70	14.83	32.31	16.90	29.43	0.64	-0.09
C2	24.58	2.82	17.90	31.26	19.11	30.20	1.10	0.07
D1	22.09	3.30	14.29	29.89	16.32	27.00	0.81	-0.08
D2	24.21	2.87	16.23	32.20	18.84	29.18	1.06	0.08
Fruit firmness (kg/cm^2)								
Weighted average of the plot	4.33							
A	4.10	0.30	3.44	4.76	3.51	4.60		0.00
B1	4.31	0.22	3.82	4.81	3.87	4.76	1.80	0.13
B2	4.26	0.17	3.88	4.65	3.89	4.67	3.09	0.28
C1	4.27	0.35	3.45	5.08	3.56	4.79	0.69	-0.08
C2	4.75	0.27	4.13	5.37	4.28	5.22	1.18	-0.19
D1	4.40	0.33	3.66	5.14	3.40	4.92	0.80	0.19
D2	4.28	0.29	3.36	5.21	3.71	4.83	1.01	0.12
Refractometric index (ºBaumé)								
Weighted average of the plot	6.86							
A	7.02	0.14	6.71	7.32	6.80	7.31		0.00
B1	6.55	0.10	6.32	6.78	6.34	6.74	3.77	0.10
B2	6.63	0.19	6.10	7.16	5.90	7.09	1.09	0.54
C1	7.22	0.17	6.81	7.63	6.99	7.54	1.40	-0.09
C2	6.88	0.10	6.64	7.12	6.70	7.12	4.04	0.21
D1	6.80	0.32	5.99	7.61	6.37	7.32	0.39	-0.05
D2	7.31	0.17	6.86	7.76	6.88	7.66	1.29	0.38

A (Simple random sampling); B1 and B2 (NDVI stratified sampling, 2 and 3 classes); C1 and C2 (ECa stratified sampling, 2 and 3 classes); D1 and D2 (combined NDVI + ECa stratified sampling, 2 and 3 classes). *MSE* (Mean Squared Error), CI_L and CI_U (lower and upper confidence interval considering normality), CI_{LB} and CI_{UB} (lower and upper CI using bootstrap), *RE* (relative efficiency), r_V (relative variance).

The choice between using scheme B1 (stratifying by using the NDVI) and scheme C2 (stratifying by using the ECa) is not easy. An analysis of the special characteristics of the plot can help to understand the sampling results for later decision-making. In the plot under study, affected by the presence of a petrocalcic horizon and high $CaCO_3$ content, some advantage was expected by stratifying the sample using a classified map of the ECa. In fact, significant inverse spatial cross-correlation was obtained between yield and ECa using the bivariate Moran's I_B statistic (Table 2). As high $CaCO_3$ content is a limiting factor of yield, with high ECa values usually associated with low yields (Martínez-Casasnovas et al., 2012; Ortega-Blue & Molina-Roco, 2016; Uribeetxebarria et al., 2018), the spatial variation of ECa could make it advisable to stratify on the basis of this information layer instead of using an NDVI map. However, given the also significant spatial correlation between NDVI and yield (Table 2), the best efficiency results, and the simplicity in managing the stratification in only two strata, the B1 scheme is the option to recommend. In fact, NDVI has been used successfully to guide sampling for yield forecasting tasks in many crops (Fortes et al., 2015; Miranda & Royo, 2003; Taylor et al., 2010).

From a practical point of view, as scheme B1 was more efficient (*RE* out of 1.20, Table 1), a similar efficiency for the SRS (scheme A) could be reached using a smaller sample size, theoretically equal to *n*(SRS)/*RE* (12/1.20). In short, stratified sampling according to two strata of NDVI allowed the sample size to be reduced by 17 % compared to the SRS for the same precision.

Table 2 Spatial cross-correlation between NDVI and ECa ancillary information layers and the sampled yield and quality variables

Ancillary information	Sampled fruit variable	Bivariate Moran's I_B coefficient*	Pseudo p-value**
NDVI	Yield	-0.147	0.011
NDVI	Fruit firmness	0.199	0.004
NDVI	Refractometric index	-0.215	0.002
ECa	Yield	-0.315	0.001
ECa	Fruit firmness	-0.059	0.173
ECa	Refractometric index	0.029	0.306

*Global spatial statistic to estimate the spatial cross-correlation between ancillary and sampled variables. Correlation calculated based on 84 sampling trees using GeoDa 1.12 software (Anselin et al., 2010).
**Significance test was based on 999 permutations to generate the reference distribution under the null hypothesis of spatial randomness. The observed statistic was then compared to this distribution to calculate a so-called pseudo p-value (0.001 is the most extreme pseudo p-value under this scenario).

3.2. Sampling to estimate fruit quality parameters

Stratified sampling schemes worked differently when estimating fruit quality parameters. Regarding fruit firmness (Table 1), scheme B2 was clearly better in both *MSE* and efficiency (*RE* greater than the other sampling schemes). A significant spatial cross-correlation between NDVI and firmness (the greater the NDVI, the greater the firmness) could explain this result (Table 2). Likewise, stratifying sampling trees based on three strata of NDVI allowed spatial classification in fruit firmness to be more effective ($1 - r_V$ = 0.28). Concerning the sugar content of the fruit (refractomeric index), the results were somewhat difficult to interpret. Again, NDVI correlated spatially in a significant way (Table 2), showing an inverse relationship. However, among the two proposed schemes (B1 and B2), it was the B1 scheme (stratification through 2 strata) that achieved the best efficiency (*RE* value of 3.77). On the other hand, this result was somewhat inconsistent with the effectiveness of the stratification. Sampling trees were optimally classified using three classes of NDVI as suggested by the goodness of stratification ($1 - r_V$ in Table 1). Therefore, there was a discrepancy in the efficiency scores between schemes B1 and B2, and reasonable doubts arise as to whether to use two or three strata to stratify the samples. This situation can occur because the sampling trees can be well segmented (reducing the average within-class variance) and, however, presenting a variance of the sample mean too high (poor value of the *RE*). Since the fundamental criterion sought is to increase the *RE*, the B1 scheme would be the recommended option in this case by making compatible the values of *RE* (Table 1) and spatial correlation (Table 2).

The relationship between NDVI and some quality parameters has been shown in other studies (Zude-Sasse et al., 2016; Martínez-Casasnovas et al., 2012). Many times, fruits achieve lower sugar content (ºBaumé) in the areas with the highest NDVI values. Inversely, vigorous canopies with high amount of leaves (and higher NDVI values) can shade the fruits affecting fruit ripening and, as happens in viticulture (Vanden Heuvel et al., 2002), producing greener fruits with higher firmness values. This would explain the significant spatial relationship between NDVI, fruit firmness and ºBaumé within the plot (Table 2).

Regarding the use of ECa as ancillary information to stratify the quality, specifically ºBaumé in fruit, the results have been contradictory. While sampling scheme C2 (three strata of ECa) has shown the highest relative efficiency (*RE* of 4.04) together with good goodness of stratification, spatial cross-correlation between both parameters (ECa and ºBaumé) was not significant (Table 2). Since spatial correlation is an essential requirement to justify the suitability of stratification, the use of ECa was not a priori an interesting option to stratify the sampling. Nevertheless, as already said before, there could be an opportunity to use it to efficiently estimate yield in this plot.

3.3. Lessons learned for future research

Estimated variance of the mean in stratified sampling (StRS) is usually expected to be less than the variance of a simple random sample (SRS) of the same size (Cochran, 1977). Once the sample size is decided (in our case, 12 sampling units), variance for StRS is finally influenced by the particular allocation of the sampling units between the strata. Two or three strata were delimited in this work using auxiliary information maps (NDVI or ECa), to then allocate the same number of sampling units (trees) for all strata (6 or 4 trees per stratum if two or three strata were used, respectively). This procedure probably resulted in a non-optimal allocation of the sampled trees and, as a consequence, in a possible greater uncertainty (or less precision) of the estimates. Cochran, (1977) managed to evaluate, for a fixed sample size *n*, the effect of the deviation from an optimal allocation of sampling units in stratified samples. According to this approach, no significant increase in the variance (or significant loss of efficiency) was expected due to having used the same number of trees per stratum (data not shown). The use of identical allocation in each stratum was for reasons of simplifying the whole process for the farmer. However, it would be advisable in future works to opt for the proportional allocation of sampling trees according to the size of the strata to possibly minimize the variance of the stratified means. Moving away from the optimal allocation of sampling trees should be especially sensitive in yield estimation. This would explain why the variance of the mean in the stratified sampling according to three strata of NDVI was unexpectedly greater than the variance of the SRS. Interesting results comparing proportional and optimum allocation can be found in Brus, (1994).

Finally, being in agreement with other studies (Meyers et al., 2011), sample stratification making use of ancillary information is a possibility to take into account in fruit growing. Cluster analysis has been the option used in this work for the construction of strata. A pending issue for future work is to check other methods to optimize strata such as, for example, the well-known rule of the cumulative root of the frequency function (see Cochran, 1977). Although both SRS and stratified sampling provide unbiased estimates of the population mean, stratifying the sample (Lark & Marchant, 2009) is a way to (i) get more precise estimates (or estimates with less uncertainty), or

(ii) reduce the sample size for a certain precision or efficiency. However, there is a major limiting factor as it is necessary for the ancillary information to be spatially correlated with the variable to be sampled. If this requirement is met, sample estimates can improve in precision. Ultimately, fruit growers and technical advisors can benefit from positive impacts on operating time and cost.

4. CONCLUSIONS

Use of ancillary data such as NDVI in stratified sampling schemes allows yield and quality parameters in a peach orchard to be estimated with greater precision (or greater efficiency). For fruit firmness, the stratification in three strata (scheme B2) is the most recommendable option, achieving almost triple the efficiency compared to simple random sampling (SRS). This means being able to reduce the sample size by almost 67 % for the same precision of the estimates. On the other hand, refractometric index may require a simpler stratification scheme using only two NDVI classes (scheme B1). In terms of yield estimation, the 20 % higher efficiency of also stratified sampling according to two strata of NDVI (scheme B1) allowed the sample size to be reduced by 17 % compared to SRS. In no case the ECa or the combined use of NDVI and ECa have provided substantial advantages compared to the use of NDVI as a single layer of ancillary information. So, the recommendation is to use NDVI as ancillary information to more efficiently estimate yield and quality variables in peach. However, and especially for yield estimates, caution must be taken at the time of allocating sampling trees by strata.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness through the project AgVANCE (AGL2013-48297-C2-2-R). The authors also thank the IRTA Experimental Station in Gimenells (Lleida, Spain) for the possibility of carrying out this sampling study on a peach orchard.

References

Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An Introduction to Spatial Data Analysis. In: M. M. Fischer, A. Getis (Ed.) Handbook of Applied Spatial Analysis, Berlin, Germany: Springer, pp. 73-89.

Brus, D. J. (1994). Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation. *Geoderma, 62*, 233-246.

Brus, D. J., & de Gruijter, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). *Geoderma, 80*, 1-44.

Carrillo, E., Matese, A., Rousseau, J., & Tisseyre, B. (2016). Use of multi-spectral airborne imagery to improve yield sampling in viticulture. *Precis. Agric, 17*, 74-92.

Cochran, W. G. (1977). *Sampling Techniques*. New York, NY, USA: John Wiley & Sons, Inc. 428 pp.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. 85 pp.

Fortes, R., Prieto, M. H., García-Martín, A., Córdoba, A., Martínez, L., & Campillo, C. (2015). Using NDVI and guided sampling to develop yield prediction maps of processing tomato crop. *Span. J. Agric. Res. 13(1)*, e02-004, 9 pages.

Guastaferro, F., Castrignanò, A., De Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precis. Agric, 11*, 600-620.

Jensen, J. R. (1996). Introductory Digital Image Processing: A Remote Sensing Perspective. Englewood Cliffs, NJ, USA: Prentice Hall. 316 pp.

Kazmierski, M., Glemas, P., Rousseau, J., & Tisseyre, B. (2011). Temporal stability of within-field patterns of NDVI in non-irrigated Mediterranean vineyards. *J. Int. Sci. Vigne*. *Vin*., *45(2)*, 61–73.

Lark, R. M., & Marchant, B. P. (2009). Using advanced methods to reduce the cost of soil sampling; ideas for the future. R&D Conference 'Precision in arable farming: current practice and future potential', Grantham, Lincolnshire, UK, 28th-29th October 2009. pp. 18–25 ref.6.

Martínez-Casasnovas, J.A., Agelet-Fernandez, J., Arnó, J., & Ramos, M.C. (2012). Analysis of vineyard differential management zones and relation to vine development, grape maturity and quality. *Span. J. Agric. Res*., *10(2)*, 326–337.

Meyers, J. M., & Vanden Heuvel, J. E. (2014). Research Note. Use of Normalized Difference Vegetation Index Images to Optimize Vineyard Sampling Protocols. *Am. J. Enology Vitic., 65(2)*, 250-253.

Meyers, J. M., Sacks, G. L., van Es, H. M., & Vanden Heuvel, J. E. (2011). Improving vineyard sampling efficiency via dynamic spatially explicit optimisation. *Aust. J. Grape Wine Res., 17,* 306-315.

Miranda, C., & Royo, J. B. (2003). A statistical model to estimate potential yields in peach before bloom. *J. Am. Soc. Hortic. Sci., 128*, 297-301.

Miranda, C., Urretavizcaya, I., Santesteban, L. G., & Royo, J. B. (2015). Sampling stratification using aerial imagery to estimate fruit load and hail damage in nectarine trees. In: J. V. Stafford (Ed.) Precision Agriculture'15, Proceedings of the 10th European Conference on Precision Agriculture, Wageningen, The Netherlands: Wageningen Academic Publishers, pp. 541-546.

Miranda, C., Santesteban, L. G., Urrestarazu, J., Loidi, M., & Royo, J. B. (2018). Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. *Agriculture, 8*(6), 78.

Monestiez, P., Audergon, J. M., & Habib, R. (1990). Spatial dependences and sampling in a fruit tree: a geostatistical approach. *Institut National de la Recherche Agronomique*, Technical Report No. 163, 30.

Moral, F. J., Terrón, J. M., & Marques da Silva, J. R. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil. Tillage. Res., 106*, 335–343.

Ortega-Blu, R., & Molina-Roco, M. (2016). Evaluation of vegetation indices and apparent soil electrical conductivity for site-specific vineyard management in Chile. *Precis. Agric. 17*, 434-450.

Peralta, N. R., & Costa, J. L. (2013). Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Comput. Electron. Agric., 99*, 218-226.

Soil Survey Staff. (2014). *Keys to Soil Taxonomy, 12th ed*. Washington, USA: USDA-Natural Resources Conservation Service.

Sudduth, K. A., Kitchen, N. R., Wiebold, W. J., Batchelor, W. D., Bollero, G. A., Bullock, D. G., et al. (2005). Relating apparent electrical conductivity to soil properties across the north-central USA. *Comput. Electron. Agric.*, 46, 263-283.

Taylor, J., Tisseyre, B., Bramley, R., & Reid, A. (2005). A comparison of the spatial variability of vineyard yield in European and Australian production systems. In: J. V. Stafford (Ed.) Precision Agriculture'05, Proceedings of the 5th European Conference on Precision Agriculture, Wageningen, The Netherlands: Wageningen Academic Publishers, pp. 907–914.

Taylor, J., Acevedo-Opazo, C., Ojeda, H., & Tisseyre, B. (2010). Identification and significance of sources of spatial variability in grapevine water status. *Aust. J. Grape. Wine. R. 16*, 218–226.

Uribeetxebarria, A., Arnó, J., Escolà, A., & Martínez-Casasnovas, J.A. (2018). Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. *Geoderma. 319*, 185–193.

Vanden Heuvel, J. E., Leonardos, E. D., Proctor, J. T. A., Fisher, K. H., & Sullivan, J. A. (2002). Translocation and partitioning patterns of 14C photoassimilate from light- and shade-adapted shoots in greenhouse-grown 'Chardonnay' grapevines (Vitis vinifera L.). *J. Am. Soc. Hortic. Sci., 127,* 912–918.

Webster, R., & Lark, R. M. (2013). *Field sampling for environmental science and management*. London and New York: Routledge, 192 pp.

Wulfsohn, D., Aravena Zamora, F., Potin Téllez, C., Zamora Lagos, I., & García-Fiñana, M. (2012). Multilevel systematic sampling to estimate total fruit number for yield forecasts. *Precis. Agric. 13*, 256-275.

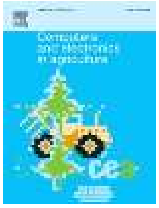
Zude-Sasse, M., Fountas, S., Gemtos, T. A., Abu-Khalaf, N. (2016). Applications of precision agriculture in horticultural crops. *Eur. J. Hortic. Sci. 81(2)*, 78-90.

Chapter 6

Assessing ranked set sampling in a fruit orchard

This chapter contains the following *under review with minor changes* paper in the journal *Computers & Electronics in Agriculture.* JCR-SCI Impact Factor: 2.427. Category: Agriculture, multidisciplinary; 1st Quartile

Uribeetxebarria, A., Martínez-Casasnovas, J.A., Tisseyre, B., Guillaume, S., Escolà, A., Rosell-Polo, J.R., Arnó, J. Assessing ranked set sampling and ancillary data to improve fruit load estimates in peach orchards. *Comput. Electron. Agric.*



Assessing ranked set sampling and ancillary data to improve fruit load estimates in peach orchards

Asier Uribeetxebarria, José A. Martínez-Casasnovas, Bruno Tisseyre, Serge Guillaume, Alexandre Escolà, Joan R. Rosell-Polo, Jaume Arnó

ABSTRACT

Fruit load estimation at plot level before harvest is a key issue in fruit growing. To face this challenge, two sampling methods to estimate fruit load in a peach tree orchard were compared: simple random sampling (SRS) and ranked set sampling (RSS). The study was carried out in a peach orchard (Prunus persica cv. 'Platycarpa') covering a total area of 2.24 ha. Having previously sampled the plot systematically to cover the entire area (104 individual trees or sampling points), both sampling methods (SRS and RSS) were tested by taking samples from this population with varying sample sizes from N = 4 to N = 12. Since RSS requires ancillary information to obtain the samples (ranking mechanism), several proximal and remote sensors already used or recently introduced in agriculture were assessed as data sources. A total of 14 variables provided by 5 different sensors and platforms were considered as potential ancillary variables. Among them, RGB images captured by an unmanned aerial vehicle (UAV), and used to estimate the canopy projected area of individual trees, proved to be the best of the options. This was shown by the high correlation (R = 0.85) between this area and the fruit load, providing RSS with the UAV-based canopy projected area the lowest sampling errors. Then, comparing relative efficiency between random sampling (SRS) and RSS, the latter enables more precise fruit load estimates for any of the considered sample sizes. Interest and opportunity of RSS can be raised from two points of view. In terms of confidence, RSS managed to reduce the variance of fruit load estimates by about half compared to SRS. Relative errors above the 10% threshold were always produced significantly fewer times using RSS, regardless of the sample size. In terms of operation within the plot, sample size could be reduced by 40%, from N = 10 for SRS to N = 6 for RSS without exceeding on average sampling errors of 10%. In summary, fruit growers can take advantage of the combined use of appropriate data (RGB images from UAV) and RSS to optimize sample sizes and operational sampling costs in fruit growing.

KEYWORDS

Ranked set sampling, Fruit growing, Sampling efficiency, Yield estimation

1. INTRODUCTION

Fruit tree crops are considered high value products (Aggelopoulou et al. 2010), but they demand more tasks from farmers compared to arable crops. Among the tasks required in peach orchards, harvesting is a complex process that, necessarily done by hand, usually also needs to be finished in a short time window. Because of that, a large amount of labor is often needed once the quality standards demanded by the market have been reached (ripening of the fruit). Hence, obtaining a reliable yield estimate in advance would improve the logistics of the entire process (Wulfsohn et al. 2012), also contributing to anticipate operational costs more effectively.

In perennial crops, it is known that regular patterns are adopted by planting commercial plots with genetically uniform plant material (Miranda et al. 2015). Therefore, it is expected to achieve regular growths and certain homogeneity in production and quality. From this point of view, making an estimate of the yield should not raise any issue. By randomly choosing a few trees, the estimation of yield should be quite accurate knowing the total number of trees within the plot. However, perennial plots are not always homogeneous and spatial variability is present due to environmental factors, as it has been shown in different studies (Berman et al. 1996; Taylor et al. 2005; Arnó et al. 2012). Moreover, since many of the factors that affect yield (or fruit load) are spatial dependent, yield spatial distribution within the orchards usually presents patterns and is not random (Aggelopoulou et al. 2013). Under this scenario of structured variability with yield potential varying for each tree and for the different productive areas within the plot, simple random sampling (SRS) strategy may be inefficient. In order to fulfill fruit grower demands, yield forecast should be made by using 5-6 sampling trees and with maximum error of 10%. Therefore, SRS may not be the best option since it does not take into account any spatial organization of the variable to estimate. Moreover, a large sample size may be needed to fulfill growers' constraints (Carrillo et al. 2016). Only the simplicity to implement and understand the results of SRS would explain why this method is still used in agriculture in forecasting tasks. To overcome this issue, new and accurate sampling methods have to be proposed to fruit growers and farm managers.

Although it is fair to mention that one of the pioneering work on sampling in fruit trees was published in the forties (Pearce, 1944), sampling methods making use of ancillary information provided by high spatial resolution sensors have been proposed in recent years. Previous analysis of the within-field variability of such information is often key when designing the sampling strategy. Even spatial variability on a smaller scale has also been the option used by other authors. So in 1990, Monestiez et al. proposed a geostatistical approach to optimally sample within the trees to improve yield estimates. Wulfshon et al. (2012) opted for a multilevel systematic sampling to estimate the number of fruits reaching errors of just 10%, then extending this same method to estimate quality parameters (Martinez Vega et al. 2013). More recently, stratified sampling based on the use of auxiliary information for the spatial stratification of the sample has also shown promising results. Examples of this research are shown in Arnó et al. (2017) and Miranda et al. (2015, 2018). Specifically, trees were stratified using clustered NDVI-based aerial images (NDVI-Normalized Difference Vegetation Index), either alone or in combination with other information layers such as the trunk cross-sectional area (TCSA). However, obtaining the representative NDVI value for an individual canopy or, even more, a manual measurement of the trees is not always a simple task (Fountas et al. 2011).

The use of ancillary information to stratify sampling has been favored by the availability and rapid access to proximal and remote sensors data. Among the sources of information most used for this purpose, it is worth mentioning the NDVI images (Meyers and Vanden Heuvel 2014), the apparent electrical conductivity (ECa) of the soil (Mann et al. 2011; Arnó et al. 2017), and time series of yield maps provided by manual harvesting or monitored by sensors on harvesters (Araya et al. 2017). Other data are also available while other technologies for acquiring crop data have been fine-tuned. It refers to RGB and multispectral cameras (Ulzii-Orshikh et al. 2017), mounted on terrestrial platforms or unmanned aerial vehicles (UAV), and mobile terrestrial laser scanners (MTLS) based on 2D LiDAR sensors (Rosell et al. 2009; Escolà et al. 2017), offering new possibilities for capturing canopy information from fruit trees. These higher resolution data allow the trees to be characterized in greater detail, but possible advantages of their use in sampling have not yet been verified.

As already mentioned, SRS is a well-known method in agriculture. It is statistically consistent, but not always the most efficient, especially in the case the variability is not randomly organized. Moreover, there is a growing interest in developing sampling methods that, in addition to provide accurate and unbiased yield estimates, allow small sample sizes to be used. Stratified sampling (Cochran 1977) may be an option, because it fulfills the characteristics described above. However, to stratify the samples based on ancillary variable maps, fruit growers have to define the strata as a preliminary step by choosing between different classifications and zoning techniques. How many strata, how the strata are obtained, and how sampling points should be allocated in each stratum are decisions to be made. Therefore, this combination of subjective factors probably ends up affecting the goodness of yield estimates. Another option is using the Ranked Set Sampling (RSS) (McIntyre 1952). This method is interesting from the operational point of view because it does not need these preliminary steps. To obtain the sample, once the number of sites to sample is set, RSS only requires applying a ranking mechanism based on the distribution of an auxiliary variable (Wolfe 2010). This ranking mechanism allows the trees (sampling points), previously taken randomly in a first iteration from the population, to be ranked from lowest to highest, according to the value that in these trees takes an ancillary variable. More details about the ranking mechanism and the different iterations of the whole process can be found in the corresponding section below (Basics of the sampling methods). With everything, the ancillary information is therefore key in the procedure since it has to present correlation with the variable to estimate (the fruit load, in the case of this work). In short, finding the best ancillary information to perform RSS in fruit orchards is a pending issue to solve.

Focusing our research on comparing RSS and SRS, the main objective of the present work was to evaluate whether the use of ranked samples significantly improves fruit load estimates in fruit growing. To do this, two main issues were addressed: i) to determine the most appropriate ancillary information to be used in the ranking process, and ii) to assess the efficiency in terms of relative error when comparing RSS to SRS for different sample sizes.

2. MATERIALS AND METHODS

2.1 Study plot

The study was carried out in a commercial peach orchard located in Aitona (45° 94’ 71” N, 0° 29’ 20” E, ETRS89, 160 m a.s.l., Lleida, Catalonia, Spain). The plot covered an area of 2.24 ha. It was planted in 2012 with white peach (*Prunus persica* cv. ‘Platycarpa’) according to a regular plantation pattern of 5x2 m (Fig. 1).

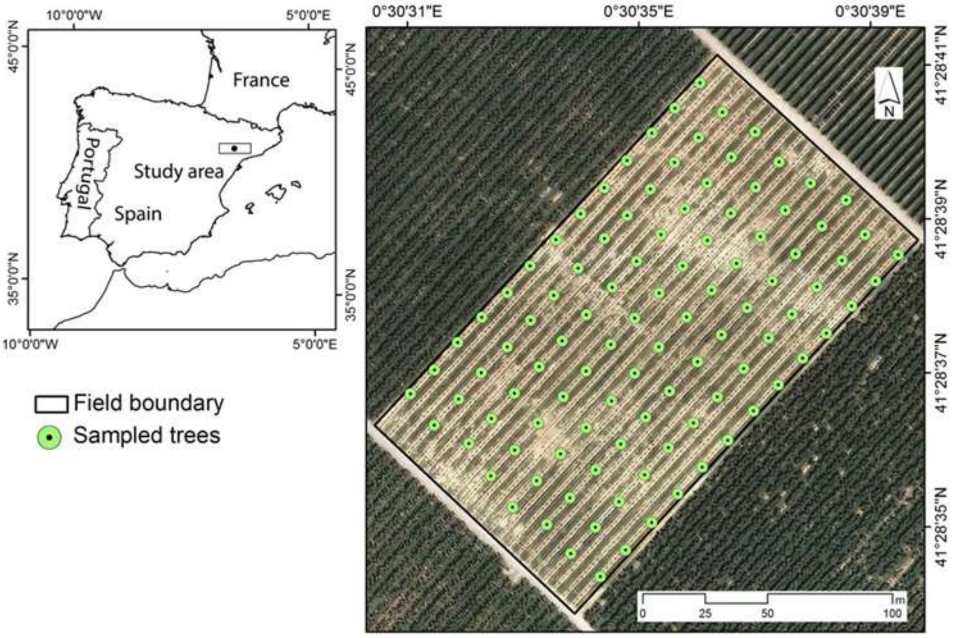


Fig 1 Location of the study area and orthophoto (2016) showing the 104 trees sampled within the plot. (Orthophoto source: Cartographic and Geologic Institute of Catalonia).

Peach trees were planted in form of “Catalan” vessel shape, which is the most common training system in Catalonia. This canopy management system produces some visible gaps between trees within the same row, being relatively easy to individualize each tree. Regarding agricultural tasks, the plot was managed like many of the commercial orchards in the region. Drip irrigation system was used for water and fertilizers supply.

The plot was representative of the so-called Ebro depression area, characterized by the presence of broad flat regions resulting from infilled valley bottoms and residual landforms. Climate is typical of hot semi-arid areas, with strong seasonal temperature variations and an annual rainfall frequently below 400 mm. As a consequence, soils in the region usually have high concentrations of calcium carbonate ($CaCO_3$) and low content of organic matter (OM). However, as a main feature, within-field soil heterogeneity is prevalent in many plots due to the levelling of original terraces in the eighties to facilitate agricultural mechanization. The plot under study resulted from this type of transformation. As for the spatial variability of trees within the plot, individualized canopy projected area values for each tree were obtained (see section 2.4.2 for more details). The average value was 2.82 m^2, with a range between 0.62 m^2 and 5.50 m^2, and a Coefficient of Variation (CV) close to 21%.

2.2. Systematic sampling of the orchard

A regular sampling grid of 15x15 m was defined in 2016 to sample the plot systematically (Fig. 1). The trees coinciding with the grid nodes were taken as sampling points. Grid size was previously defined by variographic analysis of the apparent electrical conductivity (ECa) data provided by a Veris 3100 soil sensor (Veris Technologies, Inc., Salinas, KS, USA) from an on-the-go survey of the plot. Once a strong autocorrelation of the ECa data was verified, half of the variogram range (ArcGIS 10.4.1 software) was taken as the optimum distance between sampling points (Kerry et al. 2010). On the other hand, to avoid border effect, no sample sites were positioned in a buffer zone of 15 meters from the limits that separated other plots (short sides of the rectangular shape of the plot). However, this buffer zone was not applied on the sides alongside with other fields because the cultivated peach variety and the canopy training system were the same in these adjacent plots.

In total, 104 trees (sampling points) were defined (Fig. 1), and the number of fruits was counted manually in each tree four weeks before harvest. As proposed by Miranda and Royo (2003), yield in fruit orchards is usually measured in terms of fruit load (number of fruits per tree). As yield (kg/tree) basically depends on fruit load, the latter was the variable used in this work. Covering the entire plot, these 104 trees were taken as a reference population on which to compare two sampling methods.

2.3. Basics of the sampling methods

2.3.1. Simple random sampling (SRS)

The size of the sample was initially set at N = 12 trees or sampling points (with $N \ll 104$). Randomly selected without replacement, each of the twelve measured fruit load values could be considered representative of the population (or parcel under study), and with identical probability of choice. Therefore, the mean of the sample allowed the average number of fruits per tree to be estimated without bias. However, there was a distinct possibility that such a mean did not provide a truly representative picture of the population (Wolfe 2010).

2.3.2. Ranked set sampling (RSS)

Ranked set sampling is not a new method. Initially developed by McIntyre (1952), the method was proved to be effective in improving the efficiency of pasture sampling. In this work, making use of adequate ancillary information (necessary for the ranking mechanism as it has been previously mentioned), the method was updated to evaluate its use in fruit growing.

Briefly, the method is based on an iterative sampling process (Wolfe, 2010). Following the procedure for a sample size N = 12, a first sample of size k = 12 was taken randomly without replacement from the 104 trees initially marked within the plot, and then the 12 selected trees were arranged according to some attribute (ranking mechanism). This process of sorting the sampled trees allowed a ranking to be established, usually from the lowest to the highest presence of the selected attribute. In our case, the attribute used in the ranking mechanism was an ancillary variable (or secondary variable) that, being related to the number of fruits per tree (as a relevant component of yield), should also be easy to measure and at an affordable cost. The individual (tree) with the lowest value in this ranking was finally included as the first item in the ranked set sample, and the property of interest (fruit load) was formally considered only for this unit and denoted by $Y_{[1]}$. Obviously, the remaining *k*-1 observations were discarded from this first sampling step.

In a second iteration, another sample of *k* = 12 observations was again selected and ranked following the same procedure as before. The individual (tree) ranked as the second smallest attending the attribute (ancillary variable) was now chosen, providing a second number of fruits value $Y_{[2]}$ that was added to the sample. The process was repeated until obtaining *N* = 12 measured observations $Y_{[1]}, Y_{[2]}, \cdots, Y_{[N]}$ that constitutes a balanced ranked set sample of size *N*.

The mean of the *N* values (ranked set sample mean) is, like SRS, an unbiased estimate of the population mean (Takahasi and Wakimoto 1968). However, having applied a ranking mechanism, each of the individual sample items in a balanced RSS represented a very distinctly different portion of the underlying population. Figure 2 represents this situation. The histogram represents the actual distribution of the number of fruits per tree (fruit load) in the plot under study (approximately, normal). Obtaining a sample through SRS involves sampling within that normal population (red curve in Fig. 2), and non-representative sample values may be drawn. Instead, using a perfect ranking mechanism (since the ancillary variable that provides the attribute is well correlated with the fruit load), each of the observations included in the RSS behave like mutually independent order statistics, with densities that can be represented by the individual marginal density curves in Figure 2 (in this case, *k* = 6). Therefore, because of this extra structure provided by the judgment ranking (Wolfe 2004), RSS observations were much more likely to represent the full range of variation to estimate the population mean in a more effective way. This procedure is probably better than or equivalent to stratifying according to the probability distribution of the ancillary variable (stratified sampling based on the distribution of the percentiles). Adopting the previous notation for a sample size *N* = 12, the mean was calculated using (1),

$$\overline{Y}_{RSS} = \frac{1}{N}\sum_{j=1}^{N} Y_{[j]}. \qquad (1)$$

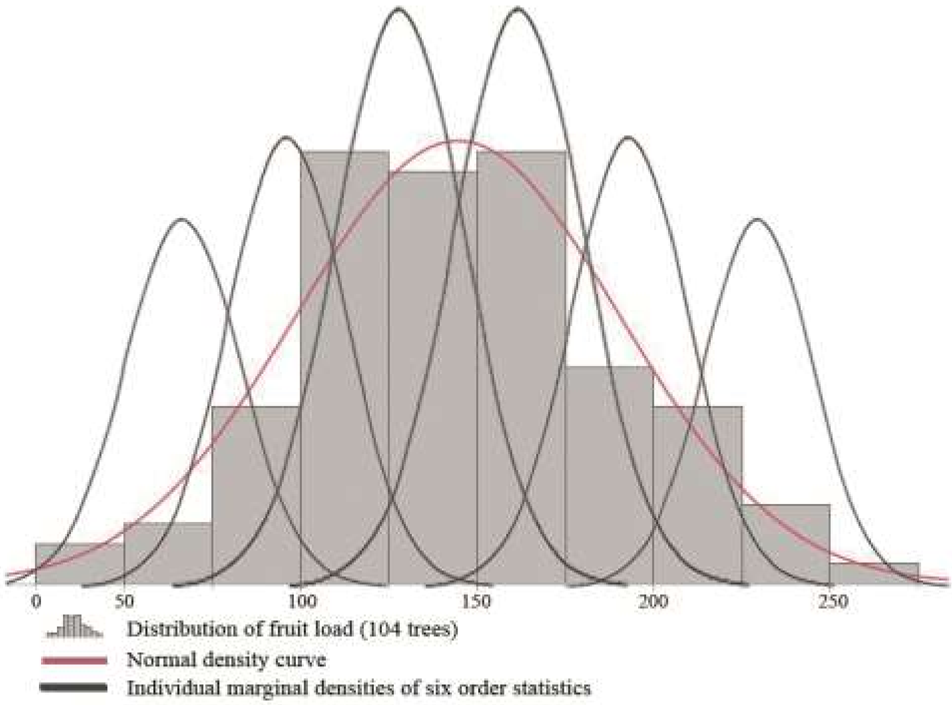


Fig 2 Distribution of the fruit load (histogram of the 104 trees sampled), normal density curve (in red) and the individual marginal densities of six order statistics $Y_{[1]}, Y_{[2]}, Y_{[3]}, Y_{[4]}, Y_{[5]}, Y_{[6]}$ (solid curves, in order of peaks, from the minimum, $Y_{[1]}$, on the left to the maximum, $Y_{[6]}$, on the right) for a random sample of size 6 from the normal distribution. Figure adapted from Wolfe's, (2010).

Regarding the variance, Wolfe, (2010) provided the formula (2):

$$Var(\overline{Y}_{RSS}) = Var(\overline{Y}_{SRS}) - \frac{1}{N^2}\sum_{j=1}^{N}\left(\mu_{[j]}^{*} - \mu\right)^2 \quad (2)$$

where $\mu_{[j]}^{*} = E\left(Y_{[j]}\right)$, for $j = 1, \ldots, N$. Since $\sum_{j=1}^{N}\left(\mu_{[j]}^{*} - \mu\right)^2 \geq 0$, it follows from Eq. (2) that the variance of the RSS mean ($\overline{Y}_{RSS}$) is, at most, equal to the variance of the SRS mean ($\overline{Y}_{SRS}$). Therefore, the RSS mean is, theoretically, a more precise estimator of the population mean μ than the SRS mean based on the same number of measured observations. The reliability of the judgment ranking process in separating order statistic expectations $\mu_{[j]}^{*}$ is a key in improving the efficiency of the RSS compared to the SRS.

2.4. Ancillary variables in the ranking mechanism

In this section, the most relevant characteristics of the ancillary variables used to rank sampled trees are shown. Both terrestrial and remote sensors, whether used or recently introduced in the framework of precision agriculture, were considered to provide these data. Table 1 shows the platforms and sensors tested with additional indication of the ancillary information provided as well as a first assessment of the technical difficulty to obtain and process the data.

Table 1 Sensors and ancillary information that were evaluated for the ranking of sampled trees in the RSS method.

Acquisition platform	Sensor	View of the canopy	Ancillary information	Geometry-based	Reflectance-based	Obtaining process
ATV	Veris 3100	Soil sensor	(B) shallow ECa			ME
			(C) deep ECa			ME
ATV	OptRx	Lateral view	(D) NDRE		X	ME
			(E) NDVI		X	ME
UAV	RGB camera	Nadir	(F) Tree canopy perimeter	X		ME
			(G) Tree canopy projected area	X		ME
Airplane	DMSC	Nadir	(H) NDVI		X	ME
			(I) $NDVI_C$	X	X	HG
	RGB camera		(J) Tree canopy perimeter	X		ME
			(K) Tree canopy projected area	X		ME
Terrestrial platform	MTLS	Lateral view	(L) Canopy impacts	X		HG
			(M) Canopy volume	X		HG
			(N) Canopy volume	X		HG
			(O) Canopy volume	X		HG

ATV (All-Terrain Vehicle); UAV (Unmanned Aerial Vehicle); DMSC (Digital MultiSpectral Camera); MTLS (Mobile Terrestrial Laser Scanner); ECa (apparent electrical conductivity); NDRE (Normalized Difference Red Edge); NDVI (Normalized Difference Vegetation Index); $NDVI_C$ (corrected value of the NDVI); ME (moderately easy); HG (hard to get)

2.4.1. Apparent electrical conductivity (ECa)

Soil apparent electrical conductivity (ECa) was measured using a Veris 3100 (Veris Technologies, Inc., Salina, KS, USA). The soil survey was conducted on March 1st, 2016, when the soil had moisture content close to field capacity. Passing through all the alleyways within the plot, two simultaneous measurements of ECa were recorded: shallow ECa (0-30 cm) and deep ECa (0-90 cm). ECa measurements were georeferenced using a Trimble AgGPS332 GPS with SBAS differential correction (EGNOS system). An acquisition frequency of 1Hz was used giving approximately 750 sampling points per hectare.

Two ECa raster maps (shallow and deep) were then obtained by ordinary kriging interpolation on a 1 m grid. By superimposing the layer of 104 polygons corresponding to the 104 trees within the plot, the respective values of shallow ECa (variable B, Table 1) and deep ECa (variable C, Table 1) were extracted. The mean ECa value was calculated for each tree at each surveying depth.

2.4.2. Tree canopy projected area and tree canopy perimeter

Tree canopy projected area and tree canopy perimeter were obtained by processing two images acquired from two different platforms: airplane and unmanned aerial vehicle (UAV). Both images were acquired on May 16, 2016, approximately at mid-day and under clear sky conditions.

Each platform had its own camera. For the UAV, images were grabbed with a 16 MP Panasonic GX 7 camera (Panasonic Corporation, Osaka, Japan) with a 20 mm “pancake” lens coupled on a Mikrokopter Oktokopter 6S12 XL eight rotor UAV (HiSystems GmbH, Moomerland, Germany). The UAV was handled in manual configuration at constant speed of 18 km·h^{-1} and constant altitude up to 100 m. Pitch and roll movements were minimized using MK HiSight SLR2 gimbal support equipped with two servo motors. The camera was configured to take an image every two seconds during the flight, being the autofocus activated and the exposure fixed. Nadir images were pre-selected to obtain an orthomosaic image with a spatial resolution of 2 cm per pixel using Agisoft Photoscan Professional software (Agisoft LLC, St. Petersburg, Russia).

Regarding the airplane image, a 4-band digital multispectral camera (DMSC) was used allowing images centred at 450 nm (blue), 550 nm (green), 675 nm (red) and 780 nm (near infrared) to be acquired. The images were pre-processed by SpecTerra (SpecTerra Services Pty Ltd, Leederville, Western Australia) to correct lens aberration, and adjust scene brightness by the bidirectional reflectance distribution function (BRDF). At the moment of photographing the plot, the airplane was flying at an altitude of 2000 m. The spatial resolution of the pre-processed image was 25 cm per pixel. The aircraft used was a CESSNA 1725 Sky Hawk operated by RS (RS Servicios de Teledetección SL, Lleida, Spain).

In both cases, images were processed by manually delimiting each of the 104 individual trees sampled within the plot. The polygons used to delimit the projected canopies to the ground allowed the area and the perimeter of each tree canopy to be calculated. Table 1 shows the four variables (symbolized with a capital letter) that were used as ancillary information on the canopy's geometry: airplane-based canopy perimeter (J), airplane-based canopy projected area (K), UAV-based canopy perimeter (F), and UAV-based canopy projected area (G).

2.4.3. Tree canopy reflectance

The aerial image provided by the multispectral camera (NDVI image, Rouse Jr et al. 1974) was post-processed according to the following procedure. First, and having eliminated the NDVI pixels with values lower than 0.45 to discard other land covers (i.e. bare soil) that were not tree canopy, all the trees (or canopies) within the plot were then easily delimited by using automatically defined polygons. Specifically, 1816 trees were individualized this way. Then, in a second step, NDVI pixels within each polygon (or canopy) were extracted by overlapping the NDVI image and the polygon layer. The average value of NDVI was then calculated and assigned to the centroid of each polygon. The third step consisted in creating a continuous map (or surface map) by interpolation (ordinary kriging) of the average NDVI values of the centroids. By superimposing this new NDVI map (1 m pixel resolution) with the polygons corresponding to the 104 trees, an average value of NDVI for each tree was obtained. This ancillary information was denoted with the letter H (Table 1). As with the previous maps for the other sensors, ArcMap 10.4.1 (ESRI, Redlands, CA, USA) performed spatial data analysis.

On the other hand, it is worth mentioning that the single information of the NDVI could be inconsistent with the expected fruit load of a tree. The bigger the canopy, the higher the expected fruit load (Mann et al. 2011). However, average NDVI could take similar values in trees of different sizes. With this in mind, formula (3) was used to correct values of the NDVI to be obtained for each tree (variable I, Table 1),

$$NDVI_C = NDVI \times \frac{TCA}{TCA_{max}} \tag{3}$$

where NDVI and $NDVI_C$ were the initial value and the corrected value of the NDVI, respectively, TCA was the tree canopy projected area obtained from the airplane image, and TCA_{max} was the largest tree canopy projected area within the plot.

Apart from the remote sensor, a terrestrial OptRx sensor (Ag Leader Technology, Ames, IA, USA) was also used to acquire canopy reflectance data using an all-terrain vehicle (ATV) as a platform. Since the readings were taken sideways passing through the alleyways of the orchard, data referred to the lateral reflectance of the canopy. Specifically, the sensor was maintained at a distance of approximately 1 m from the canopy and moved at a constant speed of 5 $km \cdot h^{-1}$. With the OptRx, only one side of each row was measured. Three different wavelengths were acquired every second (1 Hz) in the ranges of red (670 nm), red edge (728 nm), and near infrared (775 nm). So, the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Red Edge (NDRE) were finally obtained and georeferenced by also acquiring the position using a Trimble AgGPS332 GPS with SBAS differential correction (EGNOS system). An average of 169 data per row was obtained. Then, following the same procedure as before, the NDVI and NDRE data were interpolated by ordinary kriging to obtain two surface maps with a final resolution (grid) of 1 m for each vegetation index. By superimposing the surface maps with the layer of 104 polygons, the average values of NDVI and NDRE were obtained and denoted as variables E and D, respectively (Table 1). The date of the terrestrial sensor measurements was May 17th, 2016.

2.4.4. Tree volume

The use of a MTLS allowed four different ancillary variables to be obtained. The sensor used was a UTM30-LX-EW time-of-flight LiDAR (HOKUYO, Osaka, Japan) that could perform 40 scans per second (40 Hz). In addition to this 2D LiDAR sensor, the MTLS integrated a GPS1200+ (Leica Geosystems AG, Heerbrugg, Switzerland) RTK-GNSS system (a real-time kinematics global navigation satellite system receiving GPS and Glonass constellation signals). So, after processing all data, the MTLS provided measurements of the position of the impact points produced between the laser beam and the canopy. These georeferenced impacts formed a point cloud that was then analyzed to calculate the canopy volume. A more detailed description of features, components for acquisition and field use methodology can be found in Escolà et al. (2017).

Like the OptRx sensor, the MTLS laterally scanned the canopies of all the rows within the plot, but in this case, each row were measured from both sides to obtain a point cloud representative of the entire canopy. Point cloud visualization and specific computations were performed with CloudCompare (CloudCompare [GPL software] v2.6.1 2015). In short, the 3D point cloud was first classified into canopy and ground points. Points located at a height less than 0.4 m above the ground were discarded to eliminate weeds and ridges. The remaining points were then processed to individualize each tree trying to eliminate possible impacts (outliers) that were clearly outside the canopy (presenting positions more than 2.5 standard deviations away from 6 neighboring points).

Once the cloud of impact points for each tree was delimited (in our case, 104-point clouds), the first variable that was obtained was the number of canopy impacts (variable L, Table 1). CloudCompare 'octree' allowed the cubical initial volume including the entire canopy for each tree to be recursively divided into smaller cubes until visually adjusting the actual volume of the canopy. Adding the unit volume of each adjusted cube or voxel (VOlumetric piXEL), the final volume of the canopy in m^3 was obtained. Depending on either how the voxel size was set, manually or automatically, two different volumes could be obtained. The variable M (Table 1) referred to manually adjusted voxels, while the variable N (Table 1) estimated volume by adjusting a voxel size given by default parameters of the software.

The last volume (variable O, Table 1) was calculated using the '2.5D Volume' tool in the CloudCompare software. Having projected the point cloud for a specific tree on the XY plane (Z-axis projection direction), this tool allowed the volume between the 2D rasterized cloud and the ground surface (taken as arbitrary plane) to be computed. As multiple points of the canopy could fall inside each cell, the maximum height was taken for calculation.

Obviously, other methods could have been used to calculate the volume from LiDAR point clouds (Escolà et al. 2017). However, the special structure of the 'Catalan' vessel training system, open with branches in different directions and large gaps inside the canopy, could make it more convenient to use the voxelization procedure as proposed by Underwood et al. (2016) in almond. The readings with MTLS system were made on May 17th, 2016.

2.5. Evaluating the efficiency of sampling methods

To compare the SRS versus the RSS, *B* (*B* = 1000) samples from size *N* = 4 to *N* = 12 were resampled each time and for each method, with the sampling population being the 104 trees systematically sampled within the plot. By calculating the mean of each sample $\overline{Y}_i$, the variance of the distribution of means was computed using expression (4),

$$var(\overline{Y}) = \frac{1}{B}\sum_{i=1}^{B}\left(\overline{Y}_i - \overline{\overline{Y}}\right)^2 \tag{4}$$

where *B* = 1000 was the number of samples in the resampling method, $\overline{Y}_i$ was the sample mean of the fruit load, and $\overline{\overline{Y}}$ was the average fruit load for the 104 trees previously sampled covering the entire area of the plot ($\overline{\overline{Y}}$ = 144.7 fruits per tree). The latter was considered the representative value of the plot (or best estimator of the population mean), besides being practically coincident in all the resampling processes with the average of the 1000 means (data not shown).

Both sampling methods (SRS and RSS) provided an unbiased estimate of the population mean (in this case, the average of the 104 trees as a fairly representative measure of the plot mean). So in both cases, the mean squared error (*MSE*) coincided with the variance. Being the efficiency or precision the inverse of the variance, the comparison between the two sampling methods could be done using the so-called relative efficiency (5),

$$RE = \frac{Efficiency\,(\bar{Y}_{RSS})}{Efficiency\,(\bar{Y}_{SRS})} = \frac{var(\bar{Y}_{SRS})}{var(\bar{Y}_{RSS})} \tag{5}$$

once obtained $var(\bar{Y}_{SRS})$ and $var(\bar{Y}_{RSS})$ by (4). RE values greater than 1 would be expected (Chen, Bai and Sinha 2004), showing that RSS would be more precise (or more efficient) than SRS. At worst (Webster and Lark 2013), $RE = 1$ when the ranking process is completely imperfect, that is, there is no correlation between the fruit load and the ancillary variable used to rank the samples. Also, as Wolfe (2010) points out, good concomitant information is necessary to avoid error-prone ranking.

Two additional efficiency indicators were computed. First, a relative error of sampling (e_r) for each method was obtained using equation (6),

$$e_r(\bar{Y}) = \mathrm{CV}(\bar{Y}) = \frac{\sqrt{var(\bar{Y})}}{\bar{\bar{Y}}} \times 100 \tag{6}$$

being $\sqrt{var(\bar{Y})}$, the standard deviation of the mean. The last indicator allowed the number of individual samples with relative error greater than 10%, $n[e_r(\bar{Y}_i) > 10\%]$, to be computed in percentage terms over the total of 1000 samples (7),

$$PS = \frac{n[e_r(\bar{Y}_i) > 10\%]}{B} \times 100 \tag{7}$$

with the relative error of an individual sample calculated according to (8),

$$e_r(\bar{Y}_i) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{\bar{Y}})^2}}{\bar{\bar{Y}}} \times 100. \tag{8}$$

Note that this procedure was also used to test the optimal ancillary information to run the RSS. For this, the resampling of *B* = 1000 samples of size *N* = 12 using each auxiliary variable shown in Table 1 was used. Indices defined by equations (4), (6) and (7) were used to compare the different auxiliary variables used in the ranking mechanism to provide an estimate of the fruit load. In all cases, resampling was performed by programming in R software, version 3.3.2.

3. RESULTS AND DISCUSSION

3.1. Relationship between fruit load and the acquired ancillary variables

Figure 3 shows the scatter plot defined by the first two factors of a principal component analysis (PCA) showing the correlation between the ancillary variables altogether and fruit load. By the two factors, 65.9% of variability is explained. In summary, many of the ancillary variables were well summarized by the two first components of the PCA allowing two sets of variables to be discriminated.

The variables with the highest correlation to the first component, apart from the fruit load (variable A), were those obtained using remote sensing (UAV and airplane using RGB and multispectral cameras) and MTLS. So, the first component was basically related to the geometry and vigor of the trees. The soil apparent electrical conductivity (shallow and deep values) was correlated to component 2 instead of component 1. In short, component 2 should be interpreted as related to soil variability. Finally, NDVI and NDRE vegetation indices obtained from the OptRx proximal sensor were not clearly included in any of the previous groups (therefore, they showed a weak correlation with both components). As has been mentioned before, "Catalan" vessel is the adopted training system which favours open canopies to improve solar exposure in detriment of the lateral side. In short, a vegetation wall is not easily measurable from the lateral side of the canopy. Because of that, the field of view of the canopy varies between sensors (lateral from terrestrial platform and nadir from aerial platform), and NDVI readings are surely affected. Probably, the best correlation with nadir vision is due to better cover the full size of the canopy.

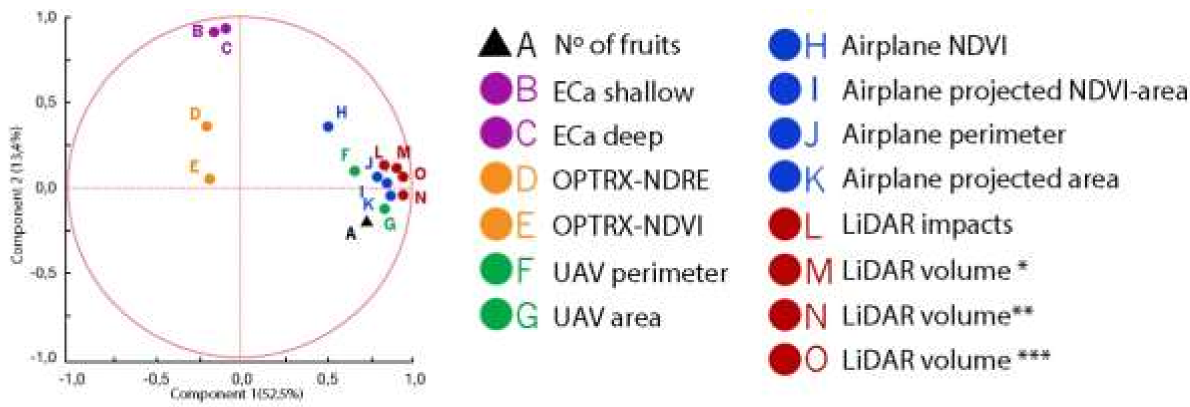


Fig 3 Scatterplot showing the correlation between the ancillary variables and fruit load (variable A) with respect to the two principal components of PCA.

Relevant information about the relationship between fruit load and all other variables was contrasted visually by the scatterplot of a PCA (Fig. 3). The surprising weak correlation between the number of fruits (variable A) and component 2 made it unwise to use the Veris sensor and the ECa as ancillary information in RSS schemes. This weak correlation was not expected when systematic sampling was designed, according to the findings of other authors in fruit growing (Käthner and Zude-Sasse 2015). In spite of this, variability at plot scale was also captured by covering the whole area of the orchard. The weak correlation could be due to drip fertigation allowing roots to grow in a controlled environment with little final soil influence (taking ECa readings in the alleyways outside the wet bulb) on canopy and fruit load (Uribeetxebarria et al. 2018). There being little correlation between ECa and fruit load, the ranking process may not be optimal to obtain a sample with the additional information needed for a more efficient estimation of the mean. Figure 3 shows a nearly zero relationship between fruit load and variables D and E obtained by the OptRx sensor. For this reason, its use in RSS was rejected. However, the variables correlated to component 1 could be of interest. In short, the options for ancillary data to be used were limited to three data sources, i) RGB cameras mounted on remote acquisition platforms (UAV and airplane, variables F, G, J and K), ii) airplane images supplied by DMSC (variables H and I), and iii) MTLS point clouds (variables L, M, N and O). The close spatial arrangement within component 1 (Fig. 3), including fruit load, helped in this interpretation.

Going into detail (Fig. 3), variable H (airplane-based NDVI) was the ancillary information with apparently the weakest correlation to fruit load among the variables of component 1. While variable H was only based on reflectance, the other variables were mainly based on geometric parameters. Even so, all the variables that were grouped as mostly correlated with component 1 were selected in order to analyze their convenience as ancillary variables in the next section. With respect to the volume of the canopy, only the data calculated with the tool 'Volume 2.5D' (variable O) were considered to avoid redundant information.

3.2. Ancillary information to be used in RSS to improve fruit load estimates

Table 2 shows the results of linear correlation between the number of fruits per tree (variable A) and the most relevant ancillary variables within component 1. As expected, the best correlation was obtained using the UAV-based canopy projected area (variable G) with a value of the linear correlation coefficient of 0.85. The worst result (coefficient of 0.21) corresponded to the airplane NDVI (variable H). The rest of the variables ranging between these two extreme values (Table 2).

Table 2 Assessment of RSS efficiency according to the ancillary variables used in the ranking mechanism. Variables are ordered according to their correlation with fruit load.

	R*	$MSE(\overline{Y})^{**}$	$e_r(\overline{Y})^{***}$	PS^{****}
UAV-based canopy projected area (G)	0.85	79.46	6.16	10.8
LiDAR-based canopy volume (O)	0.68	119.71	7.56	17.6
LiDAR-based canopy impacts (L)	0.60	118.68	7.52	18.4
Airplane-based canopy projected area (K)	0.55	149.56	8.45	22
Airplane-based canopy perimeter (J)	0.54	134.47	8.01	20.8
Airplane $NDVI_C$ (I)	0.52	144.56	8.30	22.6
UAV-based canopy perimeter (F)	0.49	147.30	8.38	22.8
Airplane NDVI (H)	0.21	158.13	8.69	25.2

*Pearson's linear correlation coefficient

**Mean squared error (or mean variance)

***Relative error of sampling

****Percentage of samples exceeding the error threshold of 10 %

As expected, Table 2 shows that *MSE* values follow almost the same order as the correlation coefficient showing that the higher the correlation, the lower the error in fruit load prediction with the RSS. The trend of the relative error ($e_r(\overline{Y})$) was similar, going from 6.16 % when the ancillary variable was the UAV-based canopy projected area to 8.69 % using the airplane NDVI to rank items within the samples. Since a relative error of 10 % may be acceptable in the fruit production sector, any of the ancillary variables could be candidates to be used in RSS schemes. However, it is also true that the sample size used in the comparison (N = 12) is not usual in practice, and has undoubtedly contributed to improve the efficiency in fruit load estimates above what is expected. Hence, it is necessary to check how the efficiency varies by reducing the sample size closer to the usual sizes used by fruit growers.

Table 2 also shows the percentage of times the relative error exceeded the 10 % threshold considering 1000 samples with a sample size N = 12. This probability is a way of knowing the risk assumed by fruit growers since they usually perform a single sampling. Values ranged from 10.8 % (variable G) to 25.2 % (variable H), and corroborate the results of *MSE* and linear correlation. The maximum difference between ancillary variables was 133 % for this parameter, being 75 % for the relative sampling error.

On the other hand, the results shown in Table 2 were in accordance with what was noticed by Nahhas et al. (2002) and Stokes (2007) when applying RSS. The better the correlation between the variable to be estimated (number of fruits per tree, in our case) and the concomitant variable (or ancillary variable) used in the ranking process, the better the sampling efficiency.

Apart from good efficiency results, the ancillary variable to be used must be easily measurable. RGB cameras embedded in UAV meet this requirement. The use of UAV in precision agriculture is a booming and low cost technology as suggested by Lelong et al. (2008). For this reason, the UAV-based canopy projected area (variable G, Tables 1 and 2) was finally chosen as the optimal ancillary variable to use in RSS schemes in fruit growing. The advantage of this variable is the high resolution of the images compared to other remote data sources.

3.3. Resolution of ancillary variables in RSS: the key factor

In peach orchards, fruit production and quality are clearly influenced by canopy lighting conditions (Tang et al. 2015; Minas et al. 2018). Seeking to enhance floral induction and fruit growth, fruit growers adopt canopy open-center training systems (such as the typical 'Catalan' peach vessel) to maximize exposure to solar radiation. In this way, peach trees with larger canopy projected area usually have a greater number of larger fruits (Marini 2002). In addition, the incidence of unproductive interior branches is lower, avoiding what usually happens in canopies formed into traditional or spherical vessels (Marini et al. 1995). In RSS, an ancillary information system that manages to individualize the canopy of trees in a precise manner will be essential, i) to ensure an optimal and accurate ranking process of the auxiliary variable, and ii) to cover all the variability of the objective variable (number of fruits) (Fig 2).

Continuing the thread of the previous paragraphs, variables that best meet previously mentioned characteristics are those based on geometrical properties of the canopy (Table 2). UAV-based canopy projected area (variable G) was the best rated ancillary variable. In contrast, LiDAR-based canopy volume (O) or impacts (L) and airplane-based canopy projected area (K) obtained less satisfactory sampling efficiency results. What could be the reason for this difference? Probably, the lack of resolution (57528 impacts per tree) is not the main reason for a lower correlation of LiDAR derived variables with fruit load. The lowest correlation comes from the different orientation of each sensor. While the RGB camera mounted on UAV (variable G) provides a top view of the canopy, MTLS provides a lateral view of the trees. The top view allows the area exposed to sunshine to be more precisely delimited. The efficiency when intercepting the sunrays is the key factor for transpiration and photosynthesis, therefore it is strongly related to the fruit load (Da Silva et al. 2014).

Figure 4A shows the canopy of a peach tree as it is captured with an aerial image from airplane (0.25 m per pixel), and the same canopy when it was captured using UAV and resolution of 0.02 m (Fig. 4B). The difference is obvious. While the delimitation of the canopy projected area was more difficult using the airplane image, the UAV-based canopy projected area allowed the actual canopy projected area to be better defined. This would explain why the use of higher resolution images should lead to more accurate tree ranking processes and more efficient fruit load estimates. To provide an explanatory example, 45 pixels were needed on average to characterize the canopy projected area using the airplane image instead of the 6500 pixels for the UAV image. In other words, the average area of 2.60 m^2 per tree using the higher resolution image became an area of 2.81 m^2 when a larger pixel size was used. The difference was therefore 8%, although large deviations for some trees could be obtained (maximum differences of up to 53% between UAV and airplane were found). Because of that, the use of images of poor resolution can alter the ranking process. In fact, ranking of individuals varied (data not shown) according to the resolution of the ancillary variable that was used (for example, G or K). The moderate Spearman correlation coefficient (0.71) between both variables allowed this property to be concluded.

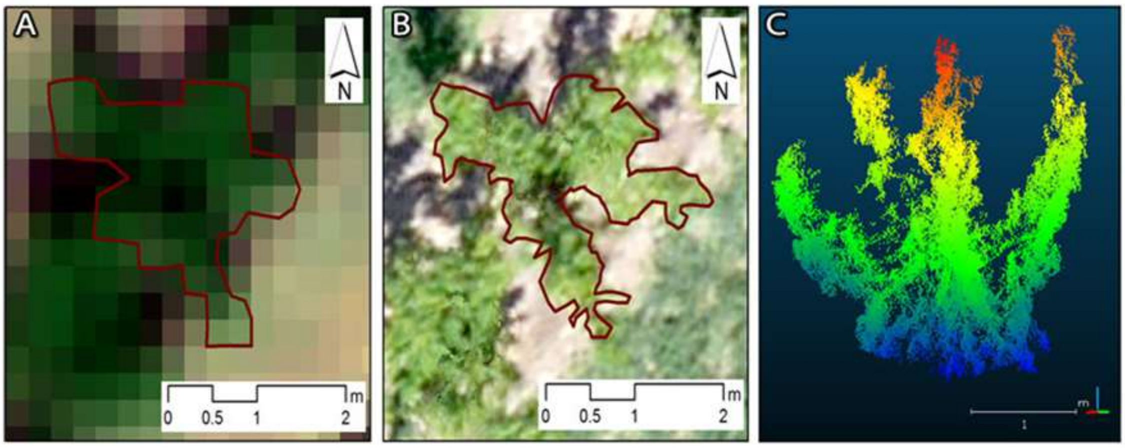


Fig. 4. Comparison of tree canopy projected area delimited for the same peach tree, A) using an airplane-based image of 0.25 m per pixel (top view), B) using an UAV-based image of 0.02 m per pixel (top view), C) 3D reproduction of the same canopy using an MTLS (side view).

MTLS are another different technology that allows canopy of trees to be characterized with high operative resolution. Figure 4C shows the point cloud obtained from a lateral LiDAR-based scan of the same tree. Then, by 3D processing of all the data, a canopy volume expression can be obtained (variable O, Table 2). However, this parameter did not provide results as good as those obtained using the simplest UAV-based tree canopy projected area (variable G). Finding an explanation is not easy. Probably, sunlight penetration is optimal in trees with more open canopy resulting in a larger projected area. Conversely, bulky trees with smaller projected areas could have grown in height instead of laterally and more openly. This could hinder the interior lighting of the canopy and negatively affect the amount of viable fruits and yield.

Figure 5 helps to understand this phenomenon. While the two trees shown (number 104 and number 45) have similar canopy volume (3.91 m^3 and 3.83 m^3, respectively), the canopy projected area (3.92 m^2 and 2.40 m^2) and fruit load (201 fruits and 124 fruits) are very different. In fact, the correlation between canopy projected area (variable G) and fruit load is higher (R = 0.85) than the correlation between canopy volume (variable O) and fruit load (R = 0.68).

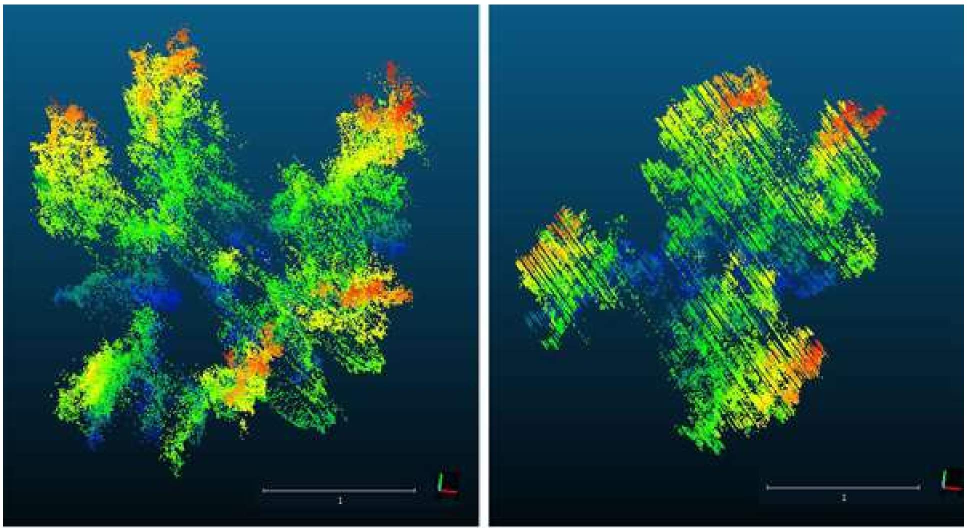


Fig. 5. Aerial projection of tree 104 (left) and 45 (right) obtained from the LiDAR-based MTLS system.

The use of NDVI showed acceptable results when adjusting the value of the vegetation index according to the tree canopy projected area. However, the exclusive use of the airplane NDVI (variable H, Table 2) was the least recommended option. The resolution of the airplane images was lower than that of UAV images or MTLS point clouds and, in addition, the reflectance alone does not account for any geometrical parameters of the canopy. In any case, the weak correlation between fruit load and NDVI is also noted in other sampling studies (Arnó et al. 2017; Miranda et al. 2018).

Concluding, resolution of acquired ancillary data was the main constraint in reliably delimiting canopy of trees (Fig. 4). UAV-based RGB images offered high resolution at competitive costs, and good correlation with the fruit load (Table 2). Therefore, RSS was only tested with UAV-based tree canopy projected area as the ancillary information. Comparison of sampling efficiency against SRS is addressed in the next section. The pending issue is to test the method in other orchards trained under systems that are more intensive forming a fruiting wall structure. Probably in this case, sections of wall vegetation rather than individualized trees should be delimited using an appropriate algorithm and a more automated method.

3.4. RSS versus SRS: sampling efficiency and sample size

Table 3 shows the mean squared error (*MSE*), and the relative efficiency (*RE*) of RSS versus SRS for different sample sizes (4 to 12 points or trees sampled within the plot). A resampling of 1000 samples allowed such statistics to be calculated in order to compare the sampling methods. RSS was run in all cases using the UAV-based tree canopy projected area as ancillary variable to drive the ranking mechanism.

Table 3. Mean squared error (*MSE*) and relative efficiency (*RE*) of the sampling methods for different sample sizes.

	Mean squared error (*MSE*)								
	N = 4	*N* = 5	*N* = 6	*N* = 7	*N* = 8	*N* = 9	*N* = 10	*N* = 11	*N* = 12
SRS	537.64	346.02	333.46	290.64	240.66	233.26	183	176.09	158.96
RSS	304.99	246.90	177.80	132.10	108.57	103.77	95.47	76.48	79.46
	Relative efficiency $(\boldsymbol{RE = var(\bar{Y}_{SRS})/var(\bar{Y}_{RSS})})$								
	N = 4	*N* = 5	*N* = 6	*N* = 7	*N* = 8	*N* = 9	*N* = 10	*N* = 11	*N* = 12
RSS vs. SRS	1.76	1.40	1.87	2.20	2.21	2.26	1.91	2.30	2.00

SRS (Simple random sampling), RSS (Ranked set sampling)

In both cases, SRS and RSS, the *MSE* decreased, as the sample size increased. As expected, the sampling methods became more efficient with an increasing sample size. However, comparatively, RSS always showed lower variance of the mean (lower *MSE*) than SRS. In short, RSS was more efficient than SRS for any of the sample sizes evaluated. Hence the *RE* was higher than one (Table 3), ranging from 1.40 (*N* = 5) to 2.30 (*N* = 11). The fact of obtaining higher values of *RE* for the larger sample sizes was possibly due to the RSS procedure. Takahashi and Wakimoto (1968) reported a loss of efficiency by decreasing the number of individuals checked in the ranking process for the same final sample size. This would explain why the *RE* of the RSS increases with increasing sample size because a higher number of trees are ranked. Thus, by increasing the sample size (*N* = 7, 8 ... 12), RSS required progressively increasing the number of trees to rank to obtain the final sample. For example, a sample of size *N* = 9 would suppose choosing 81 trees within the plot instead of the 25 (5^2) needed for *N* = 5. In this way, additional information on fruit load structure was more easily achievable by having to select and assess a larger number of trees than the one obtained in small samples. As a result, the *RE* of RSS increased compared to SRS strategy.

Considering the UAV-based canopy projected area as ancillary information, RSS allowed more accurate fruit load estimates to be obtained. The extent to which this precision was improved was quantified by the relative error of sampling (Eq. 6) as shown graphically in Fig. 6. Additionally, Figure 7 shows the effect of the sample size on the *PS* indicator for both sampling schemes. Both measures are important for fruit growers because they complement each other. Farmers collect samples at specific times during the life of the crop. Therefore, knowing the probability of exceeding the threshold and the average relative error for each sample size is important. Both curves (Figures 6 and 7) show a similar nonlinear trend. For all sample sizes, RSS has a lower error than SRS. Specifically, while in SRS the error ranged from 16.0% (*N* = 4) to 8.7% (*N* = 12), in RSS errors ranged from 12.1% (*N* = 4) to 6.2% (*N* = 12). Likewise, SRS schemes exceed the error threshold of 10% with a higher percentage (*PS*, Fig. 7) compared to RSS. While for SRS, *PS* values ranged from 52.2% (*N* = 4) to 22.3% (*N* = 12), in RSS the same parameter reached significantly lower scores, varying between 41.9% (*N* = 4) and 10.8% (*N* = 12). Therefore, using SRS practically doubles the probability of obtaining relative errors of estimation above the 10% threshold than using RSS. However, for smaller sample sizes (*N* = 4 and *N* = 5), the difference between SRS and RSS was not so obvious.

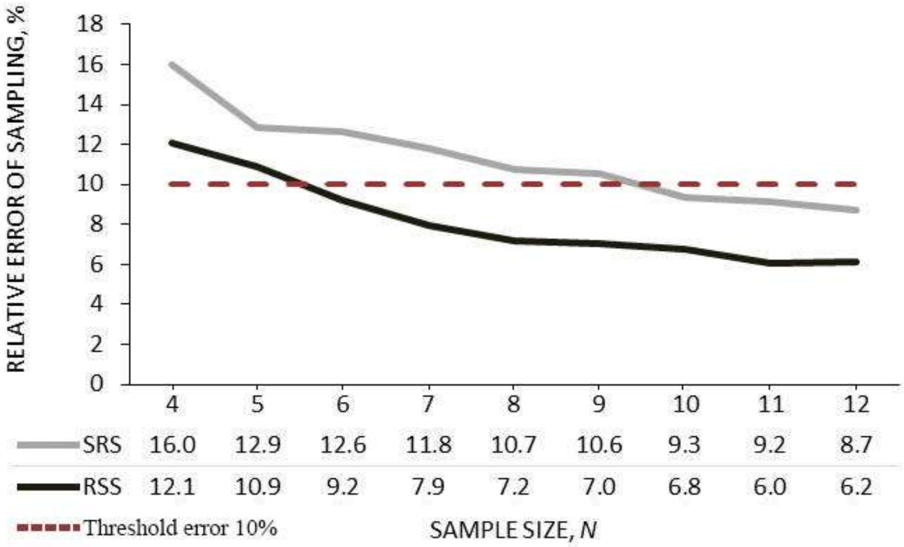


Fig. 6. Comparison of the relative error of sampling and sample size for two sampling methods: SRS and RSS. SRS (simple random sampling). RSS (ranked set sampling). Resampling of 1000 samples for each of the sample sizes was made, using the UAV-based tree canopy projected area as ancillary variable in the RSS.

Furthermore, if a relative error of no more than 10% was assumed, RSS allowed the sample size to be reduced to *N* = 6 while the SRS required a sample almost twice as much (*N* = 10) (Fig. 6). Moreover, a sample size *N* = 5 in RSS could even be justified by reaching an error of only 10.9%. Undoubtedly, being able to reduce the sampling requirements, RSS could report favorable operative as well as economic consequences in fruit sampling as those obtained by Carrillo et al. (2016) in vineyard. Although RSS requires random sampling in each iteration to generate the final ranked sample, subjectivity (and, therefore, biased decisions) should not be applied during the ranking mechanism favoring a more automated (without farmers rating) and reliable sampling process. On the other hand, the ancillary variable being available, the method is even applicable in small plots like fruit orchards (Zude-Sasse et al., 2016), avoiding to perform a stratified sampling scheme.

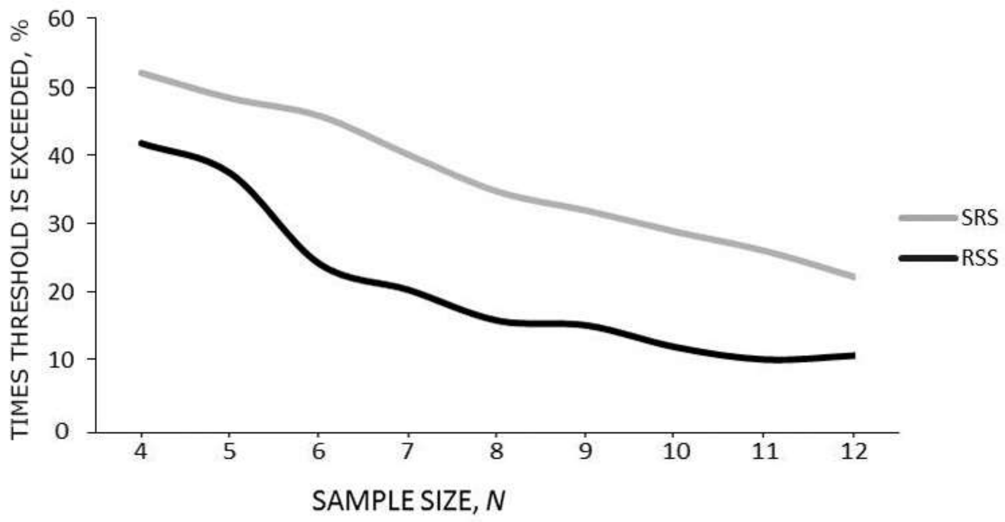


Fig. 7. Comparison of the percentage of times the error threshold of 10 % is exceeded when sample size increases for SRS and RSS.
SRS (simple random sampling). RSS (ranked set sampling). Resampling of 1000 samples for each of the sample sizes was made, using the UAV-based tree canopy projected area as ancillary variable in the RSS.

4. CONCLUSIONS

Ranked set sampling (RSS) is an interesting method to improve sampling in fruit growing. Improvements in the efficiency of peach fruit load estimates were proven by applying RSS compared to simple random sampling (SRS), at least for sample sizes from N = 4 to N = 12. However, to be optimal, RSS requires a relevant auxiliary variable, these data being the key factor in the procedure. Among different data sources currently available in agriculture, either commercialized or under development (such as MTLS), the UAV-based tree canopy projected area has been identified as the best option to use in the ranking mechanism aiming at estimate fruit number in peach production. This information can be obtained from RGB cameras that providing high-resolution images allowing the projected area of the canopy (geometric parameter) to be accurately delimited. The close correlation between this area and the fruit load per tree makes this auxiliary variable ideal in RSS schemes for peach trees in our conditions (open vessel like training systems).

In terms of relative error of sampling, RSS using samples of size N = 6 allows acceptable errors below 10% to be achieved. In contrast, SRS requires practically doubling the sample needed for a similar error. Finally, it should be noted that the work has been carried out only on one plot, although it provides relevant information as the ancillary variable to be used in RSS for future research on a larger spatial scale. In fact, UAV-based tree canopy projected area is an interesting (cheap and easy to get) ancillary information. However, since the values of the variable are obtained manually, future research is required to automatically obtain this information by developing the appropriate algorithm.

Acknowledgments

The Spanish Ministry of Economy and Competitiveness through the AgVANCE Project (AGL2013-48297-C2-2-R) funded this work. We are also grateful to Frutas Hermanos Espax SL for the possibility to carry out the research in their farm.

References

Aggelopoulou, K., Castrignano, A., Gemtos, T., Benedetto, D., 2013. Delineation of management zones in an apple orchard in Greece using a multivariate approach. Comput. Electron. Agric. 90, 119-130.

Aggelopoulou, K. D., Wulfsohn, D., Fountas, S., Gemtos, T.A., Nanos, G.D., Blackmore, S., 2010. Spatial variation in yield and quality in a small apple orchard. Prec. Agric. 11, 538-556.

Araya-Alman, M., Acevedo-Opazo, C., Guillaume, S., Valdés-Gómez, H., Verdugo-Vásquez, N., Moreno, Y., Tisseyre, B., 2017. Using ancillary yield data to improve sampling and grape yield estimation of the current season. Adv. Anim. Biosci. Precis. Agric. 8 (2), 515-519.

Arnó, J., Martínez-Casanovas, J.A., Uribeetxebarria, A., Escolà, A., Rosell-Polo, J.R., 2017. Comparing efficiency of different sampling schemes to estimate yield and quality parameters in fruit orchards. Adv. Anim. Biosci. Precis. Agric. 8 (2), 471-476.

Arnó, J., Rosell, J.R., Blanco, R., Ramos, M.C., Martínez-Casasnovas, J.A., 2012. Spatial variability in grape yield and quality influenced by soil and crop nutrition characteristics. Prec. Agric. 13 (3), 393–410.

Berman, M.E., De Jong, T.M., 1996. Water stress and crop load effects on fruit fresh and dry weights in peach (*Prunus persica*). Tree Physiol. 16, 859-864.

Carrillo, E., Matese, A., Rousseau, J., Tisseyre, B., 2016. Use of multi-spectral airborne imagery to improve yield sampling in viticulture. Prec. Agric. 17, 74-92.

Chen, Z., Bai, Z., Sinha, B.K., 2004. Ranked Set Sampling. Theory and Applications. Springer-Verlag, New York, pp. 224.

CloudCompare [GPL software] v2.6.1., (2015). http://www.cloudcompare.org. Accessed on July 25, 2017.

Cochran, W.G., 1977. Sampling Techniques. John Wiley & Sons, Inc, New York, NY, USA, pp. 428.

Da Silva, D., Han, L., Costes, E., 2014. Light interception efficiency of apple trees: a multiscale computational study based on MappleT. Ecol. Model. 290, 45-53.

Escolà, A., Martínez-Casasnovas, J.A., Rufat, J., Arnó, J., Arbonés, A., Sebé, F., Pascual, M., Gregorio, E., Rosell-Polo, J.R., 2017. Mobile terrestrial laser scanner applications in precision fruticulture/horticulture and tools to extract information from canopy point clouds. Prec. Agric. 18, 111–132.

Fountas, S., Aggelopoulou, K., Bouloulis, C., Nanos, G.D., Wulfsohn, D., Gemtos, T.A., 2011. Site-specific management in an olive tree plantation. Prec. Agric. 12, 179-195.

Kerry, R., Oliver, M. A., Frogbrook, Z.L., 2010. Sampling in Precision Agriculture. In M. A. Oliver (Ed.), Geostatistical Applications for *Precision Agriculture*. Springer, New York, pp. 36-63.

Käthner, J., Zude-Sasse, M., 2015. Interaction of 3D soil electrical conductivity and generative growth in Prunus domestica L. Eur. J. Hortic. Sci**.** 80 (5), 231-239.

Lelong, C.C.D., Burger, P., Jubelin, G., Roux, B., Labbé, S., Baret, F., 2008. Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. Sensors. 8 (5), 3557–3585.

Mann, K.K., Schumann, A.W., Obreza, T.A., 2011. Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data. Prec. Agric. 12, 457-472.

Marini, R.P., Sowers, D.S., Marini, M.C., 1995. Tree form and heading height at planting affect peach tree yield and crop value. HortScience. 30 (6), 1196–1201.

Marini, R.P., 2002. Heading fruiting shoots before bloom is equally effective as blossom removal in peach crop load management. HortScience. 37 (4), 642–646.

Martinez Vega, M.V., Clemmensen, L., Wulfsohn, D., Toldam-Andersen, T.B., 2013. Using multilevel systematic sampling to study apple fruit (Malus domestica Borkh.) quality and its variability at the orchard scale. Sci. Hortic. 161, 58-64.

McIntyre, G.A., 1952. A method for unbiased selective sampling, using ranked sets. Aust. J. Agric. Res. 3, 385-390. (Reprinted in Am. Stat. 2005. 59 (3), 230-232).

Meyers, J.M., Vanden Heuvel, J.E., 2014. Research Note. Use of Normalized Difference Vegetation Index Images to Optimize Vineyard Sampling Protocols. Am. J. Enol. Viticult. 65, 250-253.

Minas, I.S., Tanou, G., Molassiotis, A., 2018. Environmental and orchard bases of peach fruit quality. Sci. Hortic. 235, 307-322.

Miranda, C., Santesteban, L.G., Urrestarazu, J., Loidi, M., Royo, J.B., 2018. Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. Agriculture. 8 (6), 78.

Miranda, C., Urretavizcaya, I., Santesteban, L.G., Royo, J.B., 2015. Sampling stratification using aerial imagery to estimate fruit load and hail damage in nectarine trees. In: J. V. Stafford (Ed.) Precision Agriculture'15, Proceedings of the 10th European Conference on Precision Agriculture: Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 541-546.

Miranda, C., Royo, J.B., 2003. A statistical model to estimate potential yields in peach before bloom. J. Amer. Soc. Hort. Sci. 128, 297–301.

Monestiez, P., Audergon, J.M., Habib, R., 1990. Spatial dependences and sampling in a fruit tree: a geostatistical approach. Institut National de la Recherche Agronomique. Technical Report, pp. 163- 30.

Nahhas, R.W., Wolfe, D.A., Chen, H., 2002. Ranked set sampling: cost and optimal set size. Biometrics. 58, 964–971.

Pearce, S.C., 1944. Sampling methods for the measurement of fruit crops. J. R. Stat. Soc. 107 (2), 117-126.

Rosell, J.R., Llorens, J., Sanz, R., Arnó, J., Ribes-Dasi, M., Masip, J., Escolà, A., Camp, F., Solanelles, F., Gràcia, F., Gil, E., Val, L., Planas, S., Palacín, J., 2009. Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning. Agric. For. Meteorol. 149, 1505–1515.

Rouse Jr, J.W., Haas, R.H., Deering, D.W., Schell, J.A., Harlan, J.C., 1974. Monitoring the Vernal Advancement and Retrogradation (GreenWave Effect) of Natural Vegetation. Greenbelt, MD, USA, NASA/GSFC Type III Final Report, pp. 371.

Stokes, S.L., 2007. Ranked set sampling with concomitant variables. Commun. Stat. Theory. Methods. 6, 1207-1211.

Takahasi, K., Wakimoto, K., 1968. On unbiased estimates of the population mean based on the sample stratified by means of ordering. Ann. Inst. Stat. Math. 20, 1-31.

Tang, L., Hou, C., Huang, H., Chen, C., Zou, J., Lin, D., 2015. Light interception efficiency analysis based on three-dimensional peach canopy models. Ecol. Inform. 30, 60-67.

Taylor, J., Tisseyre, B., Bramley, R., Reid, A., 2005. A comparison of the spatial variability of vineyard yield in European and Australian production systems. In: J. V. Stafford (Ed.) Precision Agriculture'05, Proceedings of the 5th European Conference on Precision Agriculture: Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 907-915.

Ulzii-Orshikh, D., Lee, M., Sang-seok, Y., 2017. An yield estimation in citrus orchards via fruit detection and counting using image processing. Comput. Electron. Agric. 140, 103-112.

Underwood, J.P., Hung, C., Whelan, B., Sukkarieh, S. 2016., Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. Comput. Electron. Agric. 130, 83-96.

Uribeetxebarria, A., Daniele, E., Escolà, A., Arnó, J., Martínez-Casasnovas, J.A., 2018. Spatial variability in orchards after land transformation: Consequences for precision agriculture practices. Sci. Total. Environ. 635, 343–352.

Webster, R., Lark, R. M., 2013. Field sampling for environmental science and management. Routledge, London and New York, pp. 192.

Wolfe, D.A., 2004. Ranked Set Sampling: An Approach to More Efficient Data Collection. Stat. Sci. 19 (4), 636-643.

Wolfe, D.A., 2010. Ranked set sampling. Wiley. Interdiscip. Rev. Comput. Stat. 2 (4), 460-466.

Wulfsohn, D., Aravena Zamora, F., Potin Téllez, C., Zamora Lagos, I., García-Fiñana, M., 2012. Multilevel systematic sampling to estimate total fruit number for yield forecasts. Prec. Agric. 13, 256–275.

Zude-Sasse, M., Fountas, S., Gemtos, T.A., Abu-Khalaf, N., 2016. Applications of precision agriculture in horticultural crops. Eur. J. Hortic. Sci. 81, 78-90.

Chapter 7

General discussion

DISCUSSION

This Thesis has allowed increasing knowledge about two recurring themes of precision agriculture. First, information provided by terrestrial and aerial sensors has been used to study the spatial variability of soils and orchards of the central area of the Ebro Valley (two representative plots of peaches (*Prunus pérsica* (L.)Stokes), planted according the "Catalan" vessel training system). This information has enabled the elaboration of prescription maps, which allows proposing different to reduce the variability (Chapters 3 and 4). Secondly, and with a different objective, the information obtained from sensors has been used to propose more efficient sampling methods than simple random sampling (SRS) (Chapters 5 and 6). .

1.1. Spatial variability in fruit tree plots (peach): evaluation of land transformation on soil variation and its effects on the crop.

Moderate to high values of the Coefficient of Variation (CV) of field variables indicate a non homogeneous distribution of crop development within the fields. Thus, of the study plots, the one located in Gimenells had a CV of 41 % in the number of fruit per tree, while in the other location (Utxesa), the CV was 31%. These values are in agreement with the results presented by Aggelopoulou et al. (2010) y Fountas et al. (2011). Based on these data, and according to Kravchenko (2003), the spatial variation present in both plots makes them suitable for differentiated management.

The reason for within-yield variability is, frequently, soil variability (Mann et al., 2010). In the central area of the Ebro Valley, soil variability has been increased along years by soil movements carried out to adapt small terraced plots into larger plots. To capture spatial variability of agricultural soils is not a simple process, since obtaining and analyzing the samples is expensive, both in time and money. Therefore, soil variability is usually maped using sensors that measure proxy variables related to soil properties. The most used proxy variable for this purpose is the apparent electrical conductivity (ECa) (Corwin & Lesch, 2003). Normally, ECa is obtained from electromagnetic induction sensors (EMI) or galvanic contact sensors, like the Veris 3100 used in this research. One advantage of these sensors is that are relatively cheap and data is can be acquired easily. Mobility and adaptability to different terrains is other of advantages of these sensors (van Dam, 2012). Contact sensors present an advantage over the EMIs. The signal of the latter is affected by the presence of metal objects, such as irrigation pipes, while the signal of galvanic contact sensors is not.

The ECa signal is strongly correlated to several physic-chemical soil properties (Kafka et al., 2005). To interpret relationships between soils properties and the ECa is not simple, given the spatial dependence and interrelation of many of the properties usually analyzed in agriculture. Hence, this is the reason for using multivariate statistical methods to better interpretet these relationships. In soils with a high concentration of calcium carbonates, even with a petrocalcic horizon, the spatial variability pattern of ECa has not been studied in detail (Peralta et al., 2013) and, there are few studies focused on the analysis of these interactions (Domenech et al., 2017).

In chapter 3 of the Thesis, the problem of characterizing soil variability in a fruit orchard is addressed by the exclusive use of the ECa signal, measured with the Veris 3100soil surveyor. Different statistical and geostatistical techniques were used for data processing: i) Cluster analysis through the ISODATA algorithm to classify ECa and obtain homogeneous management zones. ii) Simulate ECa variability at different depths using the "InvVERIS" software (EMTOMO Lda. Lisbon, Portugal). This software is based on a nonlinear algorithm, smoothness-constrained (Monteiro Santos et al., 2011). The analysis of conductivity data with InvVeris allows performing a quasi-3D map to evaluate ECa variation at discrete depth intervals. iii) Multivariate analysis of variance (MANOVA) and subsequent descriptive discriminant analysis (DDA). The use of these techniques allowed knowing which soil properties affect ECa variability, and to recommend about adapt crop management to the plot variability. The methodology proposed by Moral et al. (2010) is similar to the one used in this section. But the chapter 3 of this Thesis provides a more detailed interpretation of ECa variation due to use of multivariate techniques described above and the quasi-3D model of the ECa variation in depth.

As a resultan interesting proposal is the creation of differentiated management zones based on estimation of ECa at 20-50 cm depth (Chapter 3). In this soil layer, the properties that most affected ECa were pH, organic matter content and the concentration of calcium carbonates (Chapter 3, table 2). As pH and organic matter are relatively stable over time (Shaner et al., 2008) and they contribute to soil fertility, the zones with different characteristics in these variables would allow a differentiated fertilization management (Peralta & Costa, 2013), reducing the contribution of N, as in the case described by (Koch et al., 2004, Aggelopoulou et al., 2011). Nitrogen is considered to be the nutrient that more affects quality and productivity of peach trees (Minas et al., 2018). As it is known, a low concentration of N reduces the size and quantity of fruits, besides affecting negatively certain parameters of quality. On the other hand, an excessive concentration of N reduces fruit quality, since hinder gets an ideal coloration. Moreover, excessive N fertilization implies a reduction of buds development for the next year (Daane et al., 1995). Given the importance of this element, the differentiated management of N within the orchard might bring the farmers closer to the objectives of more sustainable and competitive agriculture. In this way, the economic benefits can be increased at the same time that the environmental footprint associated with fruit growing will decrease.

Figures 6 and 4 of chapters 3 and 4 (ECa maps) show the differences created by land transformation processes made in previous decades. The variation of ECa, measured using a galvanic contact sensor has the ability to detect contrasts in soil conductivity. Both study plots showed different values where soil movements (land transformation) were carried out. This break in ECa signal continuity indicated that in these zones there were differences in soil properties. In addition to measure spatial variation of soil properties, in chapter 4 the effects of soil movements and land transformation on crop vigour were analysed.

The plot located in Utxesa (Chapter 4) was characterized using 3 different interrelated procedures; i) soil sampling (40 soil samples and subsequent laboratory analysis), ii) ECa sruvey using the Veris 3100 sensor, iii) airborne multispectral image acquisition and NDVI calculation. The analysis of relations between soil and crop produced, a priori, unexpected results since ECa and NDVI did not seem to be related. Knowing that ECa allows to detect zones with different fertility within a plot, it was expected that higher vigour trees would be located in more fertile areas (low ECa). But this was not been the case. The underlying explanation of this phenomenon was identified through a subsequent analysis. The lack of this relationship was due to drip irrigation. On one hand, ECa measures were made in the inter-row spaces where drip irrigation has not influence. On the other hand, the NDVI provides information about trees vigour, which is the reflection of the ecosystem where the roots inhabit (wet bulb), and this is influenced by fertigation. In short, information of both variables (ECa and NDVI), although measured in nearby locations, corresponds to soil properties that can vary significantly due to the different water regime with the one that they are managed.

Regarding the effects of soil movements on plot variability, the major differences between ECa and NDVI were observed in the areas where old terraces were located (figure 5 of chapter 4). These differences reinforce the hypothesis that both variables behave differently. Whereas the ECa showed discontinuities where old terraces were located, the NDVI did not reflect them. This is in line with the hypothesis of De Benedetto et al. (2013), who affirm that crop under irrigation are more influenced by water management than soil properties.

Chapter 4 of the Thesis highlights the importance that the knowledge of intra-plot variability has in fruit orchard management. Due to the small size of fruit orchards (Zude-Sasse et al., 2016), the spatial patterns of variability (Zaman & Schuman, 2006) can go unnoticed, making difficult to know them. The knowledge and identification of these spatial patterns allows a more adequate management according to the site-specific crop management (SSCM) approach. A clear example of possible benefits to orchard management according SSCM is this plot. Considering that a significant difference ($p < 0.05$) in fruit number (fruit load) between trees located in the high ECa zone and those located in the low ECa zone was found, it is noteworthy that, although there was not relationship between the ECa and the NDVI to individual level (tree), at the conglomerate level, clusters with different values of ECa, reflected significant differences in fruit production. Nevertheless, it is worth to mention that the difference between clusters was not excessive (3 kg/tree). Probably, if the plot was not managed by fertigation, the difference between clusters would be greater, since the production of the trees located in the worst soils would be lower. While those located in the high ECa zone area produced an average of 136 fruits, those located in the low ECa area produce 152, this represents a difference of 2700 kg/ha. In chapter 4 of the Thesis, it was proposed to solve the problem through the differentiated management of irrigation. As expected, in the high ECa cluster there were higher salt concentrations (Rhoades, 1999). Theoretically, a higher dose of irrigation can wash salt, diminishing its concentration (Munns, 2002). As the peach tree is sensitive to the salt concentration, the reduction of these improves the production. Results obtained in this part of the investigation are in agreement with those obtained by Lambert & Lowenberg-Deboer (2000) and Silva et al. (2007), who affirm that the expenses in the agriculture can be reduced when improving the management at the plot level.

1.2. More efficient sampling methods through the use of auxiliary information.

The knowledge of the phenological state of orchards and its variability can help to obtain more stable and higher quality crops (Mirjana & Vulić, 2005). The most precise way of knowing a parameter of a fruit is to measure it. But, to measure all the trees of a plot is not viable. On one hand, the investment in time and effort required for this task would make it totally disproportionate to the benefit obtained (Bongiovanni & Lafayette, 2004). On the other hand, in fruit growing, one of the fundamental applications of sampling is the estimation of yield in earlier phenological stages, previous to ripening (Elmendorf et al., 2016), which supposes sample destruction. A good option to solve this problem and know the phenological state of the orchard is by means of sampling. Theoretically, a good sampling has to represent all population with a few samples and this is not easy. Therefore, not all sampling techniques represent in the same way the phenological state of the orchard. In an instance, the characterization of the agricultural soil is a difficult and expensive process (Nawar et al., 2017), since the samples are collected and subsequently processed in the laboratory manually. The intrinsic cost of soil sampling together with the low efficiency of SRS used in this process means that the necessary information for the differentiated management of the plots is not obtained (Franzen et al., 2002). Since the SRS does not take into account spatial variability (Webster and Lark, 2013), an option to capture soil variability by SRS would be increasing the sampling density (Bramley & Janik, 2005). This would increase the effort significantly. Likewise, orchard characterization presents a similar problem since it also usually shows well-defined spatial patterns (Colaço et al., 2018).

Another drawback that makes doubt the samplers is to decide the number of necessary samples and where to place the sampling points to capture plot's variability. When there are not evidences about which can be the variation, an option that allows to approach this problem is the nested survey. Kerry et al. (2010) proposed to use this method to obtain maximum distance between samples. Then, taking this distance into account, realize a systematic sampling. Using this procedure, the sampler solves the problems of where and how many samples must be collected. But when the plot variability is high, the number of samples needed to capture the variability is big, since the distance between the samples is small. Unlike SRS, one problem of systematic sampling is that it is biased. In spite of, SRS is less accurate than systematic sampling.

The use of an auxiliary variable strongly correlated with the target variable opens the door to know the spatial variability of the plot more easily. Ideally, the auxiliary variable must gather the following characteristics: i) be easy to measure, ii) cheap, iii) strong spatial correlation, iv) be geolocated. Certain auxiliary variables, especially those obtained by sensors, allow to obtain many records correlated with the variable to be estimate, which facilitates the capture of spatial variability. Moreover, these variables can be used to define homogeneous management zones (Arnó et al., 2011, Córdoba et al., 2016). With all this, the usefulness of this type of variables at the time of sampling soil (Brogi et al., 2019), or vegetative (Carrillo et al., 2019) properties is beyond doubt.

In the last years, different authors have proposed different techniques to solve deficiencies of SRS and, in turn, increase sampling efficiency. These techniques can be grouped according to two different conceptual areas. On the one hand, there is the so-called "design-based" approach, which encompasses sampling based on classical statistics and assumes the independence among samples (Cochran, 1977). Another approach is known as “model-based", which is based on the principles of the geostatistics (Oliver, 2010). The latter assumes that the closest samples are more similar that those that are more distant. Therefore, there is no independence between the samples. Within the scientific community, there are defenders and detractors of both methods, as shown in the discussion article of Brus & Gruijter (1997).

Chapters 5 and 6 of this Thesis have focused on the comparison of stratified random sampling (StRS) (Chapter 5) and ranked set sampling (RSS) (Chapter 6) versus the classic SRS. Both, the StRS and the RSS have been selected against other sampling techniques because they take into account the spatial variability patterns. Both sampling schemes are part of “design-based” group. Nevertheless, the geostatistical analysis has occupied also a preferential place during the Thesis, using spatial interpolation (kriging) as a preferential option for obtaining surface maps (continuous raster coverage). Moreover, kriging has been a fundamental tool in the chapter 5, since it is the previous step to classify the plot in two or three clusters to stratify the sample.

In chapter 5, efficiency of stratified sampling (StRS) versus simple random sampling (SRS) was compared. The Gimelles study plot was clustered taking into account NDVI and ECa as auxiliary information. The relation between NDVI and tree vigour is contrasted by several studies and the ECa is the most used variable to characterize the variability of soil properties. The choice of these two variables is justified by being correlated with the target variables to estimate: i) yield (kg/tree), ii) refractometric index (°Baumé), iii) fruit firmness (kg/cm^2). The existence of a spatial correlation between the auxiliary variables and the target ones is essential for a good stratification and improve the results of the SRS since it enables design homogeneous strata. An accurate estimation of the population can be obtained using a small sample size in each stratum. The algorithm used to delimit the strata has been the ISODATA (Jensen, 1996). In chapter 5, the plot was into 2 or 3 strata, selecting in each stratum 6 or 4 trees respectively, without taking into account stratum size. The result of this identical allocation of sampling points in each stratum was a non-optimal point distribution, which could imply a lower sampling efficiency. A pending task to reduce sampling variance and improve the results in future works is to assign the sampling points in concordance to strata size. Despite this, chapter 5 confirmed the premise of Cochran (1977), who sets up that the variance estimated by StRS is lower than variance estimated by SRS for the same sample size. All parameters were better forecasted when used 2 strata of NDVI. The Gimenells plot, such as Utxesa plot, was fertigated by dripping, which creates a micro-habitat where roots are placed. This might be the reason why NDVI stratification obtained better results than ECa stratification. The reduction of sampling effort for the same precision ranged between 17% for the yield and 73% for the refractometry.

In chapter 6, efficiency of ranked set sampling to estimate fruits number was evaluated. This variable is essential in fruiticulture since it is the main component of yield (Miranda & Royo, 2003). With a strong correlation of R = 0.85, the best auxiliary variable to estimate fruit number in peach trees in “Catalan” vessel training system was the canopy projected area, obtained by means of a drone image. Furthermore, and in concordance with results obtained in chapter 4, the ECa was not a good auxiliary variable to estimate fruits number, since it behaves independently. As mentioned previously, this could be because of the drip irrigation system. The accuracy of the RSS scheme was compared respect the SRS for different sample sizes (N=4, N=12). For all sample sizes, the relative error of the RSS was less than SRS. Moreover, SRS required 10 samples to obtain an average error lower than 10%, while RSS only required 6. This was translated to an increase of 40% in the effort of sampling for the same estimation error. Besides the average error, the probability of overcoming 10 % error in each sampling was calculated. For this parameter, the RSS showed a better behaviour than SRS for all of sample sizes.

As mentioned by Gebbers & Adamchuk (2010), PA is a set of techniques that allow to apply the appropriate treatment, in the right place and at the right time. In this point, sampling plays a crucial role. It is essential to know a phenological state of the crop and act in consequence. The precise estimation of yield also requires knowing ahead in time the spatial variability attached to certain factors that determine the final yield of a fruit tree. Chapters 5 and 6 proposed two sampling schemes that make it possible to reduce the effort respect to SRS, on having increased the knowledge about crop phenology and the (spatial) distribution of its variability. Less effort in sampling implies, a lower economic cost, and this is of vital importance in a market such as agriculture, where the profit margin and profitability must be valued with increasing rigor. Fruit growers, apart from obtaining good crop estimates through increasingly more efficient sampling techniques, are also interested in knowing the possible variability of fruit quality parameters and estimate conveniently these properties in order to obtainhigher quality products with greater market value. Apart from optimizing sampling, knowing this variability can allow fruit growers to take the optimal decisions for harvest management (Taylor et al., 2007). Finally, another area where samplings are essential is the agricultural insurance. The damages suffered in a crop have been quantified historically quantified by means of a SRS. Miranda et al. (2018) have proposed recently the StRS as an alternative to SRS to quantify damages in agriculture, in consideration of improves the efficiency of the process. As mentioned previously, a more efficient sampling, improve economic results of fruit growers, by reducing effort and sample size necessary for acceptable error.

References

Aggelopoulou, K., Wulfsohn, D., Fountas, S., Gemtos, T.A., Nanos, G.D., Blackmore, S. 2010. Spatial variability of yield and quality in a small apple orchard. *Precis. Agric*. 11,538-556.

Aggelopoulou, K.D., Pateras, D., Fountas, S., Gemtos, T.A., Nanos, G.D. 2011. Soil spatial variability and site-specific fertilization maps in an apple orchard. *Precis. Agric*. 12, 118-129.

Arno, J., Martínez-Casasnovas, J.A., Ribes-Dasi, M., Rosell, J.R. 2011. Clustering of grape yield maps to delineate site-specific management zones. *Span. J. Agric*. Res. 9, 721–729.

Bongiovanni, R., Lafayette, W. 2004. Precision agriculture and sustainability. *Precis. Agric.* 5, 359–387.

Bramley, R.G.V., Janik, L.J. (2005). Precision Agriculture Demands a New Approach to Soil and Plant Sampling and Analysis - Examples from Australia. *Commun, Soil Sci, Plant Anal*. 36, 9–22.

Brogi, C., Huisman, J.A., Pätzold, S., von Hebel, C., Weihermüller, L., Kaufmann, M.S., van der Kruk, J., Vereecken, H. 2019. Large-scale soil mapping using multi-configuration EMI and supervised image classification. *Geoderma.*335 133-148.

Brus, D.J., de Gruijter, J.J. 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma.* 80, 1–44.

Carrillo, E., Matese, A., Rousseau, J., & Tisseyre, B. (2016). Use of multi-spectral airborne imagery to improve yield sampling in viticulture. *Precision. Agric*. 17, 74–92.

Cochran, W.G. 1977. Sampling Techniques. New York, NY, USA: John Wiley & Sons, Inc. 428 pp.

Colaço, A.F., Molin, J.P., Rosell-Polo, J.R. Escolà. A. 2018. Spatial variability in commercial orange groves. Part 2: relating canopy geometry to soil attributes and historical yield. *Precis. Agric.* 18, 1-18.

Córdoba, M.A., Bruno, C.I., Costa, J.L., Peralta, N.R., Balzarini, M.G. 2016. Protocol for multivariate homogeneous zone delineation in precision agriculture. *Biosyst. Eng.* 143, 95-107.

Corwin, D.L., Lesch, S.M. 2003. Application of soil electrical conductivity to precision agriculture: theory, principles and guidelines. *Agron. J*. 95: 455-471.

Daane, K.M., Johnson, R.S., Michailides, T.J., Crisosto, C.H., Dlott, J.W., Ramirez, H.T., Yokota, G.T., Morgan, D.P., 1995. Excess nitrogen raises nectarine susceptibility to disease and insects. *Calif. Agric*. 49, 13–17.

De Benedetto, D., Castrignano, A., Diacono, M., Rinaldi, M., Ruggieri, S., Tamborrino, R. 2013. Field partition by proximal and remote sensing data fusion. *Biosyst. Eng.* 114, 372–383.

Domenech, M.A., Castro-Franco, M., Costa, J.L., Amiotti N.M. 2017. Sampling scheme optimization to map soil depth to petrocalcic horizon at field scale. *Geoderma*. 290, 75-82.

Drewnowski, A., Popkin, B.,M. 1997 The nutrition transition: new trends in the global diet. *Nutr. Rev*. 55, 31–43.

Elmendorf, S.C., Jones, K.D., Cook, B.I., et al. 2016. The plant phenology monitoring design for the national ecological observatory network. *Ecosphere*. 7

Far, S.T., Rezaei-Moghaddam, K. 2017. Impacts of the precision agricultural technologies in Iran: an analysis experts' perception & their determinants. *Inf. Process. Agric.* 1–12.

Fountas, S., Aggelopoulou, K., Bouloulis, C., Nanos, G.D., Wulfsohn, D., Gemtos, T.A., et al. 2011. Site-specific management in an olive tree plantation. *Precis. Agric*. 12, 179–195.

Franzen, D.W., Hopkins, D.H., Sweeney, M.D., Ulmer, M.K., Halvorson, A.D. 2002. Evaluation of soil survey scale for zone development of site-specific nitrogen management. *Agron. J.* 94, 381–389.

Gebbers, R., Adamchuk, V.I. 2010. Precision Agriculture and Food Security. *Science.* 327, 828-831.

Griffin, T., D. Lambert, J. Lowenberg-DeBoer, (2005), Economics of Lightbar and Auto-Guidance GPS Navigation Technologies, in Precision Agriculture '05, J.V. Stafford, ed. Wageningen Academic Publishers, Wageningen, Netherlands

Jensen, J.R. 1996. Introductory Digital Image Processing: Remote Sensing Perspective. 2nd ed. Prentice-Hall, Englewood Cliffs, NJ, USA.

Kaffka, S.R., Lesch, S.M., Bali, K.M. Corwin, D.L. 2005. Site-specific management in salt-affected sugar beet fields using electromagnetic induction. *Comp. Elect. Agric*. 46, 329-350.

Kerry, R., Oliver, M.A., Frogbrook, Z.L. 2010. Sampling in Precision Agriculture. Sampling in precision agriculture. M.A. Oliver (Ed.), Geostatistical Applications for Precision Agriculture, Springer Science Business Media B.V (2010), pp. 35-63.

Koch, B., Khosla, R., Frasier, W.M., Westfall, D.G., Inman. D. 2004. Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agron. J*. 96 , 1572-1580.

Kravchenko, A.N. 2003. Influence of spatial structure on accuracy of interpolation methods. *Soil. Sci. Soc. Am. J.* 67, 1564–1571.

Lambert, D., Lowenberg-Deboer, J. 2000. Precision agriculture profitability review Purdue, USA: Site Specific Management Center. P. 154.

Mann, K.K., Schumann, A.W., Obreza. T.A. 2010. Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data. *Precis. Agric*. 4, 457-472.

Martinez Vega, M.V., Clemmensen, L., Wulfsohn, D., Toldam-Andersen,T.B. 2013. Using multilevel systematic sampling to study apple fruit (Malus domestica Borkh.) quality and its variability at the orchard scale. *Sci. Hortic*. 161, 58-64.

Minas, I.S., Tanou, G., Molassiotis, A., 2018. Environmental and orchard bases of peach fruit quality. *Sci. Hort*. 235, 307–322

Miranda, C., Santesteban, L.G., Urrestarazu, J., Loidi, M., Royo, J.B. 2018. Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. *Agriculture*. 8, 78.

Miranda, C., Royo, J.B. 2003. A statistical model to estimate potential yields in peach before bloom. *J. Amer. Soc. Hort. Sci*. 128, 297–301.

Mirjana, R., Vulić., T. 2005. Important of phenological observations and predictions in agriculture. *J. Agric. Sci*. 50, 217-225.

Monteiro Santos, F.A., Triantafilis, J., Bruzgulis, K. 2011. A spatially constrained 1D inversion algorithm for quasi-3D conductivity imaging: application to DUALEM-421 data collected in a reverine plain. *Geophysics.* 76 (2), 43–55.

Moral, F.J., Terrón, J.M., Marques da Silva, J.R.M. 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil Tillage Res.* 106,335–343

Munns, R. 2002. Comparative physiology of salt and water stress. *Plant Cell Environ*. 25,239–50.

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., Mouazen, A.M. 2017. Chapter Four - Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. In:Sparks, D.L. (Ed.), *Adv. Agron*. Academic Press.175–245.

Oliver, M.A. 2010. Geostatistical Applications for Precision Agriculture, Ed. M.A. Oliver

Peralta, N.R., Costa, J.L., Balzarini, M., Angelini, H. 2013. Delineation of management zones with measurements of soil apparent electrical conductivity in the southeastern pampas. *Can. J. Soil. Sci.* 93, 205-218.

Peralta, N.R., Costa, J.L. 2013. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Comput. Electron. Agric.* 99, 218-226.

Rhoades, J.D., Chanduvi, F., Lesch, S.M. 1999. Soil salinity assessment. Methods and interpretation of electrical conductivity measurements. FAO Irrigation and Drainage Paper 57. FAO, Rome, Italy.

Shaner, D.L., Brodahl, M.K., Buchleiter, G.W. 2008. How well does zone sampling based on soil electrical conductivity maps represent soil variability? *Agron. J.* 100, 1472-1480.

Silva, C. B., Vale, S. M. L. R., Pinto, F. A. C., Muller, C. A. S., & Moura, A. D. (2007). The economic feasibility of precision agriculture in Mato Grosso do Sul State, Brazil: A case study. Precision Agriculture, 8, 255–265.

Taylor, J.A., Praat, J.P., Bollen, A.F. 2007. Spatial variability of kiwifruit quality in orchards and its implications for sampling and mapping. *Hortscience*. 42,246–250.

van Dam, R. 2012. Landform characterization using geophysics—recent advances, applications, and emerging tools. *Geomorphology.* 137, 57–73.

Webster, R., Lark, R.M. 2013. Field sampling for environmental science and management (p. 192). London and New York: Routledge

Zaman, Q.U., Schumann, A.W. 2006. Nutrient management zones for citrus based on variation in soil properties and tree performance. *Precision Agric.* 7, 45–63.

Zarco-Tejada, P.J., Hubbard, N., Loudjani, P. 2014. Precision Agriculture: an opportunity for EU farmers - potential support with the cap 2014-2020. FAO.

Zhang, N., Wang, M., Wang, N., 2002. Precision agriculture: a worldwide overview. *Comput. Electron. Agric.* 36, pp. 113-132.

Zude-Sasse, M., Fountas, S., Gemtos, T.A., Abu-Khalaf, N., 2016. Applications of precision agriculture in horticultural crops. *Eur. J. Hortic. Sci.* 81:78–90.

Chapter 8

Conclusions

Photo author: Smihael

CONCLUSIONS

This Thesis analysed spatial variability in peach orchards located in the central part of the Ebro Valley by means of sensors and technologies commonly used in the frame of Precision Agriculture. In the first part, sensor's information was used to delineate differentiated management zones within the plots. In the second part, sensor's information was used to improve conventional sampling schemes by taking into account such spatial orchard variability. The hypotheses proposed in chapter 1 have been solved in the course of the Thesis. Below, general conclusions drawn during the research are presented:

C1. On-the-go soil apparent electrical conductivity (ECa) measured by the galvanic contact sensor Veris 3100 is influenced by presence of high contents of calcium carbonate ($CaCO_3$) in the soil. Due to this, spatial variability of carbonates can be detected and quantified (Objective 1). ECa analysis using multivariate techniques (cluster analysis combined with multivariate analysis of variance) allows soil properties responsible for the significant variation of the ECa to be identified. This detailed spatial knowledge of some soil properties can help to make optimal decisions for plot management, based on delimiting areas with different soil properties or different productive potential.

C2. ECa shows discontinuity in areas where old terraces from previous plots were situated. This suggests that soil properties were altered when reparcelling new larger plots. In any case, ECa signal allows footprint by land transformation to be conveniently detected (Objective 2). Concerning the Normalised Difference Vegetation Index (NDVI), it is difficult to trace such plot transformations, since this index did not show spatial correlation with ECa in the analysed plot. Possibly, canopy variation (or vigour/ NDVI variation) is mostly due to the use of a drip irrigation system, which creates a wet micro-habitat in the root development zone.

C3. Compared to simple random sampling (SRS), stratified random sampling (StRS) allows sampling efficiency to estimate yield and quality parameters in fruit orchards (Objective 3) to be improved. For that, and to get the best results, NDVI was used as auxiliary information. Specifically, stratifying sampling points according to homogeneous NDVI areas allows, for the same precision, to reduce by 20% the sample size as compared to SRS. In any case, the ECa or the combined use of NDVI and ECa have provided substantial advantages as compared to the use of NDVI as a single layer of ancillary information.

C4. In fruit orchards, ranked set sampling (RSS) manages to improve efficiency as compared to simple random sampling (SRS), at least for sample sizes from N=4 to N=12 (Objective 4.1). Having an auxiliary variable strongly correlated with the target variable (fruit load, in the present case), is key to success using ranked samples. UAV-based tree canopy projected area has been identified as the best option as ancillary variable to estimate fruit load in peach orchards, when "Catalan vessel" is the used training system (Objective 4.2). Sample size can be reduced by 40% using RSS compared to SRS, without exceeding sampling errors of 10%. This provides a scheme to optimize sampling costs in fruit growing.

In this Thesis, research was focused on peach orchards under “Catalan” vessel training system, where individual canopies are easily identified and delimited. The challenge for future works is to investigate the proposed methods applying the know-how learnt in this Thesis to other fruit species trained in vegetation walls and more intensive plantation patterns.

The aim of the present PhD Thesis is the analysis of spatial variability in peach fruit orchards of the central area of the Ebro Valley. In the first part, soil variability and tree vigour have been studied to obtain homogeneous management zones. In the second part, advanced sampling schemes, as stratified random sampling (StRS) and ranked set sampling (RSS) has been used to estimate different parameters. Auxiliary information provided by several sensors nourishes the sampling schemes. During the research, different multivariate and geostatistical analysis techniques have been implemented to extract information from the data.