

Mètodes de mostreig per a l'estimació de la collita en fructicultura

Mida de la mostra, precisió del mostreig i Coeficient d'Error

La mida de la mostra influeix l'exactitud i la precisió de les estimacions. Cas que els arbres (on es farà el compteig de fruits) se seleccionin mitjançant un mostreig aleatori simple, l'equació (1) permet establir la mida de la mostra o número d'unitats de mostreig (arbres) que es necessita fixant un determinat error de mostreig (també anomenat percentatge d'error, PE) i un nivell de confiança del 95% (Anderson et al., 2018),

$$n = \frac{4 \times CV^2}{PE^2} \quad (1)$$

on n és el nombre d'arbres a mostrejar, i CV és el Coeficient de Variació de la collita (càrrega de fruits per arbre) el qual ha estat prèviament establert (o és un paràmetre conegut per l'experiència en la zona). Tal i com es desprèn de l'equació (1) i a diferència de la recomanació mostrada en la Taula 1, queda clar que la mida de la mostra no està en principi influenciada pel número d'arbres (o collita esperada), sinó més aviat per la variabilitat que presenta la collita (càrrega de fruita) dins la parcel·la (CV) i per l'error màxim que el fructicultor i/o tècnic estan disposats a assumir (PE).

Per altra banda, la precisió d'un mètode de mostreig s'estableix mitjançant la variància de l'estimador. En el cas que ens ocupa (estimar la càrrega de fruita), l'estimador fa referència a la càrrega mitjana de fruita per arbre (o mitjana de la mostra). Es demostra que, en un mostreig aleatori simple (d'ara en endavant simple random sampling, SRS), la variància de la mitjana mostral s'aproxima amb l'equació (2) (Webster & Lark, 2013),

$$v(\hat{Y}_{SRS}) = \frac{s^2}{n} \times \left(1 - \frac{n}{N}\right) \quad (2)$$

sent aquesta fórmula aplicable a poblacions finites (o plantacions d'arbres fruiters) amb un número N total d'arbres i variància mostral, s^2 . L'arrel quadrada de (2) és l'anomenat error estàndard de l'estimador (o mitjana mostral, \hat{Y}_{SRS}), proporcionant la inversa d'aquest valor una mesura de la



precisió o eficiència del mètode (Peña, 2005). Alternativament al percentatge d'error (PE) i el nivell de confiança (95%) que han estat abans comentats, el Coeficient d'Error (CE) compara l'error estàndar amb el valor real esperat de la collita (Eq. 3):

$$CE(\hat{Y}) = \frac{\sqrt{v(\hat{Y})}}{Y} \times 100. \quad (3)$$

Assumint normalitat, aquest estadístic proporciona una informació força interessant al ser esperable que, en almenys el 68% de les mostres que es realitzin, l'error de mostreig o PE(%) no sobrepassi el valor especificat de CE(%). Per a una probabilitat del 95%, l'equació (4) mostra la relació entre ambdós estadístics:

$$CE(\%) = \frac{PE(\%)}{1,96} \quad (4)$$

El sector frutícola demanda mètodes de mostreig que permetin assolir valors de PE del 10% amb una confiança (probabilitat) del 95% (Miranda et al., 2018). Això és equivalent a exigir mètodes de mostreig que garanteixin valors de l'estadístic CE de aproximadament el 5%. Aquest requisit pot comportar tenir que mostrejar un gran nombre d'arbres en parcel·les amb gran variabilitat, tal i com mostra la Figura 1. D'aquí la necessitat de desenvolupar mètodes avançats de mostreig que, assegurant el valor acceptat de PE del 10%, suposin tenir que mostrejar únicament un nombre limitat d'arbres (entre 5 i 25, com a primera aproximació).

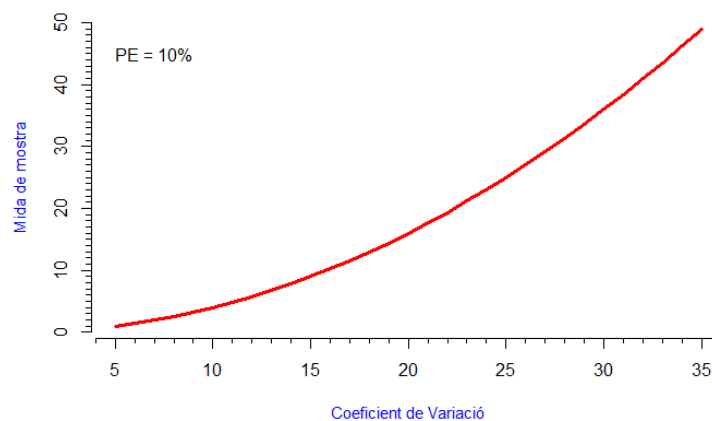


Figura 1. Mida de mostra en un mostreig aleatori simple per a un percentatge d'error (PE) del 10 % i probabilitat (confiança) del 95 %.

Mostreig estratificat

Malgrat que el mostreig aleatori és un mètode prou conegut, pot presentar l'inconvenient de requerir mides de mostra molt grans si les parcel·les



presenten una variabilitat espacial significativa (Figura 1). A més, és sempre inevitable un cert biaix per part dels fructicultors i/o tècnics a l'hora de seleccionar els arbres a mostrejar, sent per tant qüestionable la representativitat de les unitats de mostreig.

Una manera de guanyar precisió, i així optimitzar el mostreig, és estratificar la mostra. Per a que això sigui possible, és necessari disposar d'informació auxiliar o complementària de la plantació (normalment, una característica vegetativa del dosser foliar) que, i) sigui relativament fàcil d'obtenir, ii) estigui altament correlada amb el paràmetre de collita que es vol estimar i, iii) proporcioni la informació a una resolució espacial adequada.

Assumida la correlació positiva entre la càrrega de fruits i la mida dels arbres, la possibilitat de classificar-los per mida o volum en diferents estrats hauria de permetre seleccionar unitats de mostreig més representatives de la parcel·la, i així poder obtenir una estimació més acurada de la collita. Tal i com destaquen Webster y Lark (2013), l'avantatge d'estratificar resulta de la menor variabilitat que és esperable dins els estrats (o classes) en comparació a la variabilitat de la plantació en el seu conjunt. Amb la condició de ponderar la mitjana estimada de cada estrat segons l'àrea coberta dins la parcel·la, l'error estàndard¹ de la mostra estratificada es pot aproximar ara amb l'equació (5),

$$v(\hat{Y}_{STRS}) = \sum_1^K \frac{w_k^2 \times s_k^2}{n_k} \times \left(1 - \frac{n_k}{N_k}\right) \quad (5)$$

on $w_k = N_k/N$ és el factor de ponderació per a cada estrat o relació entre el número d'arbres (o unitats de mostreig) dins l'estrat k (N_k), i el número total dins la parcel·la (N), s_k és la variància dins de cada estrat, i n_k és el número d'arbres a mostrejar segons l'estrat corresponent. Operant d'aquesta manera, és esperable que l'estimació de la collita (càrrega de fruits) sigui més precisa si, com ja ha estat mencionat, la variable auxiliar utilitzada per classificar els arbres està espacialment correlada amb la collita. Llavors, la comparació entre dos mètodes de mostreig, per exemple, mostreig estratificat i mostreig aleatori, és només qüestió de calcular l'anomenada

¹ La variància d'una mostra estratificada ha de ser, en teoria, menor que la variància d'una mostra aleatòria.

Agafant com a mitjana estimada en un mostreig estratificat el valor $\hat{Y}_{STRS} = \frac{1}{K} \sum \hat{Y}_k$, la variància d'aquest estimador és $Var(\hat{Y}_{STRS}) = \frac{1}{K^2} \sum Var(\hat{Y}_k) = \frac{1}{K^2} \sum \frac{s_k^2}{n_k}$.

Operant adequadament i assumint que es mostreja el mateix número d'arbres per estrat ($n' = n_k$ sent $\sum n' = n$), s'obté l'expressió $Var(\hat{Y}_{STRS}) = \frac{1}{K^2} \sum \frac{s_k^2}{n_k} = \frac{1}{n} \left(\frac{1}{K} \sum s_k^2\right)$. El valor dins el parèntesi (mitjana de les variàncies de classe) ha de ser menor o igual que la variància de la mostra en el seu conjunt, resultant que $Var(\hat{Y}_{STRS}) = \frac{1}{n} \left(\frac{1}{K} \sum s_k^2\right) \leq \frac{s^2}{n} = Var(\hat{Y}_{SRS})$. En la demostració no s'ha tingut en compte la correcció per població finita.



eficiència relativa entre ambdós mètodes, és a dir, el quocient entre les precisions corresponents (o inverses dels errors estàndard), $RE = v(\hat{Y}_{SRS})/v(\hat{Y}_{STRS})$. En teoria, al comparar el mostreig estratificat amb l'aleatori cal esperar valors de RE majors que 1, la qual cosa permetria reduir la mida total de la mostra en una proporció igual a $100 \times (1 - 1/RE)$.

Exemples recents d'aplicació del mostreig estratificat en fructicultura els podem trobar en poma (Aggelopoulou et al., 2010) i en préssec (Miranda et al., 2018; Uribeetxebarria et al., 2019a). En ambdós casos, la informació auxiliar utilitzada per a l'estratificació de la mostra prové d'imatges aèries multiespectrals que proporcionen índexs de vegetació (com l'NDVI – Normalized Difference Vegetation Index) que normalment són bons indicadors del vigor dels arbres (i, possiblement, de la mida de la seva capçada). Aplicant aquesta metodologia, Uribeetxebarria et al. (2019a) aconseguiren reduir la mida de la mostra un 17% en comparació al mostreig aleatori. La Figura 2 mostra la parcel·la de pressecger en la qual es va portar a terme aquesta experimentació. Resultats similars han estat reportats per Miranda et al. (2018), havent-se documentat en aquest cas reduccions de mostra de fins el 20-35% quan els arbres (també de pressecger) s'estratificaven en base a l'ús simultani de la secció transversal del tronc i l'índex RVI (Red Vegetation Index). Cal remarcar, però, dues qüestions del mostreig estratificat: primer, no sempre resulta fàcil trobar una informació auxiliar que estigui altament correlada amb la collita i, segon, continua sent necessari seleccionar a l'atzar els arbres a mostrejar dins de cada estrat o classe.

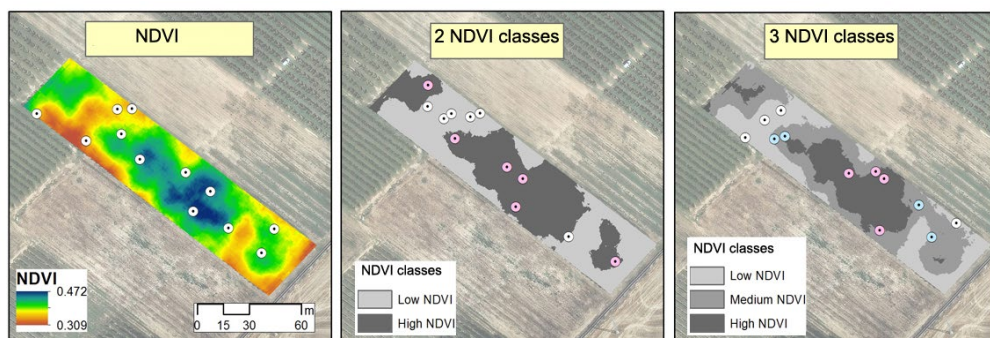


Figura 2. Esquemes de mostreig que utilitzen una mida de mostra de 12 arbres o unitats de mostreig: (esquerra) mostreig aleatori simple, (centre) mostreig estratificat mitjançant dues classes d'NDVI (6 arbres per classe o estrat), (dreta) mostreig estratificat mitjançant tres classes d'NDVI (4 arbres o punts de mostreig per classe) (adaptat d'Uribeetxebarria et al., 2019a)².

² La classificació de l'índex NDVI en dues i tres classes (clusters) es realitza normalment mitjançant l'algorisme de k -mitjanes.



Ranked set sampling

Igual que el mostreig estratificat, aquest mètode fa ús també de la informació proporcionada per sensors propers o remots per tal de facilitar la selecció d'aquells arbres (o unitats de mostreig) que millor representin el conjunt de la parcel·la. Utilitzat per [Uribeetxebarria et al. \(2019b\)](#) en presseguer, el mètode és essencialment una variant del mostreig estratificat. La diferència rau en la resolució de l'estratificació, ja que amb aquest mètode cadascun dels arbres de la mostra es pot considerar com a pertanyent a un estrat diferent (o subpoblació) dins la parcel·la. La forma en que s'obté la mostra s'explica a continuació.

De manera resumida ([Wolfe, 2010](#)), el mostreig consisteix en prendre iterativament n mostres aleatòries de mida n dins la parcel·la, és a dir, contenint cada cop n arbres o unitats de mostreig. Conegut el valor d'una variable auxiliar determinada (per exemple, l'**NDVI** de la capçada o alguna altra característica vegetativa relacionada amb la collita), es procedeix a ordenar els arbres de menor a major en cadascuna de les mostres segons el valor absolut de la citada variable auxiliar o secundària. Per tant, un cop aplicat aquest procés d'ordenació o ranking, es disposa de n mostres seqüencials (de la primera a la última n) dins les quals els arbres han estat ordenats també des de l'arbre amb menys valor de la variable auxiliar (arbre 1) fins l'arbre amb major presència d'aquesta característica (arbre n). Assumint, com ja s'ha dit, una forta correlació entre la característica auxiliar mesurada i la collita, la mostra final seleccionada (mostra ordenada o ranked sample) conté només n arbres, exactament els arbres la posició i dels quals dins la mostra coincideix amb la iteració o mostra i corresponent. En altres paraules, se selecciona l'arbre ordenat 1 de la mostra 1, l'arbre ordenat 2 de la mostra 2, i així successivament fins a seleccionar l'arbre que ocupa la posició n en la última mostra iterada. Efectivament, la condició clau és utilitzar una variable auxiliar adequada per tal d'incloure en la mostra arbres (o unitats de mostreig) representatius de la variabilitat intraparcèl·laria de la collita, fent així possible una estimació no esbiaixada i precisa.

La variància d'una mostra ordenada ve donada per (6) ([Wolfe, 2010](#)),

$$v(\hat{Y}_{RSS}) = v(\hat{Y}_{SRS}) - \frac{1}{n^2} \sum_1^n (\mu_{[j]} - \mu)^2 \quad (6)$$

on $\mu_{[j]}$ representa el valor esperat de la collita per a la posició (subpoblació) j ($j = 1, \dots, n$), i μ la mitjana real desconeguda de la parcel·la. Mitjançant l'estimació de (6) en compteig de fruits en presseguer, [Uribeetxebarria et al. \(2019b\)](#) han demostrat que aquest mètode és doble precís que el mostreig aleatori, aconseguint per tant reduir la mida de la mostra un 50%.



Concretament, fou possible obtenir Coeficients d'Error (CE) del 10% utilitzant mides de mostra de tan sols 5 arbres en una parcel·la de 2,24 ha.

Pel que fa a la informació auxiliar que s'ha de disposar de cada arbre (o unitat de mostreig), els mateixos autors (Uribeetxebarria et al., 2019b) recomanen utilitzar l'àrea projectada zenital de la capçada a partir de l'ús d'una càmera RGB muntada en un vehicle aeri no tripulat (dron). Una altra possibilitat és disposar d'una imatge des d'avioneta, tot i que la delimitació ajustada de la capçada dels arbres es veu ara dificultada donada la menor resolució espacial de la imatge aèria (Figura 3). Finalment, pot ser també interessant disposar del volum de la capçada generat a partir de l'escaneig lateral dels arbres amb sensors LiDAR terrestres (Escolà et al., 2017).

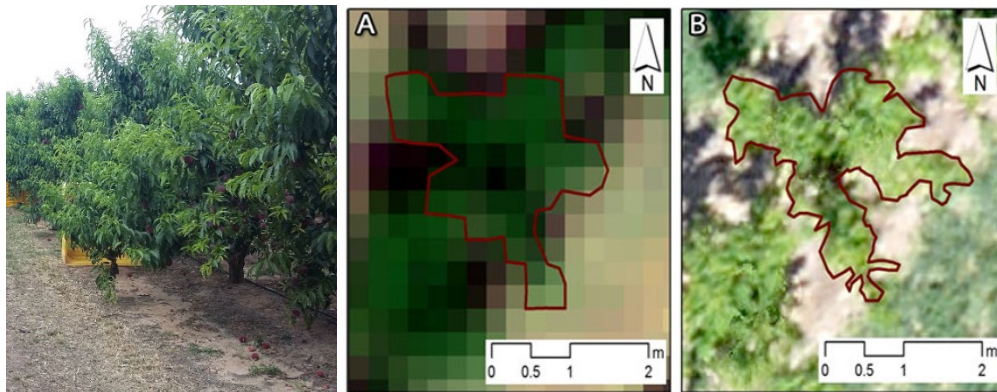


Figura 3. Delimitació en presseguer (esquerra) de l'àrea projectada zenital per a un mateix arbre, (A) utilitzant una imatge aèria des d'avioneta amb resolució 0,25 m/píxel, (B) utilitzant una imatge des de dron amb resolució 0,02 m/píxel (adaptat d'Uribeetxebarria et al., 2019b).

Mostreig sistemàtic multifase

El darrer mètode de mostreig que es presenta es coneix amb la denominació anglesa de multilevel systematic sampling. Conceptualment, consisteix en aplicar un mostreig sistemàtic en diverses fases, on cadascuna de les etapes o fases està anuada a l'anterior. Wulfshon et al. (2012) proposen aquest mètode per al compteig de fruits en pomera a través d'un procediment construït en tres etapes, i amb el qual aconseguen errors de mostreig (Coeficients d'Error, CE) que no superen el 10% a nivell de parcel·la.

Les fases proposades poden ser tres: arbres dins la parcel·la (primer nivell de mostreig), branques principals dins l'arbre (segon nivell) i, dins aquestes branques principals, es marca un tercer nivell que correspon a les branques secundàries o elements de fructificació. La idea és aplicar un mostreig sistemàtic específic per a cadascun d'aquests nivells. Amb aquest procediment, s'arriba a seleccionar dins l'arbre una mostra final de branques



del tercer nivell (branques de fructificació) on es realitza el compteig dels fruits. És important remarcar que, dins de cada nivell o fase, es fixa prèviament un interval de mostreig o període; en el cas que es comenta, s'hauria d'establir un interval entre arbres a mostrejar dins la filera, un altre interval entre branques principals dins cada arbre seleccionat i, finalment, un tercer interval per seleccionar les branques secundàries. La **Figura 4** mostra com s'executa aquest mostreig sistemàtic aniuat en un arbre seleccionat prèviament en la primera fase.

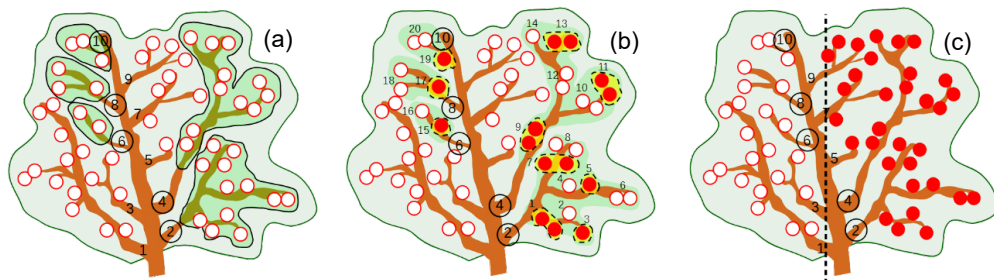


Figura 4. (a) Fase 2. Selecció de les branques principals utilitzant un període de mostreig $m_1 = 2$, que correspon a una fracció de mostreig $1/m_1 = 1/2$. Per tant, se selecciona sistemàticament una de cada dues branques començant aleatòriament per la branca número 2. (b) Fase 3. Selecció de les branques de fructificació dins les branques seleccionades anteriorment. S'utilitza novament un període de mostreig $m_2 = 2$, però ara començant per la branca lateral número 1. La mostra final (cercles omplerts) conté els fruits de les branques 1, 3, 5, 7, 9, 11, 13, 15, 17 i 19. Una estimació no esbiaixada del nombre total de fruits resulta de multiplicar el període combinat de mostreig $m = m_1 \times m_2$ i el número de fruits comptats en la mostra (adaptat de Wulfshon et al., 2012). (c) En mostrejos aleatoris i/o estratificats, es pot optar per comptar només els fruits de la meitat de la capçada, duplicant llavors la mida recomanada de mostra obtinguda de l'equació (1).

A nivell de parcel·la, és important que el fructicultor o tècnic faci una primera avaluació de la variabilitat espacial de la collita. Aquest és un pas necessari a l'hora de fixar els períodes de mostreig més convenients que seran aplicats després en les diferents etapes o fases. En definitiva, es tracta d'escalar apropiadament aquests períodes atenent a ambdues variabilitats, entre arbres dins la parcel·la i també entre branques dins l'arbre. Evidentment, caldrà mostrejar amb major intensitat (disminuint el període de mostreig) en aquella fase on major sigui la variabilitat entre les unitats de mostreig corresponents.



Referències

- 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. *Precision Agriculture*, 11, 538-556.
- Anderson, N.T., Underwood, J.P., Rahman, M.M., Robson, A., & Walsh, K.B. 2019. Estimation of fruit load in mango orchards: tree sampling considerations and use of machine vision and satellite imagery. *Precision Agriculture*, 20, 823-839.
- 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. *Precision Agriculture*, 18, 111-132.
- Miranda C, Santesteban LG, Urrestarazu J, Loidi M, & Royo JB. 2018. Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. *Agriculture*, 8(6), 78.
- Orden PRE/1950/2005, de 17 de Junio, por la que se aprueba la norma específica de peritación de daños en la producción de frutales, amparados por el Seguro Agrario Combinado. Boletín Oficial del Estado, 23 de Junio, 2005, no. 149, pp. 22172-22176.
- Peña D. 2005. Fundamentos de Estadística. Madrid: Alianza Editorial, S.A. 683 pp.
- Uribeetxebarria A, Martínez-Casasnovas JA, Escolà A, Rosell-Polo JR, & Arnó J. 2019a. Stratified sampling in fruit orchards using cluster-based ancillary information maps: a comparative analysis to improve yield and quality estimates. *Precision Agriculture*, 20, 179-192.
- Uribeetxebarria A, Martínez-Casasnovas JA, Tisseyre B, Guillaume S, Escolà A, Rosell-Polo JR, & Arnó J. 2019b. Assessing ranked set sampling and ancillary data to improve fruit load estimates in peach orchards. *Computers and Electronics in Agriculture*, 164, 104931.
- Webster R, & Lark RM. 2013. Field sampling for environmental science and management. London and New York: Routledge.
- Wolfe DA. 2010. Ranked set sampling. *Wires Computational Statistics*, 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. *Precision Agriculture*, 13, 256-275.

L'autoria d'aquest document és del [Grup de Recerca en AgròTICa i Agricultura de Precisió](#) i ha estat elaborat per Jaume Arnó amb la coordinació d'Àlex Escolà i el suport de Carla Román.



Aquesta obra està sota una llicència de Creative Commons Reconeixement-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Activitat finançada a través de l'Operació 01.02.01 de Transferència Tecnològica del Programa de desenvolupament rural de Catalunya 2014-2022