UAB
Universitat Autònoma
de Barcelona

Dipòsit digital
de documents
de la UAB

This is the **published version** of the bachelor thesis:

Rovira Molí, Pol. *Ajust Dinàmic de Sèries Temporals Predictives mitjançant Retroalimentació de Dades Reals*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Grau en Enginyeria de Dades)

This version is available at https://ddd.uab.cat/record/337328

under the terms of the CC BY-SA license.

Ajust Dinàmic de Sèries Temporals Predictives mitjançant Retroalimentació de Dades Reals

Pol Rovira Molí

1674288@uab.cat · Grau en Enginyeria de Dades · Tutor: Antonio Lozano Bagén · Curs 2025/26

Resum— La predicció precisa del creixement porcí es veu limitada per la discrepància entre els models teòrics inicials i l'evolució real dels animals en granja. Aquest treball desenvolupa un Filtre de Kalman Estès (EKF) d'Estat Augmentat per al recalibratge dinàmic de l'asímptota de pes final (A), emprant com a única retroalimentació els tiquets de l'escorxador, que estan esbiaixats per la selecció no aleatòria dels exemplars. L'arquitectura recursiva aïlla aquest biaix sistemàtic, incorporant el coneixement a priori que sempre és positiu, i corregeix el paràmetre crític de la corba logística a cada observació. S'ha avaluat la sensibilitat davant errors d'estimació inicial de ± 25 kg i pautes de freqüència de 3, 6, 9 i 12 tiquets, fins al buidatge complet de la granja. Els resultats mostren que la freqüència no afecta igual tots els casos: quan el ramader sobreestima l'asímptota, fer més enviaments millora la convergència, mentre que en els altres casos l'empitjora una mica. A més, el biaix estimat serveix com a diagnòstic de la convergència del filtre.

Paraules clau— Filtre de Kalman Estès (EKF), recalibratge dinàmic, corba logística, estimació de paràmetres, biaix de selecció, creixement porcí, simulació estocàstica.

Abstract— Accurate prediction of swine growth is limited by the discrepancy between initial theoretical models and the actual evolution of the animals. This work develops an Augmented State Extended Kalman Filter (EKF) for the dynamic recalibration of the final weight asymptote (A), using slaughterhouse weight tickets as the sole feedback, which are biased by the non-random selection of specimens. The recursive architecture isolates this systematic bias, incorporating the prior knowledge that it is always positive, and corrects the critical parameter of the logistic curve at each observation. Sensitivity has been evaluated against initial estimation errors of ± 25 kg and frequency schedules of 3, 6, 9 and 12 tickets, up to the complete emptying of the farm. Results show that frequency does not affect all cases equally: when the asymptote is overestimated, more shipments improve convergence, whereas in the other cases it slightly degrades it. The estimated bias serves as a diagnostic of the filter's convergence.

Index Terms— Extended Kalman Filter (EKF), dynamic recalibration, logistic curve, parameter estimation, selection bias, swine growth, stochastic simulation.

Agraïments— Voldria agrair aquest treball, en primer lloc, al meu tutor, l'Antonio Lozano, qui ha sigut de gran ajuda i m'ha donat tota la confiança i el recolzament durant el semestre, fent que em sentís acompanyat i còmode en tot moment. Seguidament, als meus pares, en Miquel i la Montse, i al meu germà, en Gerard, els quals m'han donat suport durant aquests quatre anys de carrera: com més difícil es posava el camí, més hi eren per impulsar-me i no deixar que em rendís en cap moment. També a tots els amics que he fet durant l'època universitària, que han fet que siguin quatre anys inoblidables, plens de moments i memòries. A la meva parella, la Mireia, que em dona el suport que necessito amb gestos petits però d'un significat molt gran. I, per acabar, voldria dedicar-me el treball a mi mateix, ja que sempre m'he sabut refer dels problemes i he tirat sempre endavant.

1 Introducció

En la indústria ramadera actual, el sacrifici dels animals d'engreix és el moment que tanca tot el cicle productiu i en bona part determina la rendibilitat de l'explotació. Decidir *quan* enviar cada animal a l'escorxador és, per tant, una decisió econòmica de primer ordre, comuna a la majoria d'espècies de producció càrnia. Aquest projecte se centra en el creixement i el sacrifici dels porcs, on aquesta decisió és especialment sensible i on es disposa d'un model de creixement ben caracteritzat.

En l'engreix porcí, el procés culmina amb l'enviament dels animals a l'escorxador, un moment on el benefici depèn d'assolir un pes objectiu específic en el moment adequat. Un *lot* és el conjunt d'animals d'una mateixa nau que s'engreixen alhora i que el ramader gestiona com una unitat. Enviar massa d'hora significa perdre dies de creixement potencial i no aprofitar el pes òptim que l'animal encara

podria assolir; enviar massa tard implica un manteniment innecessari de l'animal — amb el consum de pinso i els costos associats que això comporta — un cop ja ha superat el seu punt de creixement més eficient.

La principal dificultat és que el ramader no pot pesar individualment els milers d'animals del lot de forma freqüent. En conseqüència, les decisions d'enviament es basen en models teòrics de creixement, habitualment corbes logístiques, que el ramader parametritza a partir de l'experiència prèvia amb la raça i les condicions de la granja. El paràmetre més sensible d'aquest model és l'asímptota A, que representa el pes màxim que els animals podrien assolir. Una estimació incorrecta d'aquest valor desvia tota la corba predictiva i fa que les decisions d'enviament es basin en una projecció equivocada de la realitat del lot.

L'únic feedback de pes real i massiu disponible és el que prové dels *tiquets de bàscula de l'escorxador*: cada cop que el ramader envia un camió d'animals a sacrificar, l'escorxador els pesa i li retorna un comprovant (el tiquet) amb el pes mitjà de la càrrega. Aquestes dades, però, presenten dos problemes fonamentals. En primer lloc, contenen soroll de mesura inherent a la instrumentació. En segon lloc, i de forma molt més rellevant, estan sistemàticament esbiaixades per la tria visual del ramader: en seleccionar preferentment els animals aparentment més pesats per als primers enviaments, les observacions sobrerepresenten els extrems superiors de la distribució del lot i no reflecteixen la seva mitjana real.

Aquest projecte proposa una solució a aquest problema doble mitjançant un **Filtre de Kalman Estès (EKF) d'Estat Augmentat** [1, 6]. L'EKF no estima el pes puntual del lot, sinó que recalibra directament el paràmetre d'asímptota A de la corba logística a cada tiquet rebut. Simultàniament, estima el biaix de selecció del ramader com a variable d'estat interna, partint del coneixement a priori que aquest biaix és sempre positiu. D'aquesta manera, el sistema és capaç de separar l'error sistemàtic introduït per la tria humana del senyal real de creixement biològic, i així obtenir una corba corregida que s'acosta al pes real del lot.

2 Estat de l'Art

El projecte s'emmarca en la intersecció de tres àrees teòriques: la modelització biològica del creixement animal, la teoria de l'estimació estadística de sistemes dinàmics, i el tractament del biaix de selecció en dades de mostreig no aleatori.

2.1 Models de creixement sigmoide

Segons la recerca, el creixement animal segueix patrons en forma de S [3, 2]. Una corba en forma de S (o sigmoide) creix lentament al principi, s'accelera en una fase central i torna a frenar-se en apropar-se a un valor màxim. Existeixen diversos models matemàtics que descriuen aquesta forma; els més habituals en producció animal són la corba logística i la de *Gompertz* [11]. Totes dues tenen tres paràmetres i una forma de S molt semblant, però es diferencien en on situen el punt d'inflexió (el moment de màxim creixement): la logística és simètrica respecte a aquest punt, mentre que la de Gompertz és asimètrica i el situa més aviat. En aquest context s'utilitza la corba logística per la seva simplicitat i perquè descriu bé el creixement porcí [12]. La corba logística és el model estàndard en producció porcina per la seva capacitat de capturar les tres fases del cicle biològic: el creixement accelerat de la fase juvenil, el punt d'inflexió de màxima taxa de creixement, i la desacceleració progressiva a mesura que l'animal s'acosta al seu límit biològic. L'asímptota A d'aquest model és el paràmetre de major impacte pràctic, ja que determina el "sostre" de la corba i, per tant, el calendari d'enviaments. Una variació de tan sols 10 kg en A pot desplaçar la data de sacrifici òptima entre 5 i 15 dies. De fet, diversos treballs apliquen models logístic i de Gompertz a dades reals de porcs per estimar el pes de sacrifici òptim [8].

2.2 Biaix de selecció

El biaix de selecció és un fenomen estadístic que apareix quan la mostra observada no és representativa de la població total [4]. En la producció porcina, aquest fenomen es manifesta de forma crítica: la tria visual que realitza el ramader per enviar els exemplars aparentment més grossos en els primers camions introdueix un biaix sistemàtic positiu [14]. Com a conseqüència, les dades de l'escorxador reflecteixen consistentment pesos molt superiors a la mitjana real del lot restant a la granja. Si un model predictiu s'ajusta directament sobre aquestes observacions sense una correcció prèvia, calcularà una asímptota artificialment elevada, induint a decisions logístiques prematures i subòptimes. Aquest biaix no és constant durant tot el cicle: a mesura que els animals més grossos van sortint de la granja, el lot que queda és més homogeni i el biaix va baixant. Com que aquest canvi és lent, el filtre el pot anar seguint com una variable d'estat més.

2.3 Filtre de Kalman

El Filtre de Kalman [1] és un mètode que serveix per estimar el valor d'una magnitud que no podem mesurar directament o que mesurem amb error, combinant dues fonts d'informació: d'una banda, el que prediu un model teòric del sistema; de l'altra, les mesures reals que anem rebent, que sempre porten una mica de soroll. Una analogia útil és la d'un conductor que vol saber a quina velocitat va: pot fiar-se del que marca el velocímetre (la mesura, que pot tenir imprecisions) o del que esperaria segons l'accelerador que prem (el model). El més intel·ligent no és creure cegament cap dels dos, sinó combinar-los donant més pes al que en cada moment sigui més fiable. Això és exactament el que fa el Filtre de Kalman: en lloc de fiar-se del tot del model o del tot de les mesures, calcula una mitjana ponderada de tots dos segons com de fiable és cadascun.

El filtre funciona de manera *recursiva*: va actualitzant la seva estimació cada vegada que arriba una mesura nova, sense haver de recalcular-ho tot des del principi, sinó simplement corregint l'estimació anterior. Per això és especialment útil en sistemes que reben dades a poc a poc al llarg del temps, com és el nostre cas amb els tiquets de l'escorxador, que arriben enviament rere enviament.

En cada pas, el filtre repeteix dues etapes. Primer *prediu* quin valor hauria de tenir la magnitud segons el model i quanta incertesa hi associa (com més temps passa sense una mesura nova, menys segur n'està). Després, quan arriba una mesura real, *corregeix* aquella predicció: calcula la diferència entre el que esperava i el que realment ha mesurat — aquesta diferència s'anomena *innovació*, perquè és la

informació nova que aporta la mesura — i ajusta l'estimació en funció d'un factor de ponderació anomenat *Guany de Kalman*. Aquest guany decideix quant cas fer a la mesura nova: si la mesura és molt fiable, el filtre la segueix de prop; si és sorollosa, la té poc en compte i confia més en el model. D'aquesta manera, amb cada mesura el filtre va afinant la seva estimació i en redueix la incertesa.

El Filtre de Kalman clàssic, però, només funciona bé quan la relació entre el que volem estimar i el que mesurem és lineal, és a dir, quan es pot representar amb una recta. En el nostre cas no ho és: el pes depèn de l'asímptota A a través d'una corba logística, que té forma de S i, per tant, és no lineal. Per resoldre-ho cal fer servir el Filtre de Kalman Estès (EKF), una variant que aproxima la corba per una recta just al voltant del punt on es troba l'estimació actual [6, 5]. Aquesta aproximació lineal local es calcula mitjançant la matriu Jacobiana, que no és res més que el pendent de la corba en aquell punt. A més, en aquest projecte no estimem només l'asímptota, sinó que afegim el biaix de selecció b al conjunt de variables que el filtre estima alhora. Aquesta tècnica, anomenada estimació conjunta d'estat i paràmetres [7], és la que permet al sistema separar l'efecte real del creixement de l'efecte de la tria del ramader i corregir la corba a cada nou tiquet.

L'ús del Filtre de Kalman en producció animal no és nou: ja s'ha aplicat de forma recursiva al seguiment del consum de pinso en aviram [9] i a la monitorització de la producció diària de llet en vaques [10], on també s'aprofita la innovació per detectar canvis en la tendència. La novetat d'aquest projecte és aplicar-lo al recalibratge de l'asímptota de creixement porcí incorporant el biaix de selecció com a variable d'estat.

3 Objectius

L'objectiu general del treball és dissenyar i validar un sistema de recalibratge dinàmic de la corba de creixement porcí que sigui capaç de corregir una hipòtesi inicial incorrecta sobre l'asímptota A a partir de les dades operatives de l'escorxador, que estan contaminades per biaix de selecció. Per assolir-lo, s'han definit els objectius específics següents:

- **Model biològic i simulador:** implementar un simulador estocàstic de 3.000 animals amb variabilitat biològica individual (heteroscedasticitat del 2%) i un mecanisme de biaix de selecció que repliqui el comportament real del ramader, amb una preselecció dels 500 animals més pesats i enviament aleatori de 250, actualitzant el cens a cada enviament.

- **Filtre de Kalman en estat augmentat:** dissenyar i implementar un Filtre de Kalman Estès on l'estat no sigui el pes puntual del lot sinó el vector format per l'asímptota A i el biaix de selecció b. El filtre ha de poder estimar ambdós paràmetres simultàniament, incorporant el coneixement a priori que el biaix és sempre positiu.

- **Bateria d'experiments (Bloc 1):** avaluar la capacitat de convergència del filtre davant tres hipòtesis inicials del ramader sobre l'asímptota: sobreestimació, estimació correcta i subestimació, mantenint fixa la freqüència d'enviaments.

- **Estudi de robustesa temporal i buidatge (Bloc 2):** avaluar com la freqüència d'enviaments (3, 6, 9 i 12 tiquets) afecta la qualitat de la convergència del filtre i, en el cas extrem de 12 enviaments, observar l'evolució de A i b a mesura que la granja es va buidant.

- **Validació quantitativa:** comparar l'error final en l'asímptota estimada entre escenaris, i demostrar la superioritat del recalibratge dinàmic respecte a mantenir el model teòric original sense corregir.

4 Metodologia i Disseny Experimental

La metodologia combina la simulació estocàstica controlada, per disposar d'una veritat absoluta coneguda, amb el filtratge recursiu no lineal, per recalibrar el model teòric en temps real. El disseny experimental permet avaluar de forma aïllada l'efecte de la hipòtesi inicial del ramader i l'efecte de la freqüència d'enviaments.

4.1 Model biològic i simulador

El creixement esperat d'un animal en el dia t es modela mitjançant la corba logística, on l'asímptota A és el paràmetre central d'aquest projecte:

$$P(t, A) = \frac{A}{1 + e^{-k(t-t_0)}} \tag{1}$$

On $k = 0{,}035$ dia^{-1} és la taxa de creixement (les unitats dia^{-1}, és a dir "per dia", indiquen que k mesura el ritme de creixement relatiu cada dia) i $t_0 = 154$ dies és el punt d'inflexió de la corba, el dia en què el creixement és més ràpid. Tots dos valors s'han fixat a partir de dades de referència de la raça [13]: en condicions d'engreix intensiu, els porcs solen assolir la màxima velocitat de creixement al voltant dels cinc mesos de vida, fet que situa t_0 prop del dia 154. Aquests dos paràmetres es consideren coneguts i fixos; l'única incògnita que el filtre ha d'estimar és l'asímptota A. L'asímptota A, en canvi, és la incògnita del sistema: representa el pes màxim biològicament assolible pels animals del lot, i és el paràmetre que el ramader ha d'estimar i que el filtre ha de recalibrar.

Per a les simulacions d'aquest projecte, s'ha fixat l'asímptota real a $A_r = 225$ kg. Aquest valor s'ha escollit perquè s'aproxima al pes màxim assolible en condicions normals d'engreix porcí intensiu [12], i permet generar escenaris realistes de sobreestimació i subestimació que el ramader podria trobar en la pràctica real.

El simulador genera una població de 3.000 animals, cada un amb el seu propi valor d'A extret d'una distribució normal centrada en A_r. D'aquesta manera es reprodueix la variabilitat genètica i nutricional natural del lot, evitant que tots els animals siguin còpies idèntiques del model teòric. A sobre d'aquesta variabilitat estructural, s'incorpora un soroll gaussià del 2% sobre el pes diari de cada animal, que modela la heteroscedasticitat biològica: la dispersió absoluta de pesos entre animals és més gran al final del cicle (quan el lot pesa al voltant de 200 kg) que a l'inici (quan pesa al voltant de 20 kg), cosa que fa que l'estimació de la mitjana real sigui progressivament més difícil a mesura que avança l'engreix.

4.1.1 Generació de tiquets esbiaixats

Per reproduir el comportament real del ramader, el mòdul de generació de tiquets simula el procés de selecció visual que es produeix en el moment de carregar el camió. El ramader no escull els animals de forma aleatòria: tendeix a seleccionar preferentment els exemplars que li semblen més grossos, introduint un biaix sistemàtic a les dades.

Aquesta lògica s'implementa en dues etapes. En primer lloc, del cens total d'animals disponibles en el dia d'observació, s'identifiquen els 500 de major pes, que representen el subgrup que el ramader percep visualment com a candidats òptims per a l'enviament. En segon lloc, d'aquest subgrup se n'escullen 250 de forma aleatòria per formar la càrrega efectiva del camió. El pes mitjà dels 250 animals carregats s'emmagatzema com a observació del dia corresponent.

Un aspecte clau del simulador és que el cens s'actualitza a cada enviament: els 250 animals carregats abandonen definitivament la granja, de manera que el següent tiquet es genera sobre el cens restant (per exemple, 2.750 animals després del primer enviament de 3.000). Aquesta dinàmica reprodueix el buidatge progressiu real de l'explotació i té una conseqüència important sobre el biaix: com que els animals més pesats van sortint primer, el lot restant és cada cop més homogeni i el biaix de selecció disminueix progressivament fins a apropar-se a zero quan la granja queda gairebé buida.

4.2 Filtre de Kalman Estès (EKF) d'estat augmentat

Per entendre com funciona el filtre, cal partir d'una pregunta senzilla: *quan arriba un tiquet de l'escorxador, com sap el sistema si el pes que hi apareix és alt perquè els animals del lot creixen molt, o perquè el ramader ha seleccionat els més grossos?* Aquesta ambigüitat és el problema central, i el Filtre de Kalman Estès és l'eina que permet resoldre-la.

4.2.1 Dos paràmetres, una sola mesura

El filtre no intenta estimar el pes del lot dia a dia, sinó que treballa directament sobre els dos paràmetres que expliquen el que es veu als tiquets:

- **L'asímptota** A: el pes màxim que poden assolir els animals del lot. És el "sostre" de la corba logística, i és el paràmetre que el ramader pot haver estimat malament. El filtre l'ajusta a cada tiquet rebut.
- **El biaix** b: la quantitat de quilograms que els tiquets sobreestimen sistemàticament la mitjana real del lot, degut a la selecció dels animals més pesats. El filtre no en coneix el valor exacte a priori, però sí que parteix del coneixement que és positiu: s'inicialitza a $b_0 = 8$ kg i es restringeix perquè mai pugui prendre valors negatius, ja que un biaix de selecció negatiu seria completament ineviable en aquest problema.

Quan arriba un tiquet, el pes que hi apareix és la suma d'aquests dos efectes: el pes que prediu la corba logística (determinat per A) més el biaix de selecció (b). El repte és que el filtre, a partir d'un sol número (el pes del tiquet), ha de deduir quant correspon a cadascun.

4.2.2 Per què cal la Jacobiana?

En un Filtre de Kalman estàndard, la relació entre els paràmetres interns i la mesura observada ha de ser lineal. Però en el nostre cas, el pes és una funció en forma de S de l'asímptota A (Eq. 1), que és clarament no lineal. Per poder aplicar el filtre, cal *linealitzar localment* aquesta relació en el punt on es troba l'estimació actual d'A. Aquesta aproximació lineal és la Jacobiana:

$$H = \left[\frac{1}{1 + e^{-k(t-t_0)}},\ 1 \right] \tag{2}$$

El primer terme d'H respon a la pregunta: *si l'asímptota augmentés 1 kg, quant augmentaria el pes observat avui?* La resposta depèn del dia del cicle. Al dia 154 (el punt d'inflexió de la corba), la resposta és $0{,}5$ kg: la meitat de l'increment en A es trasllada al pes observable. Al final del cicle (dia 200 en endavant), la resposta s'aproxima a 1 kg: gairebé tot l'increment en A es reflecteix directament al pes. Això té una implicació pràctica molt important: **els tiquets tardans contenen molta més informació útil sobre l'asímptota que els primerencs**, raó per la qual un enviament al dia 200 aporta una correcció d'A gairebé tan eficient com dos enviaments al dia 170.

4.2.3 Funcionament del filtre pas a pas

El filtre opera en un cicle continu de dues etapes que es repeteix cada dia del cicle d'engreix.

En l'**etapa de predicció**, el filtre avança un dia sense rebre cap dada nova. Tant A com b es mantenen iguals (ambdós canvien lentament), però la confiança en l'estimació disminueix lleugerament: com més temps passa sense un nou tiquet, menys segur és el filtre dels seus valors actuals. Aquesta incertesa creixent es representa internament com una amplada creixent de la banda $\pm 1\sigma$ visible als gràfics.

En l'**etapa d'actualització**, quan arriba un tiquet d'escorxador, el filtre compara el pes observat amb el pes que hauria d'haver observat donats els valors actuals d'A i b. La diferència entre tots dos (la innovació) és el senyal que el filtre utilitza per corregir. A través del Guany de Kalman i la Jacobiana, el filtre distribueix automàticament aquesta correcció entre A i b: si la innovació és consistent amb un error en l'asímptota, corregeix principalment A; si és consistent amb un biaix de selecció, corregeix principalment b. Després de cada actualització, si l'estimació de b caigués per sota de zero es retalla a zero, imposant així el coneixement a priori que el biaix de selecció no pot ser negatiu. Amb cada tiquet addicional, la separació entre tots dos efectes millora i la incertesa es redueix.

Els hiperparàmetres del filtre s'han triat seguint aquestes idees. La matriu de covariància inicial $P_0 = \mathrm{diag}(25, 50)$ reflecteix una confiança moderada en la hipòtesi inicial de l'asímptota i una incertesa més gran sobre el biaix, que es desconeix amb més detall. La matriu de soroll de procés $Q = \mathrm{diag}(0{,}05,\ 0{,}02)$ és deliberadament petita, ja que tant A com b són magnituds de variació molt lenta: el seu valor real amb prou feines canvia entre un dia i el següent. Finalment, el soroll de mesura $R_t = 15 + 0{,}10\,(t - 60)$ creix linealment amb el dia del cicle per modelar la heteroscedasticitat: com que la dispersió de pesos entre animals augmenta cap al final de l'engreix, els tiquets tardans es processen amb una confiança lleugerament menor per evitar correccions brusques. Això crea un compromís, perquè els tiquets tardans són alhora els que donen més informació sobre l'asímptota (la Jacobiana és propera a 1).

L'equilibri entre aquests dos efectes condiciona com de ràpid convergeix el filtre.

4.2.4 Definició de les mètriques (MAE i RMSE)

Un cop el filtre ha recalibrat l'asímptota, cal una manera de mesurar com de bé ho ha fet, per poder comparar els diferents experiments amb un número objectiu. Per a això s'utilitzen dues mètriques d'error: l'error absolut mitjà (MAE) i l'arrel de l'error quadràtic mitjà (RMSE). Totes dues indiquen, en quilograms, com de lluny queda l'estimació del filtre respecte a la realitat: com més petites són, millor ha recalibrat el filtre.

Abans de definir-les cal precisar **què es compara**, ja que amb les dades disponibles es podrien fer diverses comparatives. En aquest context, les mètriques comparen el **pes estimat pel filtre** amb el **pes real** (Veritat Absoluta), avaluats únicament als dies en què hi ha un tiquet. Per a cada dia d'enviament t_i:

- El *pes estimat* és la corba logística avaluada amb l'asímptota que el filtre ha recalibrat fins aquell dia: $\hat{P}_i = P(t_i, \hat{A}_{t_i})$.
- El *pes real* és la corba logística amb l'asímptota veritable: $P_i^r = P(t_i, A_r)$.

A partir de l'error puntual $e_i = |\hat{P}_i - P_i^r|$, les mètriques es defineixen com:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} e_i, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2} \quad (3)$$

on N és el nombre d'enviaments. Tot i mesurar el mateix error, cada mètrica aporta una informació diferent. El **MAE** és la mitjana directa dels errors: ens diu, de mitjana, quants quilograms s'equivoca el filtre, tractant tots els errors per igual. El **RMSE** eleva cada error al quadrat abans de fer la mitjana, cosa que fa que els errors grans pesin molt més que els petits; per això el RMSE és especialment sensible a una sola desviació molt gran.

La utilitat de tenir-les totes dues està en comparar-les. Quan el MAE i el RMSE són semblants, vol dir que el filtre s'equivoca de manera estable durant tot el cicle, sense pics. Quan el RMSE és força més gran que el MAE, indica que hi ha algun enviament puntual amb un error molt més alt que la resta. Així, el MAE dona la magnitud típica de l'error i el RMSE avisa si hi ha errors aïllats greus; aquest és el criteri que es farà servir per llegir els resultats.

Cal remarcar, finalment, que les mètriques *no* es calculen contra el pes dels tiquets, ja que aquests incorporen el biaix de selecció i no representen la mitjana real del lot; compararhi mesuraria el biaix, no la qualitat de la recalibració de l'asímptota.

4.3 Disseny de la bateria d'experiments

S'han dissenyat dos blocs d'experiments. El **Bloc 1** estudia l'impacte de la hipòtesi inicial del ramader sobre l'asímptota, fixant 3 enviaments als dies 170, 185 i 200 i variant la hipòtesi inicial: sobreestimació ($A_p > A_r$), estimació correcta ($A_p = A_r$) i subestimació ($A_p < A_r$). En tots tres casos, l'asímptota real del lot és la mateixa ($A_r = 225$ kg).

El **Bloc 2** estudia l'impacte de la freqüència d'enviaments, comparant quatre configuracions (3, 6, 9 i 12 enviaments) per a cadascuna de les tres hipòtesis inicials. La configuració de 12 enviaments correspon al buidatge pràcticament complet de la granja i permet observar l'evolució de A i b al llarg de tot el cicle.

5 Resultats i Discussió

S'han realitzat dotze experiments combinant tres hipòtesis inicials del ramader sobre l'asímptota ($A_p \in \{200, 225, 250\}$ kg) amb quatre freqüències de tiquets (3, 6, 9 i 12 enviaments), amb l'asímptota real fixada a $A_r = 225$ kg en tots els casos. Totes les mètriques MAE i RMSE segueixen la definició de la Secció 4.2.4. La Taula 1 recull els resultats complets.

TAULA 1: Resultats dels dotze experiments: asímptota final estimada ($\hat{A}$), biaix final estimat ($\hat{b}$), MAE i RMSE del pes estimat respecte al pes real ($A_r = 225$ kg en tots els casos).

Hip.	Env.	$\hat{A}$	$\hat{b}$	MAE	RMSE
Sobre. (250)	3	243,5	0,00	14,77	14,79
	6	239,7	0,00	13,93	13,94
	9	236,0	0,00	12,70	12,78
	12	230,8	0,00	10,90	11,22
Control (225)	3	225,0	8,15	0,04	0,05
	6	223,8	7,32	0,36	0,53
	9	221,7	6,43	0,94	1,33
	12	218,2	5,70	1,84	2,64
Subest. (200)	3	205,8	20,29	14,93	14,96
	6	205,6	20,27	15,41	15,45
	9	203,9	19,55	15,85	15,92
	12	200,8	18,45	16,40	16,59

5.1 Bloc 1: impacte de la hipòtesi inicial (A_p)

Aquest bloc compara les tres hipòtesis inicials amb la freqüència de referència de 3 enviaments (dies 170, 185 i 200). El filtre parteix d'un biaix inicial positiu ($b_0 = 8$ kg) restringit a valors no negatius.

5.1.1 Escenari A: sobreestimació ($A_p = 250$ kg)

Quan el ramader sobreestima l'asímptota en 25 kg, el model teòric inicial projecta una corba per sobre de la realitat biològica del lot. L'EKF detecta que els tiquets són inferiors a la predicció i interpreta que l'asímptota és excessiva. Com que el biaix està restringit a valors no negatius, el filtre no pot "amagar" l'excés de predicció convertint b en negatiu: en aquest escenari $\hat{b}$ queda fixat a $0{,}00$ kg i tota la discordança recau sobre l'asímptota. Amb 3 enviaments, però, la correcció és parcial i l'asímptota estimada es queda en $243{,}5$ kg, lluny de $A_r = 225$ kg, amb un MAE de $14{,}77$ kg i un RMSE de $14{,}79$ kg (Fig. 1). Que $\hat{b}$ es quedi a zero ja és una pista: vol dir que el filtre no pot corregir del tot la sobreestimació amb tan pocs tiquets.

5.1.2 Escenari B: estimació correcta ($A_p = A_r = 225$ kg)

Quan el ramader encerta la hipòtesi inicial, les tres corbes (veritat absoluta, predicció base i Kalman) se solapen pràcticament (Apèndix A, Fig. 7). El filtre no ha de

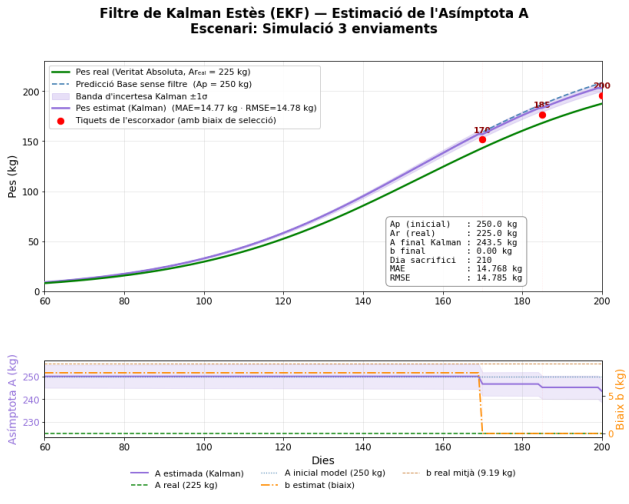

Fig. 1: Bloc 1, sobreestimació ($A_p = 250$ kg) amb 3 enviaments. El biaix estimat queda fixat a zero per la restricció $b \geq 0$ i l'asímptota només es corregeix parcialment ($\hat{A} = 243{,}5$ kg).

corregir A: l'asímptota estimada es manté a 225,0 kg. L'única component que treballa és l'estimació del biaix, que s'estabilitza en un valor positiu de $\hat{b} = 8{,}15$ kg, molt proper al biaix real mitjà observat (9,19 kg). El MAE final és de tan sols 0,04 kg i el RMSE de 0,05 kg, la precisió més alta de tots els experiments, confirmant que quan la hipòtesi del ramader és correcta el sistema separa netament el biaix de l'asímptota.

5.1.3 Escenari C: subestimació ($A_p = 200$ kg)

Quan el ramader subestima l'asímptota en 25 kg, el model prediu que els animals s'aturaran de créixer massa aviat. L'EKF observa que els tiquets superen la predicció, però la discordança total inclou tant l'error en A com el biaix real. El filtre incrementa A de forma parcial (fins a 205,8 kg) i alhora **infla el biaix molt per sobre del seu valor real**: $\hat{b} = 20{,}29$ kg davant d'un biaix real mitjà de 9,19 kg (Apèndix A, Fig. 10). Aquesta sobreestimació del biaix és el mecanisme amb què el filtre compensa el dèficit d'asímptota que no pot recuperar. El MAE resultant és de 14,93 kg i el RMSE de 14,96 kg.

Els tres casos mostren un comportament asimètric: en la sobreestimació el biaix es queda a zero, i en la subestimació es dispara molt per sobre del valor real. En tots dos casos el biaix absorbeix l'error d'A que el filtre no pot corregir, i per tant el seu valor final ens diu com de bé ha convergit.

5.2 Bloc 2: efecte de la freqüència d'enviaments

Aquest bloc analitza com afecta la freqüència d'enviaments (3, 6, 9 i 12 tiquets) la qualitat de la recalibració, comparant les tres hipòtesis inicials de manera conjunta. La idea és veure si rebre més tiquets ajuda el filtre a convergir millor, i si ho fa igual en tots els escenaris. La Taula 1 recull tots els valors; aquí se n'analitza la tendència.

El primer fet, comú a totes les hipòtesis, és que en augmentar la freqüència el **biaix real mitjà va disminuint** (9,19 → 8,52 → 7,28 → 5,37 kg per a 3, 6, 9 i 12 enviaments). Això passa perquè, a cada enviament, surten de la granja els animals més pesats i el lot restant queda més homogeni, de manera que els tiquets posteriors estan cada cop menys esbiaixats. Aquesta és la clau per entendre per què la freqüència no afecta igual els tres escenaris.

5.2.1 La freqüència no té el mateix efecte segons la hipòtesi

En augmentar el nombre d'enviaments de 3 a 12, els tres escenaris reaccionen de manera diferent:

- En la **sobreestimació** ($A_p = 250$), el MAE *baixa* de forma sostinguda: 14,77 → 13,93 → 12,70 → 10,90 kg. Més freqüència, més precisió.
- En el **control** ($A_p = 225$), el MAE *puja* lleugerament: 0,04 → 0,36 → 0,94 → 1,84 kg. Més freqüència, una mica menys de precisió (tot i que l'error es manté sempre per sota de 2 kg).
- En la **subestimació** ($A_p = 200$), el MAE també *puja* suaument: 14,93 → 15,41 → 15,85 → 16,40 kg.

És a dir, augmentar la freqüència només ajuda quan la hipòtesi inicial és massa alta; en els altres dos casos, o bé empitjora una mica o bé és gairebé indiferent.

5.2.2 Per què la freqüència ajuda en sobreestimació

La causa és el buidatge. Com que el biaix real baixa amb cada enviament, els tiquets tardans són observacions gairebé netes del pes real del lot. A més, com s'ha vist a la Secció 4.2.4 amb la Jacobiana, al final del cicle el pes observat reflecteix gairebé directament l'asímptota. La combinació de tots dos factors fa que, en la sobreestimació, aquests tiquets tardans aportin informació molt fiable sobre A i permetin al filtre baixar l'asímptota cap al valor real (243,5 → 230,8 kg). El cas extrem es veu a la Fig. 2, i les freqüències intermèdies a l'Apèndix A (Figs. 5 i 6).

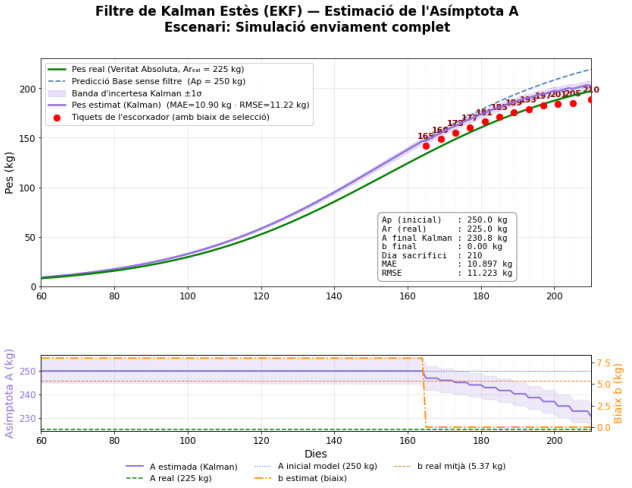

Fig. 2: Sobreestimació ($A_p = 250$ kg) amb 12 enviaments (buidatge complet). En els enviaments tardans el biaix real és molt baix; l'asímptota baixa fins a 230,8 kg i el MAE és el mínim de l'escenari (10,90 kg).

5.2.3 Per què la freqüència empitjora en els altres casos

En el control i la subestimació passa el contrari, i pel mateix motiu del buidatge. Quan l'asímptota ja és correcta (control), els tiquets tardans, que tenen un biaix real baix, surten per sota del que el filtre esperaria amb el seu biaix estimat de ~ 8 kg; el filtre interpreta en part aquesta caiguda com si l'asímptota fos una mica més baixa, i l'estira cap avall (de

225,0 a 218,2 kg). Tot i així, l'error continua per sota de 2 kg, de manera que la degradació és pràcticament irrellevant (Fig. 3; la resta de freqüències, a l'Apèndix A).

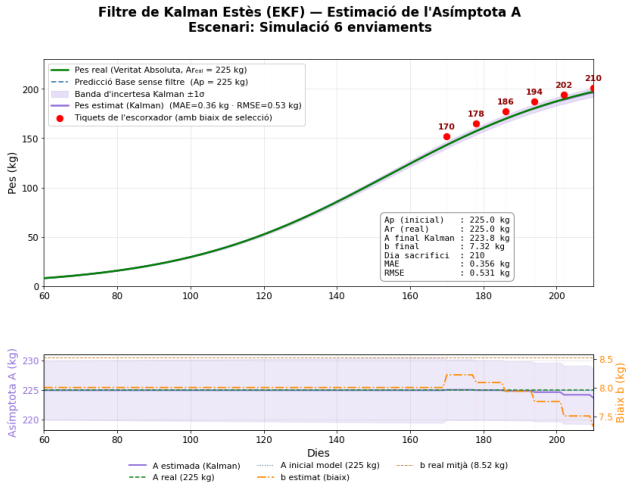

Fig. 3: Control ($A_p = 225$ kg) amb 6 enviaments. Partint d'una asímptota correcta, el filtre estima netament el biaix de selecció i manté un error molt baix ($\hat{A} = 223{,}8$ kg, MAE $= 0{,}36$ kg).

En la subestimació, els tiquets tardans de biaix baix tampoc no ajuden a recuperar l'asímptota, perquè el filtre ja ha "compromès" la major part de l'error en un biaix estimat molt inflat ($\sim$ 18–20 kg) que actua com a ancoratge. Per això el MAE creix lleugerament amb la freqüència en lloc de baixar (Fig. 4; la resta de freqüències, a l'Apèndix A).

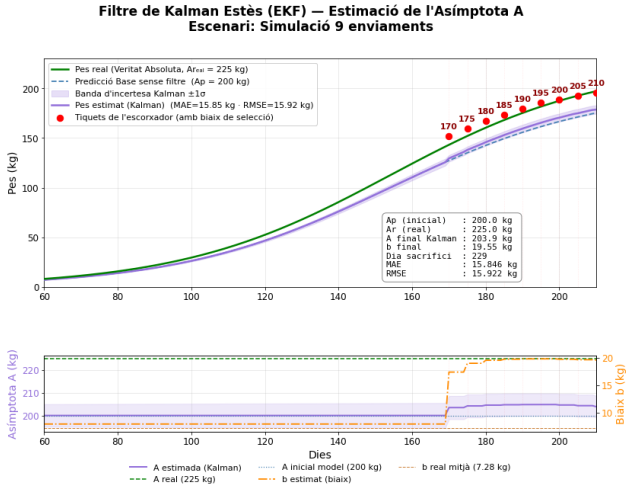

Fig. 4: Subestimació ($A_p = 200$ kg) amb 9 enviaments. El biaix estimat queda inflat i ancorat en un valor alt, de manera que els tiquets tardans no recuperen l'asímptota ($\hat{A} = 203{,}9$ kg, MAE $=$ 15,85 kg).

5.2.4 Evolució del biaix durant el buidatge

El panell inferior de les figures de 12 enviaments mostra clarament aquest buidatge: a mesura que surten els animals més pesats, el biaix de selecció real decreix de forma sostinguda fins a apropar-se a valors molt baixos en els últims camions, quan el tiquet i la mitjana real del lot gairebé coincideixen. Aquesta dinàmica, observable en els tres escenaris, valida qualitativament el model de selecció proposat i és la que explica per què la freqüència té efectes oposats segons la hipòtesi inicial.

5.3 Síntesi dels resultats

Els dotze experiments revelen tres patrons principals. En primer lloc, la hipòtesi inicial segueix sent el factor dominant: l'escenari de control assoleix sempre errors molt baixos ($<$ 2 kg), mentre que un error inicial de ± 25 kg manté el MAE en l'ordre de 11–16 kg. En segon lloc, i a diferència del que s'havia observat en estudis preliminars amb pocs tiquets, **l'efecte de la freqüència no és uniforme**: en la sobreestimació, augmentar els enviaments millora la convergència de manera notable (de 14,77 a 10,90 kg de MAE) gràcies als tiquets tardans de biaix baix; en el control i la subestimació, en canvi, la freqüència degrada lleugerament la precisió, perquè el buidatge estira l'asímptota cap avall. En tercer lloc, el biaix estimat actua com a diagnòstic robust de la convergència: saturat a zero en la sobreestimació, inflat per sobre del real en la subestimació, i estable i proper al real en el control.

Val la pena explicar per què passa aquesta asimetria. Quan l'asímptota inicial és massa alta (sobreestimació), la restricció $b \geq 0$ impedeix que el biaix absorbeixi l'excés, de manera que tota la pressió correctora recau sobre A; com que els tiquets tardans del buidatge tenen un biaix real molt baix, són gairebé mesures directes del pes real i ajuden a baixar l'asímptota cada cop millor. En canvi, quan l'asímptota inicial és massa baixa (subestimació), el filtre disposa de tota la llibertat per inflar el biaix, i ho fa fins a valors propers als 20 kg; un cop el biaix s'ha quedat fixat en aquest valor alt, els tiquets tardans (amb biaix baix) no encaixen amb l'estat del filtre i no aconsegueixen corregir l'asímptota, així que l'error es manté o fins i tot puja una mica. El control queda al mig: com que l'asímptota ja és correcta, l'únic efecte del buidatge és que $\hat{A}$ baixa una mica, perquè el filtre confon la caiguda del biaix real amb un sostre una mica més baix.

Aquests resultats canvien una conclusió que havíem tret abans: amb pocs tiquets i sense modelar el buidatge, semblava que la freqüència gairebé no influïa. Ara que sí modelem el buidatge, veiem que la freqüència depèn del signe de l'error inicial. El que importa no és només quants tiquets es reben, sinó en quin moment del buidatge arriben i amb quin biaix.

A la pràctica, l'asímptota estimada influeix en la decisió d'enviament, però cal tenir en compte que la corba logística és poc sensible a canvis d'A al final del cicle, quan els animals ja s'acosten al seu sostre. Per tant, un error de pocs quilograms en $\hat{A}$ (com en el control) gairebé no afecta el calendari d'enviaments, mentre que els errors de 10–16 kg dels casos amb hipòtesi equivocada sí que el poden desplaçar prou.

Finalment, val la pena comparar el filtre amb el model base sense corregir, que és la corba teòrica inicial del ramader (la línia "Predicció Base sense filtre" de les figures). Aquest model base manté l'error inicial durant tot el cicle: en els escenaris de ± 25 kg, l'asímptota es quedaria a 250 o 200 kg de forma permanent, amb un error d'uns 25 kg respecte al valor real. El filtre, en canvi, redueix aquest error en tots els casos: fins a un MAE de 10,90 kg en la sobreestimació amb buidatge i a menys de 2 kg en el control. Així, fins i tot en els escenaris menys favorables, el recalibratge dinàmic millora clarament la predicció respecte a mantenir el model teòric original, cosa que valida l'objectiu principal del treball.

6 CONCLUSIONS

Per concloure, s'ha dissenyat, implementat i validat un Filtre de Kalman Estès d'Estat Augmentat per al recalibratge dinàmic de la corba logística de creixement porcí. Les Podriem dividir les conclusions en tres nivells: tècnic, experimental i pràctic.

A nivell tècnic, l'EKF d'Estat Augmentat que estima simultàniament l'asímptota A i el biaix de selecció b ha demostrat ser una arquitectura robusta. La clau del seu funcionament és la Jacobiana de l'Eq. (2), que permet distribuir la innovació de cada tiquet entre la correcció de l'asímptota i la del biaix de forma òptima. La incorporació del coneixement previ que el biaix és positiu ($b \geq 0$, amb $b_0 = 8$ kg) evita resultats impossibles i, com hem vist als experiments, fa que el valor final del biaix serveixi per saber si el filtre ha convergit bé.

A nivell experimental, els dotze experiments confirmen que la hipòtesi inicial és el factor més important per a la precisió, però aporten una cosa nova sobre la freqüència: el seu efecte és diferent segons el signe de l'error inicial. En la sobreestimació, el buidatge progressiu de la granja és beneficiós, perquè els enviaments tardans tenen un biaix de selecció molt baix i actuen com a observacions netes del pes real, permetent que el filtre corregeixi l'asímptota a la baixa i reduint el MAE de $14{,}77$ a $10{,}90$ kg. En canvi, en el control i la subestimació la freqüència degrada lleugerament la precisió, ja que la caiguda del biaix durant el buidatge estira l'estimació de l'asímptota cap avall. El biaix estimat resumeix bé aquesta diferència: es queda a zero quan l'asímptota inicial és massa alta, i es dispara molt quan és massa baixa.

A nivell pràctic, aquests resultats es tradueixen en una guia d'actuació senzilla per al ramader, articulada al voltant de tres situacions. Si sospita que ha **sobreestimat** l'asímptota. Per exemple, perquè els primers tiquets surten sistemàticament per sota del que esperava, li convé *augmentar la freqüència d'enviaments*, sobretot en la fase final del buidatge: aquests tiquets tardans tenen un biaix de selecció molt baix i permeten al filtre corregir l'asímptota a la baixa, reduint l'error de manera notable. Si, en canvi, creu que la seva **hipòtesi inicial és bona**, no necessita multiplicar els enviaments: amb pocs tiquets el filtre ja separa correctament el biaix de l'asímptota i manté l'error per sota de 2 kg; fer-ne més només introdueix una lleugera degradació per l'efecte del buidatge. Finalment, si sospita que ha **subestimat** l'asímptota, augmentar la freqüència no l'ajudarà: el filtre tendeix a inflar el biaix estimat i aquest valor actua d'ancoratge que impedeix recuperar l'asímptota, de manera que el més recomanable és revisar directament la hipòtesi de partida.

Un avantatge operatiu addicional és que el ramader pot utilitzar el **valor final del biaix estimat com a senyal de diagnòstic** sense conèixer la veritat absoluta: un valor saturat a zero indica sobreestimació; un d'inflat (voral dels 8 kg) apunta a subestimació; i un d'estable i proper a aquest valor confirma que la hipòtesi inicial era encertada. Amb tot, el factor determinant és partir d'una bona estimació inicial d'A, ja que la corba logística perd sensibilitat al final del cicle i un error gran és difícil de recuperar només amb més observacions.

Com a línies futures, es preveu: (i) validar el sistema amb dades reals d'escorxadors[cite: 2]; (ii) incorporar variables auxiliars com el consum de pinso o la temperatura ambiental; (iii) desenvolupar una interfície de visualització en temps real per al tècnic de granja; i (iv) analitzar la sensibilitat del filtre als hiperparàmetres (P_0, Q i R_t) mitjançant múltiples llavors aleatòries.

Declaració d'ús d'Intel·ligència Artificial

En aquest treball s'hawn utilitzat eines d'intel·ligència artificial com a suport per a l'optimització del codi dels models i per a la revisió estilística de la redacció de la memòria. L'ús d'aquestes tecnologies s'ha limitat estrictament a una funció d'assistència i supervisió de la feina realitzada per l'autor.

BIBLIOGRAFIA

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[2] J. Berkson, "Limitations of the Application of the Logistic Function to Bio-Assay," *Journal of the American Statistical Association*, vol. 39, no. 227, pp. 357–365, 1944.

[3] F. J. Richards, "A Flexible Growth Function for Empirical Use," *Journal of Experimental Botany*, vol. 10, no. 2, pp. 290–301, 1959.

[4] J. J. Heckman, "Sample Selection Bias as a Specification Error," *Econometrica*, vol. 47, no. 1, pp. 153–161, 1979.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[6] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *Technical Report TR 95-041*, University of North Carolina, 1995.

[7] E. A. Wan and A. T. Nelson, "Dual Extended Kalman Filter Methods," in *Kalman Filtering and Neural Networks* (S. Haykin, ed.), ch. 5, pp. 123–173. New York, NY: John Wiley & Sons, 2001.

[8] Y. Wang *et al.*, "Estimation of Growth Curves and Suitable Slaughter Weight of the Liangshan Pig," *Asian-Australasian Journal of Animal Sciences*, vol. 28, no. 9, pp. 1252–1258, 2015.

[9] W. B. Roush, T. L. Cravener, and F. Zhang, "Computer Monitoring of Slope-Ratio Assays Using a Kalman Filter," *Computers and Electronics in Agriculture*, vol. 8, no. 1, pp. 1–13, 1992.

[10] J. Van Bebber, N. Reinsch, W. Junge, and E. Kalm, "Monitoring Daily Milk Yields with a Recursive Test Day Repeatability Model (Kalman Filter)," *Journal of Dairy Science*, vol. 82, no. 11, pp. 2421–2429, 1999.

[11] B. Gompertz, "On the Nature of the Function Expressive of the Law of Human Mortality," *Philosophical Transactions of the Royal Society of London*, vol. 115, pp. 513–583, 1825.

[12] M. E. Turner, E. L. Bradley, K. A. Kirk, and K. M. Pruitt, "A Theory of Growth," *Mathematical Biosciences*, vol. 29, no. 3–4, pp. 367–373, 1976.

[13] I. McMillan, R. Friars, and J. Quinton, "Age Correction of Performance Traits in Swine," *Journal of Animal Science*, vol. 46, no. 6, pp. 1565–1572, 1978.

[14] H. D. Patterson and R. Thompson, "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrika*, vol. 58, no. 3, pp. 545–554, 1971.

APÈNDIX A RESULTATS COMPLETS DE LA BATERIA D'EXPERIMENTS

A continuació es presenten la resta de figures de la bateria d'experiments que no apareixen a la secció de resultats, completant les quatre freqüències (3, 6, 9 i 12 enviaments) per a cadascuna de les tres hipòtesis inicials. En tots els casos l'asímptota real és $A_r = 225$ kg.

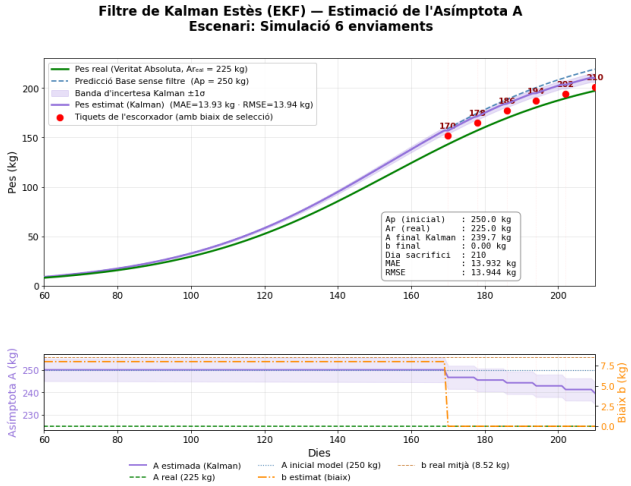

Fig. 5: Sobreestimació ($A_p = 250$ kg) amb 6 enviaments ($\hat{A} = 239{,}7$ kg, MAE $= 13{,}93$ kg).

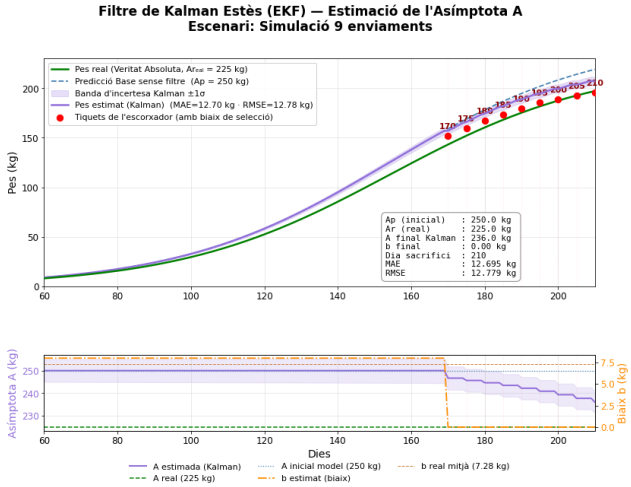

Fig. 6: Sobreestimació ($A_p = 250$ kg) amb 9 enviaments ($\hat{A} = 236{,}0$ kg, MAE $= 12{,}70$ kg).

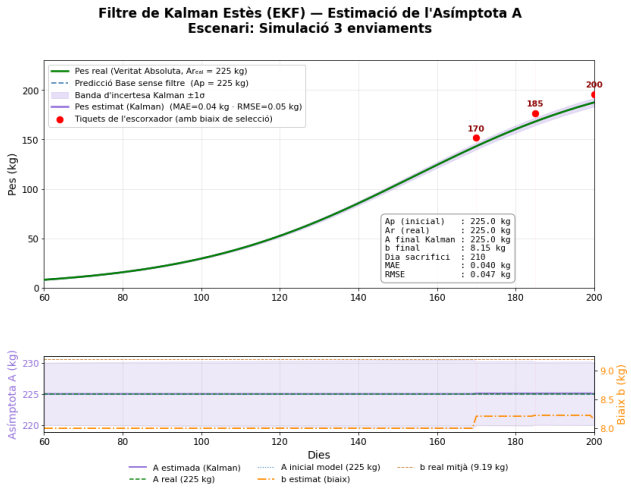

Fig. 7: Control ($A_p = A_r = 225$ kg) amb 3 enviaments. Les tres corbes se solapen i el filtre corregeix únicament el biaix de selecció ($\hat{b} = 8{,}15$ kg, MAE $= 0{,}04$ kg).

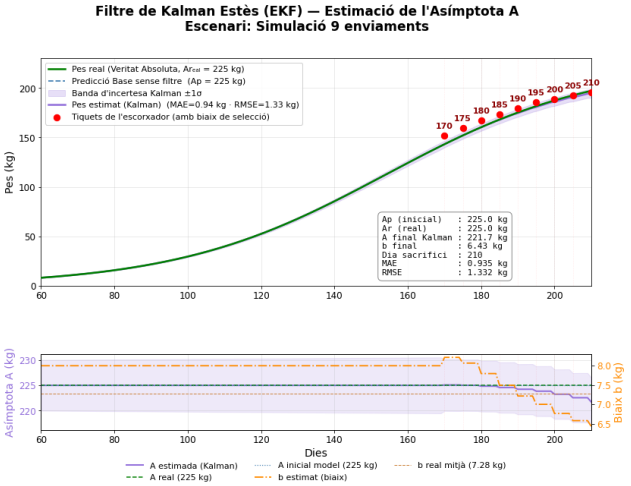

Fig. 8: Control ($A_p = 225$ kg) amb 9 enviaments ($\hat{A} = 221{,}7$ kg, MAE $= 0{,}94$ kg).

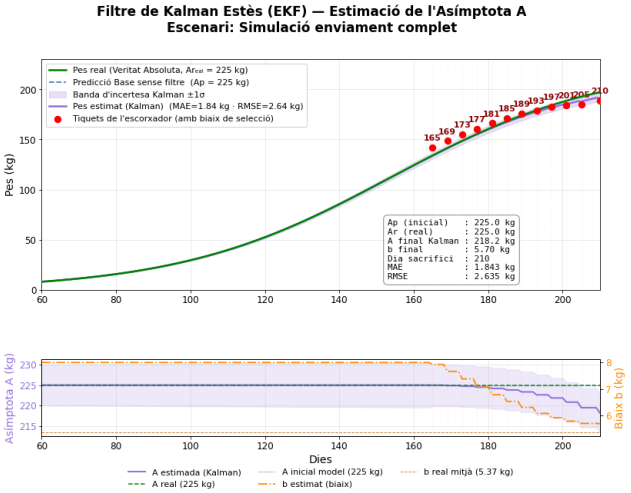

Fig. 9: Control ($A_p = 225$ kg) amb 12 enviaments, buidatge complet ($\hat{A} = 218{,}2$ kg, MAE $= 1{,}84$ kg).

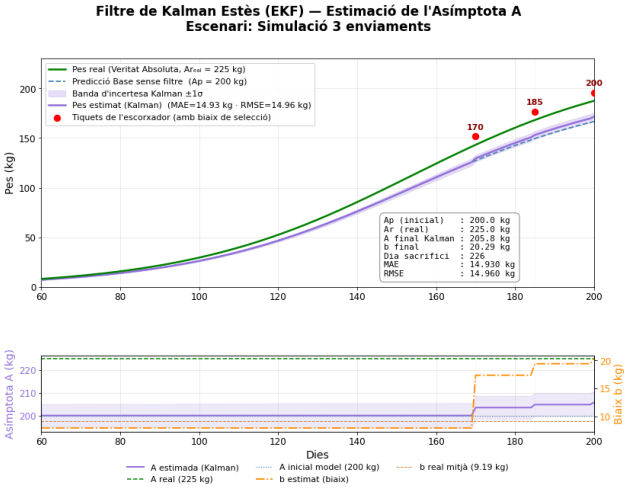

Fig. 10: Subestimació ($A_p = 200$ kg) amb 3 enviaments. L'asímptota puja de 200 a 205,8 kg i el biaix s'infla fins a 20,29 kg per compensar el dèficit.

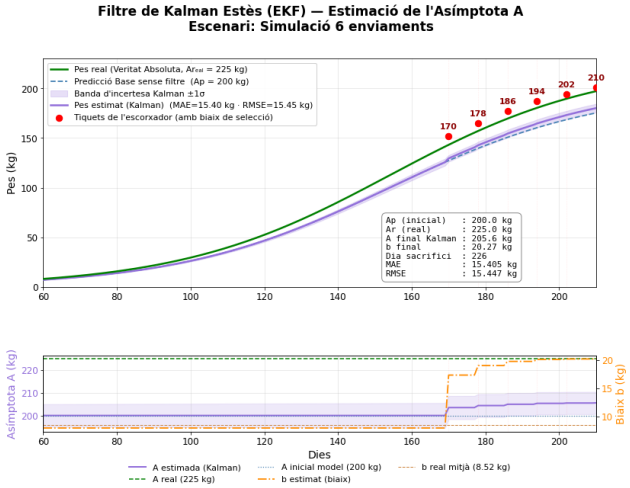

Fig. 11: Subestimació ($A_p = 200$ kg) amb 6 enviaments ($\hat{A} = 205{,}6$ kg, MAE $= 15{,}41$ kg).

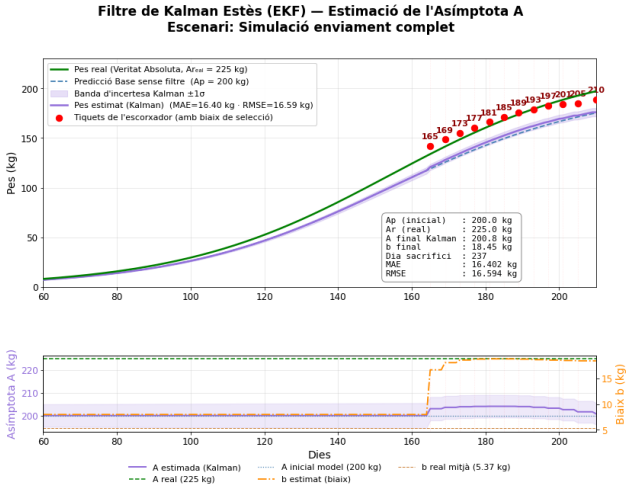

Fig. 12: Subestimació ($A_p = 200$ kg) amb 12 enviaments, buidatge complet ($\hat{A} = 200{,}8$ kg, MAE $= 16{,}40$ kg).