

Kengetallen

E-38

Pseudo-records

▪ **Inleiding**

In December 2014 heeft GES een nieuwe methode voor fokwaardeschatting geïntroduceerd: het pseudo-record systeem (het PSR systeem). In dit systeem wordt alle informatie (ouders, nakomelingen, eigen waarneming en genomische informatie) meegenomen in één berekeningsprocedure. Dit systeem produceert *genomically enhanced breeding values* (met genomics versterkte fokwaarden, GEBV).

Het pseudo-record systeem is een manier van fokwaardeschatting waarin genomische fokwaarden (*direct genomic values* of DGV) worden gebruikt als een vierde bron van informatie, naast gegevens over eigen prestaties, ouders en nakomelingen. De DGV van gegenotypeerde dieren worden behandeld als een observatie aan een virtuele kenmerk (het pseudo-kenmerk) dat is gecorreleerd met het bestaande kenmerk. Zo wordt een DGV een observatie aan het DNA van een dier voor dit pseudo-kenmerk. Deze observatie wordt het pseudo-record van een dier genoemd. Omdat we de DGV gebruiken als een observatie aan een pseudo-kenmerk, is het pseudo-record van een dier gelijk aan zijn DGV (Stoop et al., 2014).

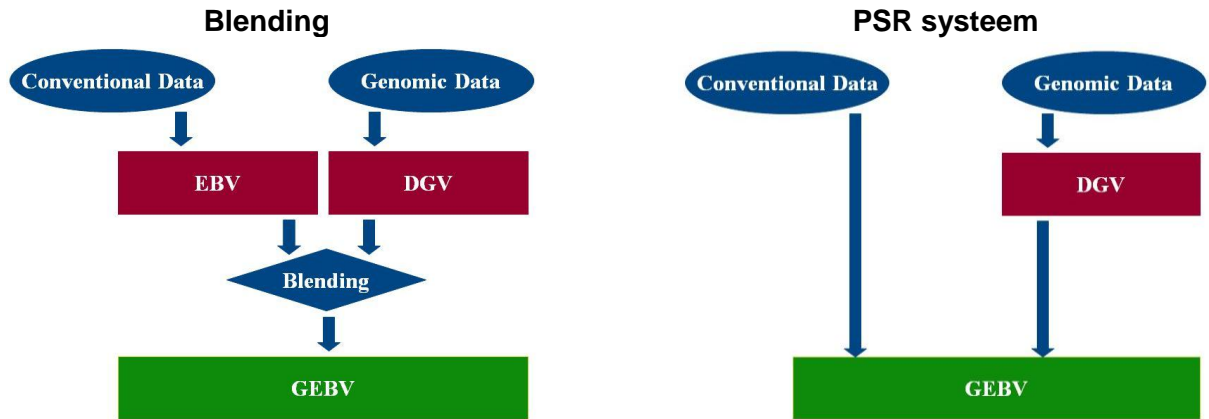
In de fokwaardeschatting kunnen pseudo-kenmerken hetzelfde behandeld worden als bestaande kenmerken, wat resulteert in één integrale fokwaardeschattings-methode waarin alle kenmerken gelijktijdig worden geëvalueerd. De correlatie tussen bestaand kenmerk en pseudo-kenmerk zorgt voor een natuurlijke integratie van genomische data met conventionele data (dat wil zeggen: gegevens van ouders, nakomelingen en eigen prestaties). De resulterende fokwaarden voor bestaande kenmerken zijn daarmee GEBV: *genomically enhanced breeding values* (met genomics versterkte fokwaarden).

Bijvoorbeeld, gegenotypeerde stieren hebben een DGV voor het kenmerk 'melkproductie', die wordt gebruikt als pseudo-record: eigen waarneming (het eigen DNA) voor het kenmerk 'pseudo-melkproductie'. Een jonge stier, gegenotypeerd maar zonder dochters, zal meer informatie beschikbaar hebben dan alleen het gemiddelde van zijn ouders (de verwachtingswaarde). Via de correlatie tussen de kenmerken 'melkproductie' en 'pseudo-melkproductie', zal het PSR systeem een fokwaarde voor 'melkproductie' opleveren, die is samengesteld uit de verwachtingswaarde en de informatie die beschikbaar is omtrent zijn DNA, in de vorm van zijn DGV.

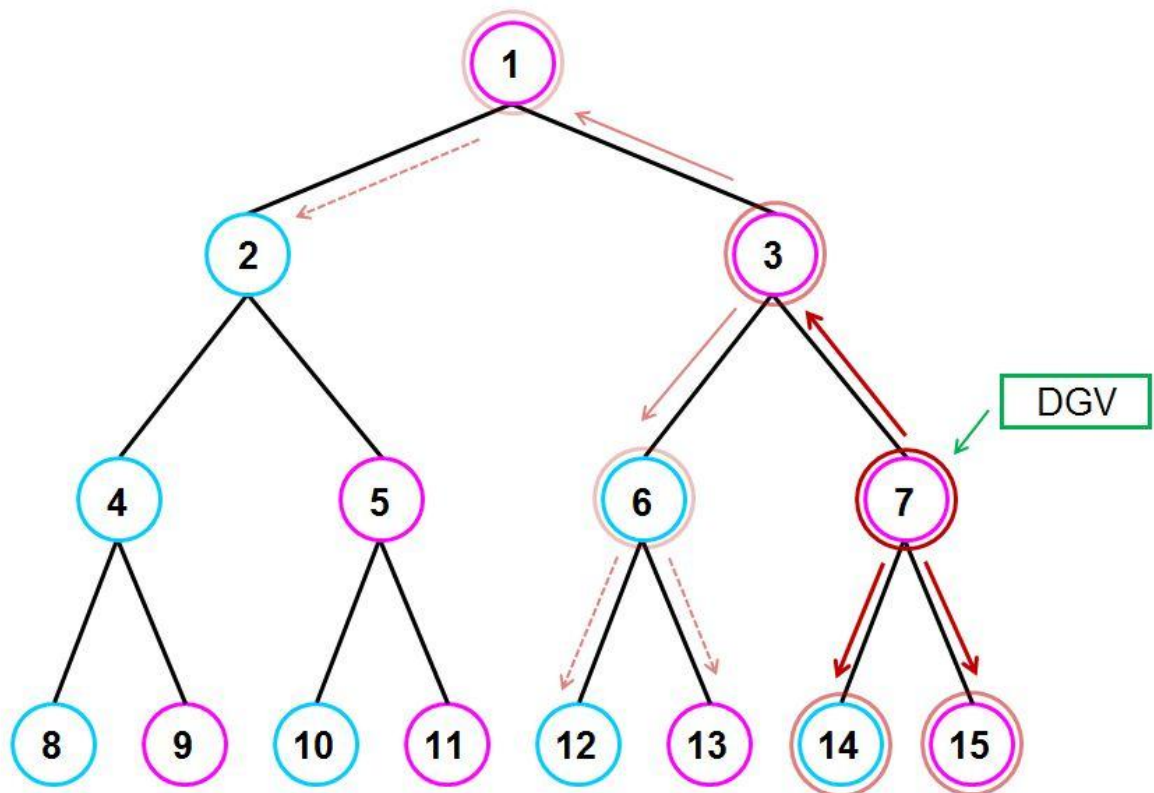
Deze methode heeft een aantal voordelen ten opzichte van de oude integratie-methode, *blending* genaamd. De drie belangrijkste zijn:

1. Vereenvoudiging van het integratie proces
2. Efficiënter gebruik van genomische data
3. Voorkomen van over- of onderschatting door genomische pre-selectie

Ad 1) Bij *blending* bestaat de fokwaardeschatting uit twee gescheiden processen, de conventionele fokwaardeschatting en de schatting van genomische fokwaarden. Deze worden gecombineerd in een nabewerking, de daadwerkelijke *blending* van fokwaarden. Met het PSR systeem is een aparte integratie-stap niet meer nodig, omdat de genomische data (DGV) direct wordt geïntegreerd met conventionele data om GEBV te verkrijgen. In Figuur 1 is het verschil in proces tussen *blending* en *pseudo-records* schematisch weergegeven.



Figuur 1. De integratie van genomische en conventionele data bij *blending* (l.) en bij gebruik van pseudo-records (r.) in schematische vorm. Bij **blending** bestaat de **GEBV** schatting uit twee gescheiden processen, de conventionele **EBV** (gebaseerd op **conventionele data**) en de genomische evaluatie (**DGV** o.b.v. **genomische data**), die worden gecombineerd in een aparte nabewerking, de eigenlijke **blending** van fokwaarden. Met pseudo-records is een aparte integratie stap niet langer nodig, omdat de **genomische data** (**DGV**) direct wordt geïntegreerd met de **conventionele data**.



Figuur 2. Illustratie van het effect van een DGV in het PSR-systeem op de stamboom van een gegetypeerd dier. Stier 7 is gegetypeerd en heeft een DGV die wordt gebruikt in het PSR-systeem. Dit resulteert in een toegenomen betrouwbaarheid van de fokwaarde, aangeduid met de rode cirkel, omdat het nu GEBV is. Echter, een deel van deze toename (ongeveer $\frac{1}{4}$ ervan) wordt doorgegeven aan zijn nakomelingen (dieren 14 en 15) en zijn ouders (stier 3, in dit geval), aangeduid met de lichter gekleurde cirkels. Deze fokwaarden zijn feitelijk ook GEBV. Een kwart van de toename van betrouwbaarheid voor vader 3 wordt verder doorgegeven in de pedigree. Dit effect verdwijnt vrij snel: de toename van de betrouwbaarheid in dieren 12 en 13 is slechts ongeveer $\frac{1}{64}$ van de toename van betrouwbaarheid in stier 7.

Ad 2) Als een dier een geldig DGV heeft zullen zijn GEBV meer betrouwbaar zijn dan de conventionele fokwaarden. In het PSR-systeem is deze toename van betrouwbaarheid niet langer beperkt tot het dier zelf. Zijn ouders en nakomelingen (indien niet gegenotypeerd) profiteren ook door de structuur van de stamboom. Dit wordt geïllustreerd in Figuur 2.

Ad 3) Genomische selectie kan een over- of onderschatting veroorzaken in conventionele fokwaardeschatting als in de groep dieren met een genomische waarde de gemiddelde genetische aanleg van nakomelingen, paringsdieren of leeftijdsgenoten verschilt van die van dieren met een conventionele fokwaarde. Dit kan gebeuren doordat de *Mendelian sampling* term (verschil tussen fokwaarde en verwachtingswaarde) in getypeerde dieren nauwkeuriger bekend is dan in conventioneel geschatte dieren, of door niet willekeurige paring van dieren gebaseerd op genomische fokwaarden. Door genomische waarden van getypeerde dieren direct in te passen in de pedigree van de gehele populatie wordt deze over- of onderschatting vermeden.

▪ Het principe van pseudo-records

Het basisidee in het PSR systeem is dat de DGV van een getypeerd dier gebruikt wordt als een observatie aan een pseudo-kenmerk, het pseudo-record. Als pseudo-records beschikbaar zijn voor een bepaald kenmerk, wordt het pseudo-kenmerk gebruikt in een multi-kenmerk analyse, waar het is gecorreleerd met het bestaande kenmerk (Mäntysaari en Strandén; 2010). Als dit correct wordt gedaan kan het pseudo-kenmerk in de fokwaardeschatting hetzelfde worden behandeld als een bestaand kenmerk.

In formule-vorm ziet het systeem van bestaand kenmerk – pseudo-kenmerk er uit als:

$$\begin{bmatrix} Y \\ D \end{bmatrix}_i = \begin{bmatrix} a_Y \\ a_D \end{bmatrix}_i + \begin{bmatrix} e_Y \\ e_D \end{bmatrix}_i$$

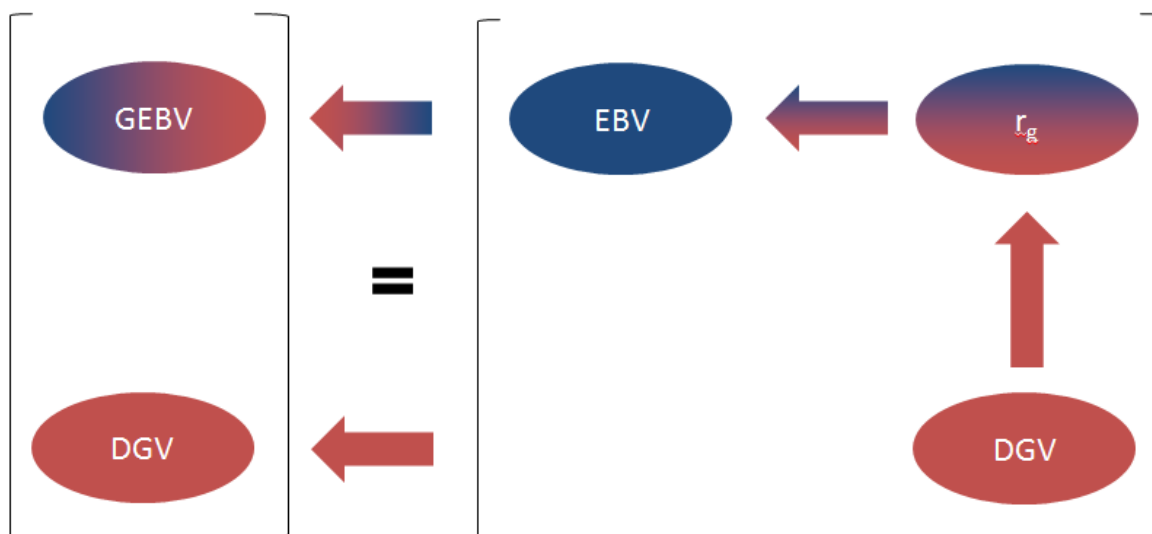
en

$$\text{var} \begin{bmatrix} a_Y \\ a_D \end{bmatrix} = \begin{bmatrix} \sigma_g^2 & r_g \sigma_g^2 \\ & \sigma_g^2 \end{bmatrix}$$

Waar Y en D observaties zijn aan het bestaande en het pseudo-kenmerk. De variabelen a_Y en a_D zijn fokwaarden voor Y en D . In deze context correspondeert a_Y met de GEBV en a_D met de DGV van een dier. Als laatste zijn e_Y en e_D de schattingsfouten voor Y en D .

De tweede formule beschrijft de genetische relatie tussen Y en D , waar σ_g de genetische spreiding is. Omdat de DGV een genomische fokwaarde is voor kenmerk Y is de spreiding van het pseudo-kenmerk D gelijk aan de genetische spreiding van het bestaande kenmerk Y .

De integratie van genomische data met conventionele gegevens vindt plaats in de tweede formule hierboven. Deze formule neemt de genomische fokwaarde (a_D) en gebruikt de correlatie r_g (rechtsboven in de rechter matrix) om de informatie van a_D te integreren in a_Y (zie Figuur 3). De genetische correlatie van het pseudo-kenmerk met het bestaande kenmerk (r_g) is evenredig aan de betrouwbaarheid van de genomische informatie in de DGV. Vandaar dat de hoeveelheid informatie uit de DGV (a_D) in de fokwaarde a_Y (de GEBV) wordt beperkt tot de nauwkeurigheid waarmee de DGV werd geschat.



Figuur 3. Schematische weergave van de manier waarop genomische data (DGV) de fokwaarde van een dier verandert in een GEBV volgens het eerder beschreven systeem van formules. Rechts in de figuur zorgt de genetische correlatie r_g genomische informatie uit de DGV door wordt gegeven naar de conventionele fokwaarde EBV, zodanig dat deze in het eindresultaat (links) GEBV is. Omdat voor het gecorreleerde genomische kenmerk een erfelijkheidsgraad van 1 wordt aangenomen, is de DGV links gelijk aan de DGV rechts.

Het pseudo-kenmerk heeft een aangenomen erfelijkheidsgraad van (bijna) één. De fokwaarde a_D van een getypeerd dier is daardoor gelijk aan de DGV van het dier. Omdat de erfelijkheidsgraad ≈ 1 , hebben gerelateerde dieren (ouders of nakomelingen) met een DGV geen invloed op de fokwaarde a_D . Daarmee worden dubbeltellingen van (genomische) informatie en overschatting van de betrouwbaarheid van de GEBV voorkomen.

▪ Dataselectie

De observaties aan een pseudo-kenmerk in het PSR systeem worden gevormd door de DGV uit een genomische evaluatie. Dieren met DGV hebben een corresponderend pseudo-record in het PSR systeem, als:

- Het dier een stier is en
 - 10 maanden of ouder is en eigendom van een KI organisatie die deelneemt aan de genomische fokwaardeschatting
 - Een Eurogenomics dier is
 - Geen KI stier is en afgevoerd
- Of als het dier vrouwelijk is.

Momenteel zijn er twee KI organisaties die deelnemen aan de Nederlandse nationale genomische evaluaties: CRV en KI Kampen. Eurogenomics is een Europees netwerk van KI bedrijven die genomische fokwaardeschattingen doen. Dit netwerk is opgericht om DGV nauwkeuriger te kunnen schatten door referentie-populaties samen te voegen. Om dit te bereiken hebben de deelnemers aan Eurogenomics afgesproken genotype informatie van hun stieren uit te wisselen voor gebruik in genomische evaluaties. In de praktijk betekent dit dat een Eurogenomics stier een Nederlands/Vlaamse DGV in het PSR-systeem zal hebben.

▪ Bepaling van parameters

Voor elk pseudo-kenmerk moet de conventionele fokwaardeschatting worden uitgebreid met genetische en residuele variantiecomponenten om de link tussen pseudo-kenmerk en bestaand kenmerk te modelleren. Om dit correct te doen is de (genetische) correlatie tussen een pseudo-kenmerk en geassocieerd bestaande kenmerk nodig. De benodigde correlatie is gelijk aan de wortel van de betrouwbaarheid van het totale merker effect r_{mark}^2 (Stoop et al., 2014). Daardoor verandert het hierboven beschreven PSR model in:

$$\begin{bmatrix} Y \\ D \end{bmatrix}_i = \begin{bmatrix} a_Y \\ a_D \end{bmatrix}_i + \begin{bmatrix} e_Y \\ e_D \end{bmatrix}_i$$
$$\text{var} \begin{bmatrix} a_Y \\ a_D \end{bmatrix} = \begin{bmatrix} \sigma_g^2 & \sqrt{r_{mark}^2} \cdot \sigma_g^2 \\ & \sigma_g^2 \end{bmatrix}$$

Merk op dat r_g is vervangen door $\sqrt{r_{mark}^2}$. In het PSR systeem mag een pseudo-kenmerk alleen het bestaande kenmerk beïnvloeden waarmee het geassocieerd is. In een multi-trait fokwaardeschatting zijn daardoor de covarianties tussen een pseudo-kenmerk en alle andere, niet-geassocieerde, kenmerken evenredig aan de correlatie tussen pseudo-kenmerk en het bestaande kenmerk. Voor een grondiger, technische beschrijving van het afleiden van correcte parameter matrices voor gebruik in een fokwaardeschatting met meerdere (pseudo-) kenmerken, verwijzen we naar Appendix A.

▪ Methode of GEBV schatting

Vanwege de aard van het PSR-systeem, kan het alleen worden toegepast op fokwaardeschattingen die werken met behulp van een stelsel van lineaire vergelijkingen (*generalized linear models*). Op dit moment wordt het PSR-systeem toegepast in de volgende fokwaardeschattingen:

- Productie
- Exterieur
- Vruchtbaarheid
- Uiergezondheid
- Klauwgezondheid
- Karakter & Melksnelheid
- Levensvatbaarheid
- Geboortegemak
- Vleesindex

Pseudo-kenmerken worden in alle fokwaardeschattingen waarin het PSR-systeem wordt toegepast op dezelfde manier geanalyseerd. Het statistisch model dat wordt gebruikt is:

$$y_{pi} = \mu_p + \text{animal}_{pi} + \text{error}_{pi}$$

Waar:

- μ : het gemiddelde
- y : de DGV voor PSR kenmerk p van dier i
- animal : het (genetische) dier-effect
- error : het residuele effect (schattingsfout)

Omdat $h^2 \approx 1$, zal het effect *animal* gelijk zijn aan het pseudo-record, de DGV. Omdat de schatting van DGV afhankelijk is van nauwkeurige en actuele conventionele fokwaarden, vindt de genetische evaluatie met het PSR-systeem plaats aan het einde van het schattingsproces. Het proces van genetische evaluatie gebeurt in de volgende volgorde:

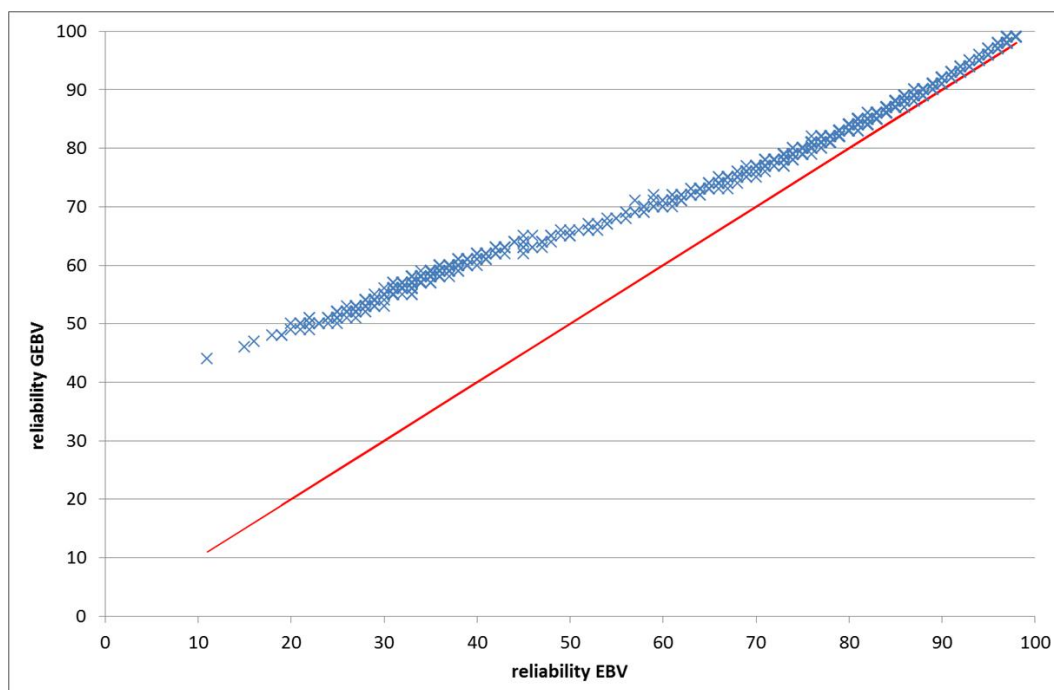
- 1) Conventionele fokwaardeschatting NL/VL
- 2) DGV schatting op basis van 1) en de vorige MACE run
- 3) Fokwaardeschatting NL/VL met behulp van het PSR systeem

Dit zorgt ervoor dat de PSR evaluatie met de meest recente DGV als pseudo-observaties begint.

Tabel 1 geeft een overzicht van alle kenmerken waarvoor DGV (en dus pseudo-records) beschikbaar zijn. De tabel toont per kenmerk de betrouwbaarheid van het merker-effect en de genetische correlatie tussen pseudo-kenmerk en bestaande kenmerk. Merk op dat voor een aantal fokwaardeschattingen een pseudo-kenmerk voor een index ontbreekt (o.a. totaal exterieur, uiergezondheidsindex). Deze indexen bestaan uit onderliggende kenmerken waarvoor al pseudo-records bestaan. Een pseudo-kenmerk van een index voegt dan geen nieuwe informatie toe en wordt daarom weggelaten. De uiteindelijke indexen worden gevormd uit de GEBV van de onderliggende kenmerken, inclusief de toegenomen betrouwbaarheid.

▪ GEBV betrouwbaarheden

Ten opzichte van conventionele fokwaarden laten GEBV een toegenomen betrouwbaarheid zien, met name voor jonge stieren met weinig of geen dochters. Zie Figuur 3, bijvoorbeeld, waar de betrouwbaarheid voor de GEBV Uiergezondheid (in blauw) is gegeven ten opzichte van de betrouwbaarheid van de conventionele fokwaarde (rode lijn) voor stieren met een DGV. Voor stieren met al een relatief betrouwbare index (rond 70%) is de afstand tussen de rode lijn en de blauwe "x"-tekens bescheiden. Voor jonge stieren zonder dochters neemt de betrouwbaarheid het meeste toe.



Figuur 3. Betrouwbaarheid van GEBV voor Uiergezondheid versus die van conventionele fokwaarden (EBV) voor getypeerde stieren. GEBV betrouwbaarheden in blauwe "x", rode lijn geeft de betrouwbaarheid van de conventionele EBV weer.

Tabel 1. Overzicht van alle kenmerken waarvoor pseudo-records beschikbaar zijn, hun erfelijkheidsgraad (h^2) de betrouwbaarheid van de marker-gegevens (r^2_{mark}) en de genetische correlatie tussen pseudo- en bestaande kenmerk (r_g). De weergegeven waarden zijn op basis van de December 2014-evaluatie.

BVE	Kenmerk	h^2	r^2_{mark}	r_g
Productie	kg Melk	0,57	0,688	0,829
	kg Vet	0,58	0,639	0,799
	kg Eiwit	0,50	0,605	0,778
Exterieur	Hoogtemaat	0,52	0,686	0,828
	Voorhand	0,24	0,641	0,801
	Inhoud	0,31	0,588	0,767
	Openheid	0,11	0,620	0,787
	Conditie-score	0,30	0,653	0,808
	Kruisligging	0,34	0,642	0,801
	Kruisbreedte	0,40	0,669	0,818
	Beenstand achter	0,15	0,465	0,682
	Beenstand zij	0,23	0,639	0,799
	Klauwhoek	0,14	0,610	0,781
	Beengebruik	0,14	0,482	0,694
	Vooruieraanhechting	0,27	0,600	0,775
	Voorspeenplaatsing	0,38	0,653	0,808
	Speenlengte	0,38	0,659	0,812
	Uierdiepte	0,38	0,713	0,844
	Achteruierhoogte	0,23	0,602	0,776
	Ophangband	0,23	0,661	0,813
	Achterspeenplaatsing	0,32	0,662	0,814
	Frame	0,28	0,621	0,788
	Robuustheid	0,14	0,266	0,516
Uier	0,29	0,623	0,790	
Beenwerk	0,16	0,503	0,709	
Uiergezondheid	Celgetal	0,37	0,632	0,795
	Subkl. mastitis	0,06	0,644	0,803
	Klin. mastitis	0,06	0,569	0,754
Afkalven	Direct geboortegemak	0,07	0,717	0,847
	Maternaal geboortegemak	0,05	0,384	0,620
	Directe levensvatbaarh.	0,04	0,326	0,571
	Maternale levensvatbaarh.	0,09	0,643	0,802
Vruchtbaarheid	Non-return 56	0,04	0,485	0,696
	Interval afkalv. – 1 st inseminatie	0,17	0,625	0,791
	Tussenkalftijd	0,15	0,624	0,790
	Interv. 1 ^{ste} – laatste inseminatie	0,08	0,635	0,797
Karakter & melksnelheid	Melksnelheid	0,23	0,587	0,766
	Karakter	0,12	0,482	0,694
Klauwgez.	Klauwgezondheidsindex	0,18	0,317	0,563
Vleesind.	Vleesindex	0,25	0,564	0,751

▪ Literatuur

Mäntysaari, E.A. & Strandèn, I. (2010) Proc. 9th WCGALP, Leipzig, Germany.

Stoop W.M., H. Eding, M.L. van Pelt, L.C.M. de Haer en G. de Jong, (2014) Proc. 10th WCGALP, Vancouver, Canada

▪ Appendix A

Deze appendix is een meer gedetailleerde, technische uitleg over de afleiding van parameter matrices voor gebruik in een fokwaardeschatting met het PSR-systeem. In het eerste deel wordt de afleiding van matrices uitgelegd wanneer meerdere pseudo-kenmerken bij een fokwaardeschatting zijn betrokken. In het tweede worden pseudo-kenmerken van indexen behandeld, die correleren met een aantal onderliggende bestaande kenmerken.

Basis formule van gecorreleerde genomische kenmerken

Het basis idee voor opname van pseudo-records in een nationale fokwaardeschatting werd geformuleerd door Mantysaari en Strandèn (2010):

$$\begin{bmatrix} Y \\ P \end{bmatrix}_i = \begin{bmatrix} Z_Y & 0 \\ 0 & Z_P \end{bmatrix} \begin{bmatrix} a_Y \\ a_P \end{bmatrix}_i + \begin{bmatrix} e_Y \\ e_P \end{bmatrix}_i$$

$$\text{var} \begin{bmatrix} a_Y \\ a_P \end{bmatrix} = \mathbf{G} = \begin{bmatrix} \sigma_{g,Y}^2 & r_g \sigma_{g,Y}^2 \\ & \sigma_{g,Y}^2 \end{bmatrix}$$

$$\text{var} \begin{bmatrix} e_Y \\ e_P \end{bmatrix} = \mathbf{E} = \begin{bmatrix} \sigma_{e,Y}^2 & 0 \\ & (1-h^2)\sigma_{g,Y}^2 \end{bmatrix}$$

Waar Y en P het bestaande en het pseudo-kenmerk zijn, a_Y en a_P de fokwaarden voor Y en P en e_Y en e_P de error op Y en P . Z_Y en Z_P zijn incidentie matrices die observaties aan dieren linken.

Conventionele en genomische fokwaarde informatie wordt gekoppeld door middel van een genetische correlatie r_g die gelijk is aan de wortel van de betrouwbaarheid van de DGV (minus stamboom informatie). Meestal wordt de toename in betrouwbaarheid uitgedrukt als een toename in *expected daughter contributions* (EDC) van DGV ten opzichte van stamboom informatie, dat wil zeggen oudergemiddelden of stier-indexen. Als erfelijkheidsgraad h^2 wordt (bijna) één aangenomen. Vandaar dat de residuele variantie voor gecorreleerde genomische kenmerken een kleine fractie $(1-h^2)$ van de genetische variantie is.

Basis structuur van de genetische covariantie matrix

De matrix \mathbf{G}_c is een genetische covariantie-matrix van N conventionele kenmerken. De matrix \mathbf{G}_{22} is dat deel van \mathbf{G}_c dat kenmerken bevat waarvoor pseudo-records bestaan. Dan is de matrix \mathbf{G}_c op te delen als:

$$\mathbf{G}_c = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}$$

De vector \mathbf{r}^2 is een vector met betrouwbaarheden van de merker effecten r_{mark}^2 van P pseudo-kenmerken verkregen uit een validatie van genomische fokwaarden. Om de correlaties tussen pseudo-kenmerk en corresponderend bestaand kenmerk te verkrijgen transformeren we \mathbf{r}^2 naar een diagonale matrix \mathbf{R} , zodanig dat $R_{i,i} = \sqrt{r_i^2}$ en $R_{i,j \neq i} = 0$.

$$\mathbf{R} = \begin{bmatrix} \sqrt{r_1^2} & 0 & 0 & 0 \\ & \sqrt{r_2^2} & 0 & 0 \\ & & \ddots & 0 \\ & & & \sqrt{r_P^2} \end{bmatrix}$$

We nemen aan dat $h^2 \approx 1.0$ voor DGV. Daardoor is de genetische covariantie-structuur tussen pseudo-kenmerken gelijk aan die tussen de bestaande kenmerken in \mathbf{G}_{22} . De correcte genetische covariantie matrix is dan:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_c & \mathbf{G}_{c,psr} \\ & \mathbf{G}_{psr} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} & \mathbf{G}_{12}\mathbf{R} \\ & \mathbf{G}_{22} & \mathbf{G}_{22}\mathbf{R} \\ & & \mathbf{G}_{22} \end{bmatrix}$$

Deze structuur zorgt dat kenmerken in \mathbf{G}_{psr} alleen hun evenknie in \mathbf{G}_{22} beïnvloeden, zonder direct effect op elk ander kenmerk.

Afleiding van \mathbf{G} voor meerdere pseudo-kenmerken

De structuur van de \mathbf{G} matrix boven beschreven is eenvoudig uit te rekenen, wanneer gebruik wordt gemaakt van een matrix Φ zodanig dat:

$$\Phi = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \\ \mathbf{0} & \mathbf{W} \end{bmatrix}$$

Waar \mathbf{I}_1 en \mathbf{I}_2 identiteitsmatrices zijn, corresponderend met \mathbf{G}_{11} en \mathbf{G}_{22} . De matrix \mathbf{W} is een matrix die de relatie tussen pseudo-kenmerk en corresponderend bestaand kenmerk beschrijft. Als elk kenmerk in \mathbf{G}_{22} correspondeert met één enkel pseudo-kenmerk, is matrix $\mathbf{W} = \mathbf{I}_2$. Met deze matrix berekenen we:

$$\mathbf{G}^* = \Phi \mathbf{G} \Phi'$$

De resulterende matrix \mathbf{G}^* is $N+P$ groot, d.w.z. covarianties van pseudo-kenmerken zijn nu opgenomen in de matrix. Echter, in \mathbf{G}^* is de correlatie tussen pseudo-kenmerk en bestaand kenmerk nog niet verrekend. Voor de correcte \mathbf{G} matrix delen we \mathbf{G}^* op, zodanig dat conventionele en pseudo-kenmerken zijn gescheiden, en vermenigvuldigen de niet-diagonale deelmatrices met \mathbf{R} :

$$\mathbf{G}_{psr} = \begin{bmatrix} \mathbf{G}_{cc}^* & \mathbf{G}_{cp}^* \mathbf{R} \\ & \mathbf{G}_{pp}^* \end{bmatrix}$$

Waar het subscript \mathbf{c} conventionele kenmerken aangeeft en subscript \mathbf{p} de pseudo-kenmerken.

Afleiding \mathbf{G} voor samengestelde kenmerken en indices

De boven beschreven methode is eenvoudig uit te breiden naar pseudo-kenmerken die lineaire combinaties van bestaande kenmerken zijn (bijv., als het pseudo-kenmerk een overall kenmerk is, terwijl de bestaande conventionele kenmerken lactatie-specifiek zijn). Om dit te doen moet de matrix \mathbf{W} met relaties tussen pseudo-kenmerk en corresponderende bestaande kenmerken de gewichten van de lineaire combinatie bevatten.

Als voorbeeld nemen we twee overall kenmerken, waarvoor pseudo-records bestaan, die elk een index zijn van twee onderliggende kenmerken. De matrix \mathbf{W} is dan:

$$\mathbf{W} = \begin{bmatrix} w_{11} & 0 & w_{12} & 0 \\ 0 & w_{21} & 0 & w_{22} \end{bmatrix}$$

Waar w_{ij} het gewicht is van de j -de kenmerk in index i . Matrix Φ is nu $\Phi = [\mathbf{I}; \mathbf{W}]$ gelijk aan de eerder beschreven methode om \mathbf{G}^* uit te rekenen. Vermenigvuldiging van \mathbf{G}_{pc} (en \mathbf{G}_{cp} !) met een geschikte \mathbf{R} resulteert in correcte covarianties tussen pseudo-kenmerken, onderliggende bestaande kenmerken en alle andere kenmerken in \mathbf{G} .

De residuele covariantie matrix

Residuele covarianties tussen pseudo-kenmerken worden verondersteld 0 te zijn. En omdat pseudo-kenmerken een erfelijkheidsgraad hebben van $h^2 \approx 1$, kan de residuele covariantie matrix \mathbf{E} berekend worden als:

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{psr} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_c & \mathbf{0} \\ \mathbf{0} & (1-h^2) \cdot \text{diag}(\mathbf{G}_{pp}) \end{bmatrix}$$

Waar $\text{diag}(\mathbf{G}_{pp})$ een diagonale matrix is met de diagonale elementen van \mathbf{G}_{pp} .