

Non parametric vs. GBLUP model for genomic evaluation with large reference population in Holstein cattle

*N. Charfeddine¹, S. T. Rodríguez-Ramilo^{1,2}, J. A. Jiménez-Montero³,
M. J. Carabaño² and Ó. González-Recio^{2,4}*



¹ CONAFE, Madrid, Spain

² INIA, Madrid, Spain

³ UPM, Madrid, Spain

⁴ Biosciences Research Division, Victoria, Australia



BACKGROUND



□ **November 2012:**

Official Genomic Evaluation was implemented in SPAIN

❖ **Genomic model:**

NON PARAMETRIC MODEL: R-BOOSTING

Ongoing monitoring showed underestimation of true values for some traits:

**Rear legs side view, Rear legs rear view, Overall Feet & legs
Longevity and Days open**

OBJECTIVE

The aim of this study was:

To compare the non parametric model :

Reproducing Kernel Hilbert Spaces Regression (RKHS)

with the most popular parametric models :

GBLUP (GB)

GBLUP with polygenic effect (GBP)

GENOTYPES



EUROGENOMICS reference population

After edits

18, 443 genotyped bulls were used

After editing : 36,971 SNPs were used

PHENOTYPES

Sire deregressed Proofs (DRP)

Reference (training) set:

Sires < 2005 with EBV
INTERBULL Jan'2009

Validation set :

Sires 2005-2007 with EBV in
INTERBULL Dec'2011
Minimum EDC= 15



Traits	h ²	Reference set	Validation set
Production traits			
Milk yield (MY)	0.28	14487	3943
Fat yield (FY)	0.28	14487	3940
Protein yield (PY)	0.28	14487	3938
Fat % (FP)	0.35	14487	3940
Protein % (PP)	0.31	14487	3938
Type traits			
Stature (STA)	0.45	14308	3782
Chest width (CW)	0.17	13127	3733
Body depth (BD)	0.27	13751	3573
Angularity (ANG)	0.25	13126	3943
Rump angle (RA)	0.34	14261	3062
Rump width (RW)	0.31	14249	3943
Rear legs side view (RLSV)	0.16	14308	3752
Rear legs rear view (RLRV)	0.11	12141	3790
Foot angle (FA)	0.12	13450	3688
Fore udder attachment (FUA)	0.19	13516	3689
Rear udder attachment (RUA)	0.23	12900	3706
Suspensory ligament (SL)	0.21	14304	3178
Udder depth (UD)	0.30	14306	3767
Fore teat placement (FTP)	0.30	14357	3685
Rear teat placement (RTP)	0.30	12044	3760
Teat length (TL)	0.29	14306	3771
Feet & legs (FL)	0.15	13664	2889
Functional Traits			
Somatic cell score (SCS)	0.175	14487	3938
Longevity (LONG)	0.115	13854	226
Days open (DO)	0.043	14198	3753

Non parametric GENOMIC MODELS



RKHS (Gianola et al.2006)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\alpha} + \mathbf{e}$$

,Where:

\mathbf{y} : vector of DR proofs in the training set

$\mathbf{X}\boldsymbol{\beta}$: parametric term ($\mathbf{1}\mu$)

$\mathbf{K}\boldsymbol{\alpha}$: nonparametric term

\mathbf{K} : is a positive definite matrix of kernels = Genomic Matrix (VanRaden, 2008)

$\boldsymbol{\alpha}$: is a vector of nonparametric coefficients distributed as $N(\mathbf{0}, \mathbf{K}^{-1}\sigma_{\alpha}^2)$

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{N}^{-1}\sigma_e^2)$$

Genomic breeding value $\mathbf{u} = \mathbf{K}\boldsymbol{\alpha}$

RKHS was solved in a Bayesian context using Gibbs Sampling

Parametric GENOMIC MODELS

GB

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

,where:

\mathbf{y} : vector of DR proofs of sires in the training set

μ : Overall mean

\mathbf{Z} : incidence matrix

\mathbf{g} : vector of random marker effects .

$$\mathbf{g} \sim N(0, \mathbf{G} \sigma_g^2)$$

\mathbf{e} : random residual effect $\mathbf{e} \sim N(0, \mathbf{I} \sigma_e^2)$

Genomic breeding value $\mathbf{u} = \mathbf{Z}\mathbf{g}$

GBP

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g}_\omega + \mathbf{e}$$

,where:

$$\mathbf{g}_\omega = \mathbf{u} + \mathbf{g}$$

\mathbf{u} : vector of residual polygenic effects

\mathbf{g} : vector of random marker effects

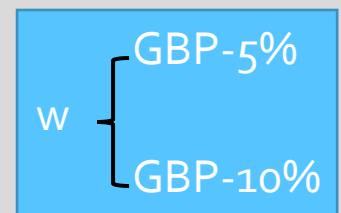
$$\text{Var}(\mathbf{g}_\omega) = [\omega\mathbf{A} + (1-\omega)\mathbf{G}] \sigma_{g_\omega}^2 \quad \mathbf{G}_\omega = \omega\mathbf{A} + (1-\omega)\mathbf{G}^*$$

\mathbf{A} : pedigree-based relationship matrix

\mathbf{G}^* : adjusted genomic relationship matrix (Gao et al., 2012)

ω : ratio of residual polygenic to total additive genetic variance

\mathbf{e} : random residual effect $\mathbf{e} \sim N(0, \mathbf{I} \sigma_e^2)$



\mathbf{G}^{-1} and \mathbf{G}_ω^{-1} were calculated using program provided by Ismo Strandén.

G-BLUP solutions were calculated using Mix99



CRITERIA FOR MODEL COMPARISON

- Weighted Pearson Correlation between Predicted DGV and DRP
- Regression coefficient of DRP on DGV
- Mean Square Error
- Selection effectiveness: Percentage of bulls truly ranked

**Mean and Confidence intervals were estimated using
Bootstrapped samples**

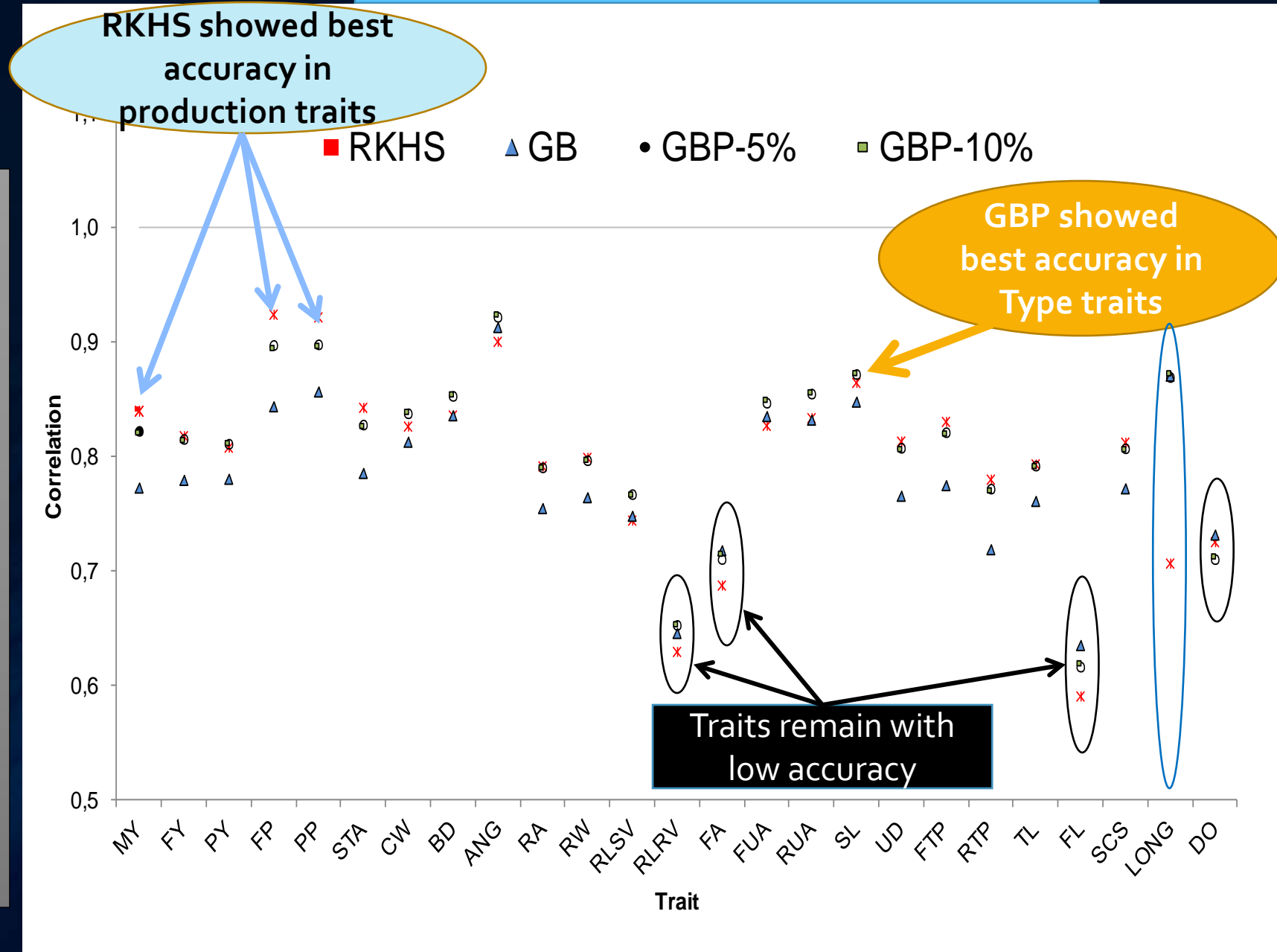
**1000 samples were drawn with
replacement from the whole testing set**

RESULTS

CORRELATION

- CI overlapped across models
- RKHS showed best accuracy than simple GB
- GB results improved including polygenic effect
- GBP-5% and GBP-10% showed the same accuracy

Figure 1. Correlation of predictions

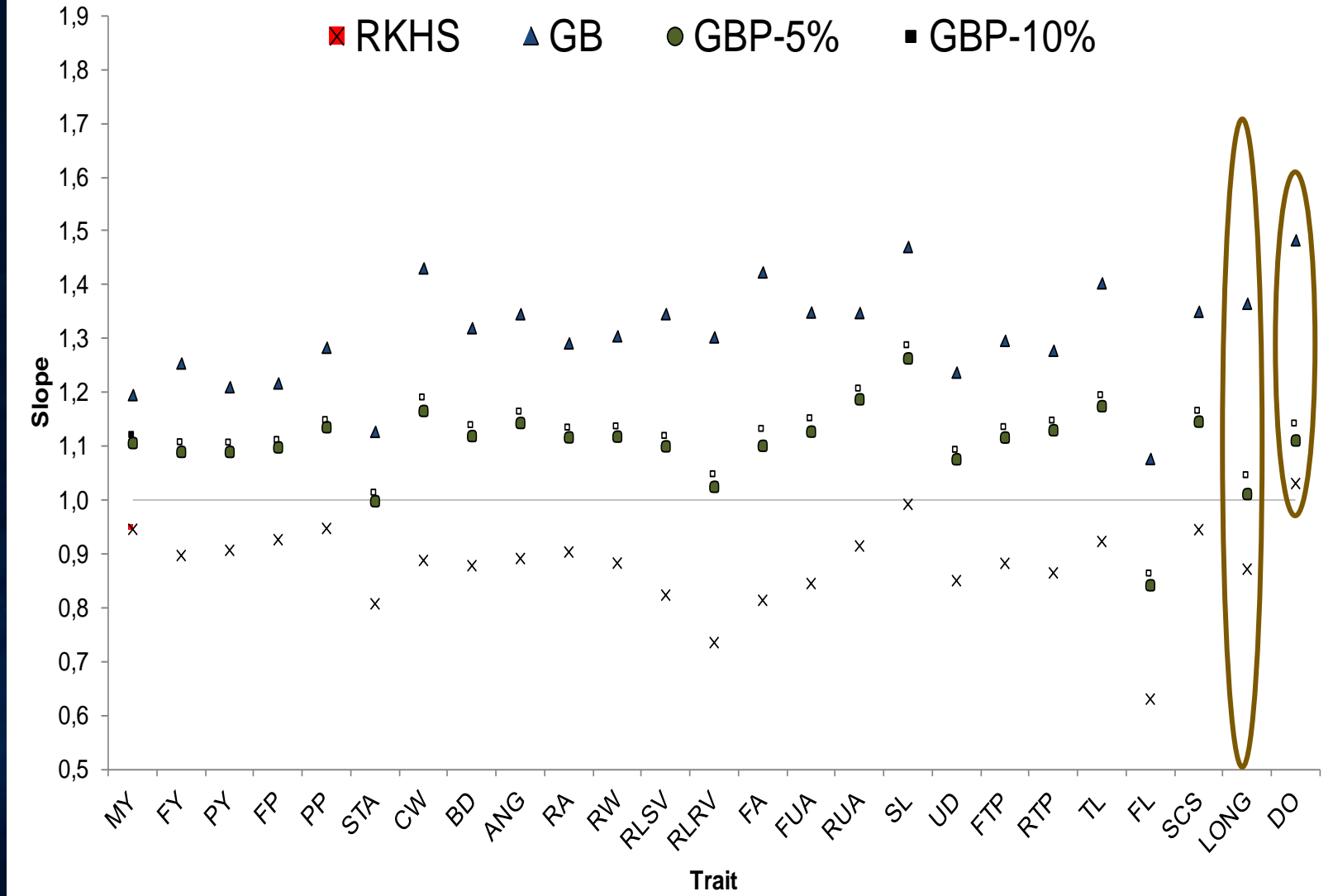


RESULTS

REGRESSION COEFFICIENT

- RKHS b-values < 1
- GB b-values > 1.
- GB led to less bias than GB
- GBP-5% showed slightly better result than GBP-10%
- LONG and DO CI overlapped across models

Figure 2. slope of predictions



RESULTS

MEAN SQUARED ERROR



Production traits

Trait	RKHS	GB	GBP-5%	GBP-10%
MY	177438.45 (169456.35 - 185618.44)	237846.66 (227314.84 - 247675.13)	196069.36 (187370.43 - 204453.62)	198435.17 (190483.75 - 207213.51)
FY	300.36 (287.43 - 315.38)	358.92 (343.71 - 374.58)	307.86 (295.32 - 322.29)	310.79 (297.58 - 325.50)
PY	210.49 (200.32 - 219.33)	271.67 (260.71 - 283.52)	227.00 (217.10 - 237.71)	229.06 (217.27 - 239.22)
FP	0.02 (0.02 - 0.02)	0.03 (0.03 - 0.03)	0.02 (0.02 - 0.02)	0.02 (0.02 - 0.03)
PP	0.01 (0.01 - 0.01)	0.01 (0.01 - 0.01)	0.01 (0.01 - 0.01)	0.01 (0.01 - 0.01)

- RKHS showed smaller MSE in production traits

RESULTS



Type traits

MEAN SQUARED ERROR

- MSE CI overlapped across models
- GBP showed smaller MSE in type traits
- GBP-5% and GBP-10% showed similar MSE

Trait	RKHS	GB	GBP-5%	GBP-10%
STA	0.51 (0.49 - 0.54)	0.54 (0.52 - 0.56)	0.47 (0.45 - 0.49)	0.47 (0.45 - 0.49)
CW	1.39 (1.33 - 1.46)	1.50 (1.43 - 1.57)	1.37 (1.31 - 1.44)	1.38 (1.32 - 1.44)
BD	1.03 (0.98 - 1.08)	1.08 (1.03 - 1.13)	0.99 (0.95 - 1.04)	1.00 (0.95 - 1.04)
ANG	0.83 (0.78 - 0.87)	0.91 (0.86 - 0.96)	0.82 (0.77 - 0.86)	0.82 (0.77 - 0.87)
RA	0.82 (0.78 - 0.86)	0.91 (0.87 - 0.95)	0.81 (0.77 - 0.85)	0.82 (0.78 - 0.86)
RW	0.84 (0.80 - 0.88)	0.94 (0.89 - 0.98)	0.84 (0.80 - 0.88)	0.84 (0.81 - 0.89)
RLSV	1.31 (1.24 - 1.37)	1.32 (1.25 - 1.39)	1.24 (1.18 - 1.30)	1.24 (1.19 - 1.31)
RLRV	2.02 (1.93 - 2.12)	2.02 (1.92 - 2.11)	1.93 (1.83 - 2.02)	1.94 (1.85 - 2.03)
FA	2.21 (2.11 - 2.32)	2.23 (2.12 - 2.34)	2.15 (2.06 - 2.27)	2.15 (2.05 - 2.25)
FUA	1.25 (1.19 - 1.31)	1.44 (1.38 - 1.51)	1.27 (1.21 - 1.33)	1.28 (1.22 - 1.33)
RUA	0.99 (0.94 - 1.04)	1.20 (1.14 - 1.25)	1.03 (0.98 - 1.08)	1.04 (0.99 - 1.09)
SL	1.30 (1.24 - 1.37)	1.57 (1.50 - 1.65)	1.40 (1.33 - 1.47)	1.42 (1.35 - 1.49)
UD	0.71 (0.68 - 0.74)	0.85 (0.81 - 0.89)	0.71 (0.68 - 0.75)	0.72 (0.69 - 0.76)
FTP	0.67 (0.64 - 0.71)	0.81 (0.77 - 0.85)	0.69 (0.66 - 0.72)	0.70 (0.67 - 0.73)
RTP	0.60 (0.58 - 0.63)	0.70 (0.67 - 0.74)	0.61 (0.58 - 0.64)	0.62 (0.59 - 0.64)
TL	0.79 (0.75 - 0.83)	0.90 (0.85 - 0.94)	0.80 (0.76 - 0.84)	0.81 (0.77 - 0.85)
FL	2.32 (2.22 - 2.44)	2.39 (2.27 - 2.52)	2.29 (2.17 - 2.39)	2.30 (2.18 - 2.40)

RESULTS

MEAN SQUARED ERROR



Functional traits

Trait	RKHS	GB	GBP-5%	GBP-10%
SCS	93.78 (89.75 - 98.55)	109.93 (104.78 - 114.54)	96.95 (92.72 - 101.12)	97.84 (93.66 - 102.54)
LONG	546.86 (448.09 - 649.12)	470.10 (383.25 - 553.54)	463.45 (376.62 - 548.32)	461.52 (381.03 - 550.24)
DO	409.63 (384.94 - 434.70)	413.87 (387.97 - 438.51)	410.64 (387.59 - 435.65)	412.10 (387.93 - 437.77)

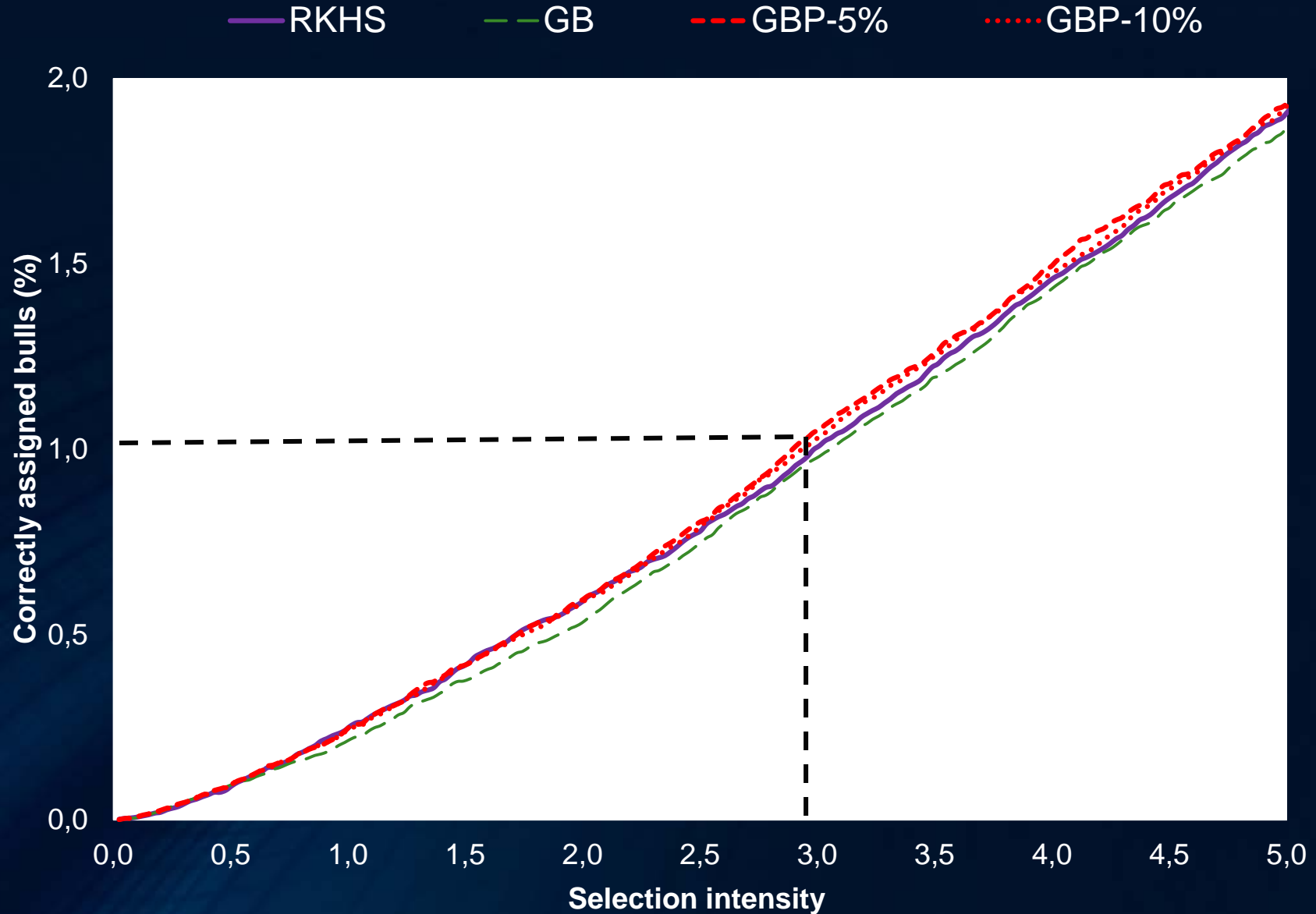
- RKHS showed smaller MSE in SCS and DO
- GBP-10% showed smaller MSE in LONG
- CI overlapped in DO
- CI >> in LONG

RESULTS

SELECTION EFFECTIVENESS

- Parametric and Non parametric models showed similar effectiveness in ranking top bulls
- Polygenic effect improve slightly bull ranking by GBLUP model

Figure 3. Percentage of correctly assigned bulls for a given selection intensity



CONCLUSIONS



- ❑ RKHS performed better results than simple GBLUP model
- ❑ Polygenic effect improved GBLUP results and matched those of RKHS
- ❑ RKHS involve a slight overestimation while GBLUP display an underestimation
- ❑ The weighting factors (5 and 10%) of polygenic effect showed similar results for most traits
- ❑ Inclusion of polygenic effect in RKHS approach could further improve its performance
- ❑ Further research is needed to improve results of some traits and to optimize the computational requirements of RKHS approach

Merci!

