Contents

Contents.....- 3 -

Index of Tables.....- 5 -

Index of figures .....- 8 -

Summary .....- 10 -

Chapter 1

General Introduction:.....- 27 -

Objectives .....- 79 -

Chapter 2

Genotyping strategies for genomic selection in small dairy cattle populations.....- 85 -

Chapter 3

Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle .....- 117 -

Chapter 4

The gradient boosting algorithm and random Boosting for genome-assisted evaluation in large data sets.....- 147 -

Chapter 5

Predictive ability of dairy cattle genotypes imputed from different density platforms.....- 177 -

Chapter 6

General Discussion..... - 207 -

Final Conclusions..... - 229 -

Index of Tables

TABLE 2.1. AVERAGE DIFFERENCES IN THE ACCURACY OF PREDICTED GBVS AND STANDARD DEVIATIONS (IN PARENTHESIS) FOR EACH SELECTIVE GENOTYPING STRATEGY[A] VERSUS THE SIRESDYD[B] STRATEGY BASED ON THE HERITABILITY AND USE OF DIFFERENT FEMALE TRAINING SETS AND POPULATION SIZES FROM A CONTEMPORARY POPULATION OF 40,000 ANIMALS ..... - 100 -

TABLE 2.2. BIAS AND MEAN SQUARE ERROR (MSE) OF GENOMIC PREDICTIONS IN THE TESTING SET FOR DIFFERENT GENOTYPING STRATEGIES, TRAINING SET SIZE AND HERITABILITY ..... - 101 -

TABLE 2.3. AVERAGES AND STANDARD DEVIATIONS OF INTERCEPTS, OF GENOMIC PREDICTIONS IN THE TESTING SET, FOR DIFFERENT GENOTYPING STRATEGY, TRAINING SET SIZE AND HERITABILITY REGRESSIONS ..... - 103 -

TABLE 2.4. AVERAGES AND STANDARD DEVIATIONS OF SLOPES OF GENOMIC PREDICTIONS IN THE TESTING SET, FOR DIFFERENT GENOTYPING STRATEGY, TRAINING SET SIZE AND HERITABILITY REGRESSIONS ..... - 103 -

TABLE 2.5. AVERAGES AND STANDARD DEVIATIONS OF COEFFICIENTS OF DETERMINATION OF GENOMIC PREDICTIONS IN THE TESTING SET, FOR DIFFERENT GENOTYPING STRATEGY, TRAINING SET SIZE AND HERITABILITY REGRESSIONS ..... - 105 -

TABLE 3.1. ACCURACY, STANDARDIZED BIAS IN MEANS, BIAS IN REGRESSION COEFFICIENTS AND MEAN SQUARED ERROR (MSE) OF GENOMIC PREDICTIONS FOR DIFFERENT EVALUATION METHODOLOGIES AND FIVE TRAITS OF ECONOMIC INTEREST IN SPANISH DAIRY CATTLE ..... - 133 -

TABLE 4.1. PEARSON CORRELATION[1] BETWEEN PREDICTED AND OBSERVED RESPONSES IN THE TESTING SET USING THE ORIGINAL GRADIENT BOOSTING ALGORITHM (MTRY=100%) OR ITS MODIFIED VERSION "RANDOM BOOSTING", FOR DIFFERENT VALUES OF PERCENTAGE OF SNPS SAMPLED AT EACH ITERATION (MTRY) AND SMOOTHING PARAMETER (V) ... - 171 -

TABLE 4.2. ESTIMATED BIAS[1] (MEASURED AS AVERAGE DIFFERENCE BETWEEN PREDICTED AND OBSERVED RESPONSES IN STANDARD DEVIATION UNITS) IN THE TESTING SET USING THE ORIGINAL GRADIENT BOOSTING ALGORITHM (MTRY=100%) OR ITS MODIFIED VERSION "RANDOM BOOSTING", FOR DIFFERENT VALUES OF PERCENTAGE OF SNPS SAMPLED AT EACH ITERATION (MTRY) AND SMOOTHING PARAMETER (V) .....- 172 -
TABLE 4.3. COMPUTATION TIME[1] (IN HOURS) TO RUN 10-FOLD CROSS VALIDATIONS (A COMPLETE GENOMIC ASSISTED EVALUATION CYCLE) REGARDING THE VALUE OF THE SMOOTHING PARAMETER (*V*) AND THE PROPORTION OF SNPS SAMPLED AT EACH ITERATION (*MTRY*) .....- 173 -
TABLE 5.1. ACCURACY FOR THE GENOMIC ESTIMATION OF TWO EVALUATION METHODS INDEXED FOR FOUR TRAITS OF ECONOMIC INTEREST IN DAIRY CATTLE AFTER THE IMPUTATION FROM 3K, 6K AND 50K TO 50K AND HD. MEAN OF THE 1000 REPLICATES AFTER BOOTSTRAPPING AND CONFIDENCE INTERVALS CONSIDERED AS THE NARROWEST GAP CONTAINING 95% OF THE REPLICATES .....- 191 -
TABLE 5.2. REGRESSION COEFFICIENTS FOR THE GENOMIC ESTIMATION OF TWO EVALUATION METHODS INDEXED FOR FOUR TRAITS OF ECONOMIC INTEREST IN DAIRY CATTLE AFTER THE IMPUTATION FROM 3K, 6K AND 50K TO 50K AND HD. MEAN OF THE 1000 REPLICATES AFTER BOOTSTRAPPING AND CONFIDENCE INTERVALS CONSIDERED AS THE NARROWEST GAP CONTAINING 95% OF THE REPLICATES .....- 193 -
TABLE 5.3. MEAN SQUARED ERRORS FOR THE GENOMIC ESTIMATION OF TWO EVALUATION METHODS INDEXED FOR FOUR TRAITS OF ECONOMIC INTEREST IN DAIRY CATTLE AFTER THE IMPUTATION FROM 3K, 6K AND 50K TO 50K AND HD. MEAN OF THE 1000 REPLICATES AFTER BOOTSTRAPPING AND CONFIDENCE INTERVALS CONSIDERED AS THE NARROWEST GAP CONTAINING 95% OF THE REPLICATES .....- 194 -

TABLE 5.4. CONFUSION MATRICES FOR THE CLASSIFICATION OF ANIMALS IN FIVE CLASSES ACCORDING TO THEIR RANKING REGARDING OBSERVED DRPS OF FOUR TRAITS OF ECONOMIC INTEREST IN DAIRY CATTLE USING TWO EVALUATION METHODS AFTER THE IMPUTATION FROM 6K TO 50K AND FROM 50K TO HD. OBSERVED AND PREDICTED CLASSES IN ROWS AND COLUMNS RESPECTIVELY ..... - 197 -

TABLE 5.5 RATE OF ANIMALS CORRECTLY CLASSIFIED ACCORDING TO THEIR RANKING IN FIVE CLASSES EACH ONE CONTAINING 20% OF THE VALUES (OVERALL), CORRECTLY CLASSIFIED IN THE FIRST CLASS (TOP 20%), OR WITHIN THE THREE HIGHEST CLASSES (TOP 60 %). RESULTS SHOWED FOR FOUR TRAITS OF ECONOMIC INTEREST IN DAIRY CATTLE USING TWO EVALUATION METHODS AFTER THE IMPUTATION FROM 3K, 6K AND 50K TO 50K AND HD ..... - 198 -

Index of figures

FIGURE 2.1 DISTRIBUTION OF SIMULATED QTL EFFECTS: (A) 0.30 HERITABILITY TRAIT SCENARIO AND (B) 0.10 HERITABILITY TRAIT SCENARIO. .....- 93 -

FIGURE 2.2. DISTRIBUTION OF THE NUMBER OF DAUGHTERS PER SIRE IN (A) 0.30 HERITABILITY TRAIT SCENARIO AND (B) 0.10 HERITABILITY TRAIT SCENARIO. .....- 95 -

FIGURE 2.3. DISTRIBUTION OF R^2 BETWEEN SINGLE-NUCLEOTIDE POLYMORPHISM (SNP) PAIRS AND PHYSICAL DISTANCE: (A) CHROMOSOME 1 FOR THE 0.10 HERITABILITY TRAIT AND (B) CHROMOSOME 7 FOR THE 0.30 HERITABILITY TRAIT. .....- 98 -

FIGURE 2.4 ESTIMATED ACCURACIES FOR GENOMIC BREEDING VALUES FOR TWO DIFFERENT HERITABILITIES (0.10 AND 0.30) IN TESTING SETS WHEN 1000, 2000, OR 5000 FEMALES IN THE TRAINING SET WERE GENOTYPED. THE FOLLOWING GENOTYPING STRATEGIES WERE USED: COWS AT RANDOM (RND), TOP YIELD DEVIATION COWS (TOPYD), TOP BREEDING VALUE COWS (TOPBV), TWO-TAILED YIELD DEVIATION COWS (TTYD), TWO-TAILED BREEDING VALUE COWS (TTBV), ALL SIRES (SIRESDYD), AND PEDIGREE INDEX WITHOUT GS. .....- 99 -

FIGURE 3.1. NUMBER OF GENOTYPED BULLS BY YEAR OF BIRTH. .....- 131 -

FIGURE 3.2. DISTRIBUTION OF MINOR ALLELE FREQUENCIES (MAF) OF THE SNPS AFTER QUALITY CONTROL. .....- 132 -

FIGURE 3.3. AVERAGE LINKAGE DISEQUILIBRIUM (MEASURED AS R^2) AND CONFIDENCE INTERVAL (ESTIMATED BY R PACKAGE GPLOTS) BETWEEN SYNTENIC MARKERS WITH RESPECT TO THEIR PHYSICAL DISTANCE. .....- 133 -

FIGURE 5.1. DIAGRAM OF THE DESIGN OF REFERENCE AND VALIDATION SETS AND PROCESS OF IMPUTATION ACCURACY EVALUATION FROM 3K AND 6K TO 50K. .....- 185 -

FIGURE 5.2. PERCENTAGE OF COMMON BULLS IN THE OBSERVED AND PREDICTED RANKINGS WHEN LESS OR EQUALS THAN TOP 10% OF GENOMICALLY EVALUATED BULLS ARE SELECTED REGARDING FAT PERCENTAGE. COMPARISON BETWEEN 50K (—) AND HD (×) GENOTYPES. .- 196 -