Running head: Predicting fertility from seminal traits

Predicting fertility from seminal traits: performance of several parametric and non-parametric procedures[1]

M. Piles[*1], J. Díez[†], J.J. del Coz[†], E. Montañés[†], J. R. Quevedo[†], J. Ramon[*], O. Rafel[*], M. López-Béjar[‡], and L. Tusell[*]

[*] Unitat de Cunicultura, Institut de Recerca i Tecnologia Agroalimentàries, Torre Marimon s/n, 08140, Caldes de Montbui, Barcelona, Spain

[†] Artificial Intelligence Center, University of Oviedo at Gijón, Asturias, Spain

[‡] Dpt. Animal Health and Anatomy, Universitat Autònoma de Barcelona, 08193, Barcelona, Spain

[1] Corresponding author: miriam.piles@irta.es

ABSTRACT: This research aimed at assessing the efficacy of non-parametric procedures to improve the classification of the ejaculates in the artificial insemination (**AI**) centers according to their fertility rank predicted from characteristics of the AI doses. A total of 753 ejaculates from 193 bucks were evaluated at 3 different times from 5 to 9 mo of age for 21 seminal variables (related to ejaculate pH and volume, sperm concentration, viability, morphology and acrosome reaction traits, and dose characteristic) and their corresponding fertility score after AI over crossbred females. Fertility rate was categorized into 5 classes of equal length. Linear Regression (**LR**), Ordinal Logistic Regression (**OLR**), Support Vector Regression (**SVR**), Support Vector Ordinal

Regression (**SVOR**), and Non-deterministic Ordinal Regression (**NDOR**) were compared in terms of their predictive ability with two base line algorithms: MEAN and MODE which always predict the mean and mode value of the classes observed in the data set, respectively. Predicting ability was measured in terms of rate of erroneous classifications, linear loss (average of the distance between the predicted and the observed classes), the number of predicted classes and the F_1 statistic (which allows comparing procedures taking into account that they can predict different number of classes). The seminal traits with a bigger influence on fertility were established using stepwise regression and a nondeterministic classifier. MEAN, LR and SVR produced a higher percentage of wrong classified cases than MODE (taken as reference for this statistic), whereas it was 6 %, 13 % and 39 % smaller for SVOR, OLR and NDOR, respectively. However, NDOR predicted an average of 2.04 classes instead of 1 class predicted by the other procedures. All the procedures except MODE showed a similar smaller linear loss than the reference one (MEAN) being SVOR the one with the best performance. The NDOR showed the highest value of the F_1 statistic. Values of linear loss and F_1 statistics were far from their best value indicating that possibly, the variation in fertility explained by this group of semen characteristics is very low. From the total amount of traits included in the full model, 11, 16, 15, 18 and 3 features were kept after performing variable selection with the LR, OLR, SVR, SVOR and NDOR methods, respectively. For all methods, the reduced models showed almost an irrelevant decrease in their predictive abilities compared to the corresponding values obtained with the full models.

Key words: fertility, non parametric methods, prediction, rabbit, seminal traits

INTRODUCTION

Artificial insemination (AI) in rabbit commercial farms is performed with pooled semen from several bucks at a high sperm dosage in order to overcome the negative effects on fertility of semen with suboptimal characteristics. This practice reduces the output of AI centers and impedes making right decisions regarding male replacement and management in AI centers. Obtaining an accurate prediction of the fertilizing potential of ejaculates would alleviate those limitations increasing the economical benefits of AI centers.

However, the relationship between the seminal traits and the result of insemination is still not clearly established. Most of the studies have shown that the proportion of the observed variance that is explained by models including the set of traits usually recorded in the AI centers is very low (Brun et al., 2002, Gadea et al., 2004, García-Tomás et al., 2006a). This could be due to: i) The experimental design. Thus, when AI is performed with semen obtained after a strong pre-selection of the ejaculates, the observed variability is reduced. ii) The variables used as fertility markers, the way how they are measured and the time when they are recorded with respect to AI time could not be adequate. iii) The methods used for variable selection and prediction could be too rigid for modeling some kind of relationships. iv) The use of variables with not relevant or redundant information may mislead the classifiers, leading to dismiss their performance. Finally, v) It could be possible that, actually, the part of the observed variance of this trait (i.e. fertility at kindling) due to the variation of the characteristics of the ejaculates accepted for AI is very low, being much more important features of the doe and environmental factors. In this case the search of a method, based on features of

74 the ejaculate, to explain a large part of the variation of the AI, would be necessarily
75 unsuccessful.
76
77 Objectives of this work were to answer the following: 1) Is it possible to improve
78 the accuracy of fertility prediction by using more flexible procedures?; 2) How much the
79 information provided by seminal variables can improve fertility prediction?; 3) Among
80 them, which are the ones with highest influence on male fertility?
81
82 **MATERIAL AND METHODS**
83
84 **Animals and data**
85
86 The research protocol was approved by the animal care and use committee of the
87 Institut de Recerca i Tecnologia Agroalimentàries (IRTA).
88
89 *Animals*
90 Males belonged to the Caldes line selected for growth rate during the fattening
91 period (Caldes line: Gómez et al., 2002a). Bucks were bred and reared in an experimental
92 farm in Caldes de Montbui (Barcelona, Spain). This farm has insulated walls and roof
93 and the proper cooling equipment to avoid animal exposure to extreme temperatures.
94 After weaning at 32 d, males were housed in collective cages of 8 individuals with a
95 photoperiod of 16 h light/day. Animals were fed a commercial diet of rabbit pellets *ad*
96 *libitum* (15.5% crude protein, 2.3% fat, 17.2% fiber) until 60 d. Subsequently, they were
97 housed on the farm of the AI centre under the same environmental conditions as the
98 experimental farm and placed beside it, and they were restricted to 180 g/d of another

commercial diet (16% crude protein, 4.3% fat, 17% fiber). Fresh water was always available.

Semen collection

All males began training to use an artificial vagina at 4.5 mo of age. A homemade polyvinyl chloride artificial vagina containing water at a temperature of 50 ºC was used. One ejaculate was collected per male each week for the first two weeks. After this period, 2 ejaculates per male were collected each week, with an interval of 30 min between collections. From 5 to 9 mo of age, all males were evaluated at three different times for seminal quality traits and their corresponding fertility score after AI over crossbred females in a commercial farm.

Evaluation of the seminal traits and AI

Ejaculates were stored in a dry bath at 35ºC until evaluation for no more than 15 min after collection. Ejaculates containing urine and calcium carbonate deposits were discarded, and gel plugs were removed. The ejaculate volume was assessed with a micropipette and the pH of the semen was determined using a 507 Crison pH-meter (Crison Instruments, SA, Alella, Barcelona, Spain). Aliquots (25µl) of ejaculate were diluted 1:4 (vol/ vol) in a commercial extender (Galap, IMV Technologies, Saint Ouen sur Iton, France) to assess the individual motility under a microscope with a phase-contrast optic (Nikon, Lewisville, TX) at 400X magnification, according to a subjective scale from 0 to 5 corresponding to a percentage of sperm showing progressive movement of: 0 to 10, 11 to 25, 26 to 50, 51 to 70, 71 to 90, or 91 to 100%, respectively (Roca et al., 2000).

To prepare the AI doses, a small pre-selection of ejaculates was performed, discarding for AI only those with individual motility lower than 2 and a percentage of dead spermatozoa higher than 50%. Semen suitable for AI was immediately prediluted 1:1 (vol/ vol) with a commercial semisolid extender (Cunigel, IMV Technologies, Saint Ouen sur Iton, France). After evaluation, the ejaculates obtained per male each day were pooled and cell sperm concentration (**Conc**; millions of spermatozoa per mL) was measured by using a sperm cell counter (NucleoCounter SP-100, Chemometec A/S, Allerød, Denmark). The resultant pool of ejaculates was divided into two parts which were diluted until 10 x 10^6 spermatozoa/ mL and 40 x 10^6 spermatozoa/ mL, respectively, to obtain AI doses at two different sperm concentrations (**DC**). The dilution rate (**Dilu**) was also recorded. Semen doses were stored in straws of 0.5 mL at 18ºC for 24 h until their use.

After 24 h, an AI dose (at each dose concentration) of each male dose was processed to artificially induce the acrosome reaction. The AI dose was tempered at 37 ºC for 30 min to allow the liquefaction of the semisolid extender. After tempering, samples were centrifuged and supernatants aspirated. The pellets were then resuspended to 200 µL with Hepes–Tyrode's Lactate (Hepes-TL). An aliquot of 50 µL was incubated at 37.5 ºC in 5%CO_2 in air for 3 h.

To induce acrosome reaction, Calcium Ionophore (A23187, Sigma Chemical., St Louis, MO, USA) was used at a final concentration of 10 µM, according to Januskauskas et al. (2000). Sperm samples with the ionophore were incubated 30-45 min at 37ºC. After incubation, samples were centrifuged and the pellets washed and resuspended to 60 µL with TALP medium.

149

150 The evaluation of sperm quality traits, sperm storage capacity and the ability of
151 the sperm to undergo the acrosome reaction, was performed from aliquots (10μL) of
152 semen samples collected at three different stages. In each of them, the aliquots were
153 placed and fixed on slides using a vital nigrosin-eosin staining (Bamba, 1988). Under a
154 light microscope (Nikon, Lewisville, TX) at 400X magnification, 200 spermatozoa were
155 evaluated from each slide to determine the following sets of sperm quality traits: 1) Sperm
156 characteristics obtained from aliquots of semen samples collected immediately after
157 pooling the ejaculate (0 h): percentages of viable spermatozoa (**VI_0**), spermatozoa with
158 normal apical ridge (**NAR_0**), morphological abnormalities of head (**HAP**), neck-midpiece
159 (**NAP**) and tail (**TAP**) and spermatozoa with presence of cytoplasmic droplet (**TD**), and
160 proximal (**PD**) and distal (**DD**) cytoplasmic droplet. 2) Sperm characteristics obtained
161 after 24 h of the storage period of the AI doses (around the insemination time):
162 percentages of viable spermatozoa (**VI_{24}**), spermatozoa with normal apical ridge
163 (**NAR_{24}**). 3) Sperm characteristics obtained after the induction of the acrosome reaction:
164 percentage of acrosome reacted spermatozoa (**AR**).
165
166 The stability of the sperm during the storage period was evaluated as difference
167 in sperm viability (**DVI**) and in percentage of spermatozoa with normal apical ridge
168 (**DNAR**) between 0 h and 24 h after ejaculate extraction. These parameters were
169 computed as DVI = VI_0 - VI_{24} and DNAR = NAR_0 - NAR_{24}, respectively. The ability of
170 the sperm sample to undergo acrosome reaction was evaluated as the percentage of
171 reacted acrosome spermatozoa after artificial induction of the acrosome reaction (**DAR**).
172 This parameter was obtained as DAR = AR - (100 - NAR_{24}). The percentage of total

reacted spermatozoa from 0h to the end of the process of acrosome reaction induction (**DAR_0**) was obtained as: $DAR_0 = AR - (100 - NAR_0)$

The total volume of the ejaculate per male and day of collection (**Vol**; mL) was calculated as the sum of the volumes of the 1st and 2nd suitable for AI ejaculates of each male and day. The total individual motility of the sperm per male and day of collection (**IM**), was calculated as $IM = \left(IM_1 \times Vol_1 + IM_2 \times Vol_2\right) \times \left(Vol_1 + Vol_2\right)^{-1}$, where IM_1 and Vol_1 and IM_2 and Vol_2 are the individual sperm motility and volume measures for the 1st and the 2nd ejaculate of the pool of each male, if present, respectively. The pH was also measured separately in each ejaculate before pooling. In cases where there were two ejaculates suitable for AI per male and per day, the resultant pH of the pool (**pH**) was calculated as follows:

$pH = -\log_{10}\left[\left(10^{-pH_1} \times Vol_1 + 10^{-pH_2} \times Vol_2\right) \times \left(Vol_1 + Vol_2\right)^{-1}\right]$. Where pH_1 and pH_2 are the pH measures for the 1st and 2nd ejaculate of the pool of each male, if present, respectively.

The AI doses were applied in a commercial farm (Montmajor, Barcelona, Spain) over crossbred does (P x V) (V line: Estany et al., 1989; Prat line; Gómez et al., 2002b). Females followed a semi-intensive reproductive rhythm: first mating at about 4.5 mo old, with subsequent 42 d reproductive cycles. All females were treated 48 h before AI with 15 IU eCG (subcutaneously; Foligon, Intervet International B.V., Booxmeer, Holland) and ovulation was induced immediately after AI with 0.02 mg of Gonadorelin (im; Fertagyl, Intervet internacional B.V. Booxmeer, Holland).

Diagnosis of pregnancy was made by palpation, 14 d after AI, and the result was confirmed at parity. A total of 6,613 AI results were obtained. Fertility (**Fert**) was defined as percentage of kindling rate by male, dose sperm concentration and AI day. In order to have a representative value of male fertility, Fert records computed with less than 4 inseminations per day and dose concentration were discarded from the analysis (representing the 11.4 % of the whole Fert data). These discarded Fert values were originated from two sources: i) from AI which results were not recovered in the farm (e.g. death or culled females, lost data records sheets, etc), and ii) from AI with doses from males that did not produce enough amount of total sperm in the day. Removing the second group of Fert data in the analyses, could lead to a possible bias in the estimation of the male fertility prediction and in the seminal traits used as explanatory variables of this prediction, especially in Vol and Conc (as the Vol x Conc leads the total amount of sperm produced per male each day). In order to check that, the ratio between the mean, median and 1^{st} and 2^{nd} quartiles of each seminal trait obtained from the whole data set and from the data set after removing the Fert values without less than 4 IA was calculated. Ratios were all close to 1 for all seminal parameters (ranging from 0.98 to 1.05). Therefore, it was confirmed that no bias in the parameters existed after performing this data edition. A total of 752 records of Fert were obtained from 193 males.

Table 1 shows a brief description of the seminal variables used in the analyses as predictors of fertility and Figure 1 shows their corresponding box plots.

Statistical analyses

Taking into account the continuous nature of fertility, defined as percentage of kindling rate, *regression* is the most straightforward approach susceptible to be applied when a fertility prediction model needs to be defined. Classical regression methods require the assumption of a specific parametric function (e.g., linear, quadratic, etc.) to model the data, which could be too rigid for modeling some kind of relationships. An alternative approach for the analysis of this kind of traits could be the use of non parametric methods (Wasserman, 2006), such as machine learning algorithms, since they do not require prior knowledge of a parametric function and can accommodate complex relationships between dependent and independent variables and intricate dependencies among explanatory variables. Besides, they are very flexible and can learn arbitrarily complex patterns when enough data are available.

Moreover, given that the objective is to get a classification of the ejaculates rather than an exact value of its potential fertility rate, models for classification (i.e. interval prediction, which is more reliable because the targets are broader) could be used. For this purpose, fertility rates are grouped into intervals. Furthermore, it is also possible to consider the ordinal nature of the intervals converting the learning process in an *ordinal regression* task. The power of these classifiers can be additionally improved using the so-called *nondeterministic classifiers* (Alonso et al., 2008 and del Coz et al., 2009), whose aim is to predict a set of classes (consecutive in case of ordinal regression) as small as possible, but (presumably) still containing the true class.

Loss functions

Classification procedures were compared in terms of 3 loss functions:

1) *Linear loss* or *absolute deviation*, which computes the absolute value of the difference between the observed and the predicted data.
2) *Error rate*, which computes the number or rate of erroneous classifications.
3) The complementary of the F_β statistic, which measures the goodness of a nondeterministic classification and can be defined as:

$$F_\beta\left(h(x),y\right)=\frac{(1+\beta)\cdot P\left(h(x),y\right)\cdot R\left(h(x),y\right)}{\beta\cdot P\left(h(x),y\right)+R\left(h(x),y\right)}=\frac{1+\beta}{\beta+\left|h(x)\right|}\cdot 1_{y\in h(x)}$$

Where, y is the true value, $h(y)$ is the prediction, $|h(x)|$ stands for the number of classes included in the prediction for an entry x, P (Precision) is the proportion of predicted classes that are relevant (which it is only one) and R (Recall) is an indicator of whether the real class is in the set of the classes predicted. Thus, F is the harmonic average of P and R, where β indicates the number of times that R is considered much important than P. A typical choice of this parameter is $\beta=1$, which means assigning equal importance to R and P ($\mathbf{F_1}$). For a test set F_1 is computed as the average of this parameter for each individual entry. F_β is an informative measure in order to compare different classifiers, but sometimes it is not easy to infer the quality of the solution. On the contrary, P and R are able to provide a better understand of the behavior of the classifier. The latter shows us the accuracy of the prediction, and, the former indicates the number of ranks in the prediction (in fact, it involves the inverse of such number).

Note that in deterministic classification the number of classes included in the prediction is always 1, and then, $F_1 = P = R =$ (*1 – error rate*).

Data in training and validation sets

The initial data set (*S*) contains $n = 752$ semen samples described by the 21 traits shown in Table 1 and the class to be learned (*Y*) is Fert (the percentage of kindling rate). Despite AI was performed with homospermic doses, the buck was deliberately omitted in the set of variables used for prediction because the objective of this research was to assess the ability of the characteristics usually measured in an ejaculate (pooled or not from several males) to predict fertility after AI. In other words, the objective was to assess the value of those measurements by themselves as fertility markers, irrespectively of the buck.

This data set was used for *regression tasks*. However for *ordinal regression* task or *nondeterministic ordinal regression task*, it was rewritten discretizing those ordered values in a set of 5 qualitative ranks (classes) of equal length for Fert: very low, low, medium, high, very high (refer to Table 2 for the intervals of fertility rate used in this study for semen classification).

Learning algorithms

To analyze these data, several methods were employed:

1) Linear Regression (**LR**)
2) Ordinal Logistic Regression (OLR)
3) Support Vector Regression (**SVR;** Vapnik, 1995). This algorithm uses the ε-insensitive loss function that ignores errors smaller than a certain threshold $\varepsilon > 0$. Another characteristic of the support vector algorithms is the existence of a parameter (*C*) that is a trade-off between the flatness of the learned function and the amount up to which deviations larger than ε are tolerated (Smola and Scholkopf, 2004). Smola and Scholkopf also show that SVR works well in environments with noise and outliers, as it is usually the case of seminal traits.

4) Support vector algorithm for ordinal regression tasks (**SVOR;** Chu and Keerthi, 2005) using the classes shown in Table 2.
5) A nondeterministic ordinal regression algorithm (**NDOR**) proposed by Alonso et al. (2008). This algorithm is able to control the number of classes to predict and the error rate by means of a trade-off parameter (β). When the number of classes in the prediction is more than one, then the classes must be consecutive.

Two base line algorithms are employed in order to test the performance of the more sophisticated ones.

6) The **MEAN**, is used to compare with regression algorithms. MEAN always returns the mean value of the classes observed in the examples of the data set. The translation from percentages of kindling rate to classes is shown in Table 2.
7) The **MODE**, can be employed to compare with ordinal regression algorithms. MODE always returns the mode value of the classes observed. This method can be also adapted to work as a nondeterministic algorithm. The idea is to force it to make predictions with a fixed number of the most frequently and consecutive classes. Thus, MODE_i will predict the i most frequent and consecutive classes.

Variable selection

Finally, the seminal traits with a bigger influence on fertility were established using a traditional *feature selection* method (stepwise regression) and machine learning oriented one. There are no specific feature (i.e. variable) selection methods for nondeterministic algorithms; however, multi-class classification feature selection algorithms can be adapted for this task. For this purpose, variables were ranked in order

to select the best variables subset. The score proposed for obtaining the ranking is the pairwise ranking error (PRE), an extension of the area under the ROC curve (AUC) able to cope with more than two ordered classes (Hanley and McNeil, 1982). Then, it is used a method based in the Recursive Feature Elimination (RFE) algorithm proposed by Guyon et al. (2012) that produces an attribute ranking. To select the best feature subset a wrapper is applied in conjunction with OLR, SVR, SVOR and NDOR.

Experimental setup

A 5-fold cross-validation was repeated twice. The same Support Vector Machine (SVM) (Vapnik, 1995) implementation was used to learn a regressor (SVR) and to obtain the posterior probabilities needed in the nondeterministic classifier (NDOR): LibSVM (Chang and Lin, 2001) with linear kernel. The ordinal regression classifier (SVOR) is that described in (Chu and Keerthi, 2005). To adjust the *C* parameter for these algorithms we performed an internal grid search (a 2-fold cross-validation repeated 5 times) with $C = 10^k$ and $k \in [-5,2]$.

RESULTS AND DISCUSSION

Figure 1 shows the Boxplots for the seminal characteristics used to predict Fert. All of them showed values close to the ones obtained in other studies in the same paternal line of rabbits (García-Tomás et al., 2006b; García-Tomás et al., 2008). The variables describing different morphological abnormalities had small values of the median (especially for HAP and DD) and they showed an asymmetric distribution of the data. Classical linear regression does not seem to be the most adequate procedure for analyses

with this type of variables and with complex relations among them because the distribution of the data is not known beforehand and the assumption of any distribution may lead to misclassify the data.

The predictive performance of all the procedures is shown in Table 3. Regarding the error rate, the procedure that considers the mode of the data distribution in the training set as the prediction for all the data in the validation set (MODE) can be considered the simplest method that could be used for classification. Thus, the result obtained with MODE was taken as a reference or maximum value to not be overcome for some other procedure in order to improve the quality of the classification. Results indicate that MEAN, LR and SVR produced a bigger percentage of wrong classified cases than MODE, whereas this percentage was 6%, 13%, and 39% smaller for SVOR, OLR and NDOR, respectively.

However, NDOR predicted an average of 2.04 classes instead of 1 class predicted by the other procedures. The average of the distance between the predicted and the observed classes (linear loss) was 1.36 for the procedure which considered the mean of the data in the training set as the predicted value for all the data in the validation set. The result obtained with this procedure was considered the maximum value that other procedures should not overcome. All the procedures except MODE showed a smaller linear loss than the reference one. Although differences between procedures for this statistic were small, SVOR showed the best performance. In the case of NDOR, this statistic was obtained as the average of the difference between the observed class and the average of the predicted classes.

The F_1 statistic allows comparing procedures taking into account that they can predict different number of classes. The maximum value for this statistic is 1.0, which corresponds to the situation where all cases are correctly predicted with just one class. The procedure with the highest value of this statistic was NDOR. However, as it occurred with the error rate, this value was far from its best value indicating that, probably, the percentage of variation in fertility explained by this group of semen characteristics is very low (Brun et al., 2002; Gadea et al., 2004; García-Tomás et al., 2006a) and it may be necessary to find other semen quality markers, or to evaluate some of the currently used ones in a more precise manner or closer to the AI time.

Predictive ability obtained with the MODE and NDOR are shown in Table 4. The NDOR was able to correctly classify 2 out of 3 semen samples predicting an average of 2 consecutive classes. The error rate of this procedure was compared with the reference procedure (MODE) modified to perform as a nondeterministic procedure (Table 4). The modified MODE would require predicting 4 out of the 5 possible classes in order to correctly classify more semen samples than the NDOR procedure. Regarding the error rate, MODE_4 had a better value than NDOR but it predicted 4 classes instead of 2.04. Regarding the other parameters the MODE has always a worse performance than NDOR, independently of the number of classes that it predicted. NDOR predicted just one class in the 35% of the semen samples, 2 classes in 47% of the semen samples, 3 classes in 8% of the samples, never predicted 4 classes, and in 10% of the samples it predicted all the classes (given that there was no information supporting any possible classification).

As a general comment of the previous results, non-parametric methods for predicting the rank of the ejaculates according to their potential fertility rate from seminal characteristics, seems to improve the quality of the prediction with respect to the obtained using the classical regression procedure. However the improvement is not high enough

394 to make decisions concerning the bucks or the ejaculates. One of the main problems could
395 be that the trait that we are trying to predict (fertility at kindling) is due to the male and
396 to the female in a different rate. Fertility at kindling is greatly conditioned by prenatal
397 survival, which is uniquely determined by the doe and other environmental factors. Then
398 the effects of the male are masked, and it is very difficult to establish a relationship
399 between seminal characteristics and this trait. On the other hand, fertility rate, as it is
400 defined here, is calculated as the rate of positive matings which does not allow
401 differentiate for each insemination between ejaculates that fertilize most of the oocytes
402 and those that fertilize only a part of them.

403 Based on all these considerations, Piles et al. (2012) propose using embryonic
404 survival and number of implanted embryos instead of fertility at kindling, in order to
405 improve the quality of the evaluation of the ejaculates by their characteristics involved in
406 fertilization and the subsequent embryogenesis processes, which are the reproductive
407 processes which probably have an important male contribution. This could be important
408 when the objective is to improve the quality of the doses produced in the AI centers or to
409 make decisions regarding buck replacement.

410 Table 5 shows which of the 22 features (21 seminal traits and Age) were kept in
411 each one of the methods after performing the feature selection. Except the NDOR, the
412 number of variables kept with the other methods after performing variable selection was
413 high. From the total amount of features included in the full model, 11, 16, 15, 18 and 3
414 features were kept after performing variable selection with the LR, OLR, SVR, SVOR
415 and NDOR methods, respectively.

416

417 The predictive ability of the resultant reduced models with the selected features is
418 also shown in Table 5. For all methods, the reduced models showed almost an irrelevant

decrease in their predictive abilities -in terms of error rate, linear loss and F_1 statistic- with respect to their corresponding values obtained with the full model (Table 3). It is highly surprising that NDOR almost did not get worse predictive ability in the reduced model compared to the full one (only a slight increase in the linear loss was observed) by using only 3 of the whole 22 features. From a practical point of view, this is a very interesting result because it implies that fertility could be equally predicted using a very small number of seminal variables without diminishing the predictive ability of the method.

Only pH and TD were always selected by the five procedures. The Age, DC, Vol, IM, NAR_0, AR and DAR were selected in four of the five models presented. Average and standard error of Age and the most relevant seminal traits for each fertility class are shown in Figure 2.

The semen pH is related to the concentration and motility of spermatozoids in the ejaculate (Hulet and Ercanbrack, 1962; Coffey, 1988; Bencheikh, 1995; Brun et al., 2002; Garcia-Tomás et al., 2006b). In rabbits, several studies have found negative correlations between pH of ejaculate and fertility (Coffey, 1988; Brun et al., 2002; Tusell et al., 2011) or litter size (More O'Ferrall and Meacham, 1968) in accordance with our results (Figure 2, panel D).

Migration of the cytoplasmatic droplet occurs in the epididymus (Pérez-Sánchez et al., 1997), but cytoplasmic droplets can be present in the ejaculated spermatozoa (Cooper and Yeung, 2003). Our results (Figure 2, panel D) are in accordance with the ones obtained with boars where the high presence of ejaculated spermatozoids with distal

droplets led to obtain reduced fertility and litter size (Waberski et al., 1994). Relationship between infertility and droplet retention has also been denoted in mice (Yeung et al., 2000) and human (reviewed by: Cooper, 2005).

Rabbit bucks reach sexual maturity at approximately 5 months and semen quality generally decreases in older rabbit bucks (> 2 years; Boiti et al., 2005). García-Tomás et al. (2009) found that, between 4.6 and 7.5 mo of age, males from the Caldes line still had an important increase in ejaculate volume and individual motility of the spermatozoa suggesting that they could have not totally reached sexual maturity according to their testis size and the percentage of seminiferous tubules with presence of lumen during that period. However, according to the fertility results of the current study, it seems that fertility of males decreases with Age because the average male age of the ejaculates with the lowest fertility is higher than the average male age of those with the highest fertility (Figure 2, Panel A). Further research is needed in order to clarify the effect of male age on fertility.

Several works in rabbits have denoted the effect of sperm concentration of the AI doses on fertility (Farrell et al., 1993; Alvariño et al., 1996; Viudes-de-Castro and Vicente, 1997). Although it is not very clear, it seems that ejaculates with higher Vol have lower fertility than the ones with lower Vol (Figure 2, Panel B). As semen is composed by seminal plasma and spermatozoa, Vol is determined by the amount of these two components. Ejaculates are diluted to prepare the AI dose considering only the spermatozoa concentration of the ejaculate and the desired amount of sperm in the AI dose. Among all the methods, feature Dilu was kept only in the stepwise LR whereas Vol and Conc were kept in two and three of the methods, respectively (Table 5). Killian et al.

(1993) suggested that the effect of dilution on the potential fertility of the doses is male specific because of individual variation in the composition of plasma and sperm quality.

Even though that a small preselection was performed discarding for AI those ejaculates with the worst scores of IM, it was possible to observe the positive effect of high IM scores on fertility (Figure 2, Panel C). Spermatozoa require good motility in order to cross the female genital tract, reach the oocytes and perform a successful fertilization. Moreover, sperm motility is a good indicator of the status and functionality of the sperm membrane (Gadea, 2005). Several authors have denoted that sperm motility is a good indicator of poor fertility; however, high values do not guarantee good fertility (Flowers, 1997; Braundmeier and Miller, 2001). Brun et al. (2002) found that mass motility score was the most influential trait on kindling rate among several quantitative and qualitative seminal traits analyzed whereas Garcia-Tomás et al. (2006a) found no clear relationship between fertility and individual sperm motility evaluated according to a subjective scale. Both studies rejected higher amount of ejaculates than in the current study using, among other variables, sperm motility scores. This could contribute to diminish the amount of variation for this trait and possibly to reduce its correlation with fertility.

NAR_0, AR and DAR features are related to the acrosomal status of the sperm (Figure 2, Panel E, G and H, respectively). Regarding NAR_0, which reflects the proportion of spermatozoa with a normal apical ridge in an untreated semen sample, is indicative of its fertilizing ability because acrosome reacted spermatozoa or abnormal acrosome spermatozoa have a short longevity and are not able to fertilize (Saake and White, 1972). Our results indicate the positive relationship between NAR_0 and fertility

(Figure 2, Panel E). Also in rabbits, a negative and moderate correlation was found between fertility and percentage of abnormal acrosomes (Courtens et al., 1994). However, the percentage of sperm with acrosomal integrity was found to be non significant when it was included in a multiple regression analysis of fertility in two paternal lines of rabbits (being one of them the Caldes line; Garcia-Tomás et al., 2006a). In that study, NAP was the feature with the most relevant effect on fertility. Conversely to NAR_0, that only evaluates the morphological acrosome status of the sperm at the time of collection, other laboratorial tests could better assess the functionality of the sperm acrosome. The NAR_0 and DAR features seem to be more informative with regards to the fertilizing capacity of fresh semen than the morphological evaluation of the sperm acrosome status, but its relation with fertility is not clear (Colenbrander et al., 2003) although there is some evidence of it in bovine (Whitfield and Parkinson, 1995). The DAR gives the proportion of spermatozoa that, after artificial induction, have undergone acrosome reaction (because the final figure of reacted spermatozoa is corrected by the initial amount of spermatozoa already reacted before the acrosome reaction induction) whereas AR only refers to the final amount of reacted acrosome spermatozoa present in the semen sample once the acrosome reaction has been artificially induced. As expected, NAR_0 and DAR are two seminal traits highly correlated (0.73). In addition, NAR_0 and AR (Figure 2, panel E and G, respectively) showed a very similar profile indicating that samples with higher percentage of normal acrosome at collection are more susceptible to satisfactory develop acrosome reaction after artificial induction.

IMPLICATIONS

519 Non-parametric methods for prediction such as Support Vector Ordinal
520 Regression and Non-Deterministic Ordinal Regression seem to improve the success in
521 the classification of the ejaculates according to their potential fertility rank predicted
522 from characteristics of the artificial insemination doses, with respect to the obtained
523 using the classical regression procedure. Moreover, Non-Deterministic Ordinal
524 Regression could use the relevant information in a more efficient manner, because it is
525 able to get the same predictive ability than other methods using a very small number of
526 seminal traits. However, the predictive ability of male fertility from seminal
527 characteristics used in this research (many of which are commonly used in the AI
528 centers for ejaculate selection) is not very different of that obtained with no seminal
529 information.
530
531 The pH, the rate of spermatozoa with presence of cytoplasmic droplet, and the
532 rate of reacted spermatozoa during the process of acrosome reaction induction could be
533 good markers for field AI buck fertility.

ACKNOWLEDGMENT

This research was supported by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA, Madrid, Spain) project RTA2005-00088-CO2. The authors thank Cristina Andreu, Irina Garcia, Mireia Campos (Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain), and Anna Pérez, Carme Requena, Oscar Perucho (Unitat de Cunicultura, IRTA, Caldes de Montbui, Barcelona, Spain) for their valuable cooperation during the experimental phase of this study, Professor M. Baselga and R. Lavara for their comments and suggestions and, Dr. F. Goyache, who made this collaboration possible.

REFERENCES

Alvariño, J. M. R., López, F. J., Del Arco, J. A., Bueno, R. Torres, R. 1996. Effects of semen concentration on rabbit artificial insemination with fresh or 24 hours stored semen. Proc. 6th World Rabbit Congr., Toulouse, France 2: 33-35.

547 Alonso, J., del Coz, J. J., Díez, J., Luaces, O., Bahamonde, A. 2008. Learning to predict one or more ranks in ordinal regression tasks. Proc
548 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), LNAI
549 5211, 39–54. Springer.
550 Bamba, K. 1988. Evaluation of acrosome integrity of boar spermatozoa by bright field microscopy using an eosin-nigrosin stain. Theriogenology
551 29: 1245-1251
552 Bencheikh, N. 1995. The effect of frequency of ejaculation on semen characteristics and sperm output in the rabbit. Ann. Zootech. 44: 263-279.
553 Boiti, C., Castellini, C., Theau-clement, M., Besenfelder, U., Liguori, L., Renieri, T., Pizzi, F. 2005. Guidelines for the handling of rabbit bucks
554 and semen. World Rabbit Sci. 13: 71-91.
555 Braundmeier, A. G., Miller, D. J. 2001. Invited Review: The Search is on: Finding Accurate Molecular Markers of Male Fertility. J. Dairy Sci. 84,
556 1915-1925.
557 Brun, J. M., Theau-Clement, M., and Bolet, G. 2002. The relationship between rabbit semen characteristics and reproductive performance after
558 artificial insemination. Anim. Reprod. Sci. 70: 139-149.
559 Chang, C. C., Lin, C. J. 2001. LIBSVM: A library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm.
560 Chu, W., Keerthi, S. S. 2005. New approaches to support vector ordinal regression. Proceedings of the 22nd International Conference on Machine
561 learning. P 145—152.

562 Coffey, D. S. 1988. Androgen action and the sex accessory tissues. In: Physiology of reproduction, Knobill. E., and J. D. Neill, (eds.). Raven Press,
563 New York, pp. 1081-119.
564 Colenbrander, B., Gadella, B. M., Stout, T. A. E. 2003. The predictive value of semen analysis in the evaluation of stallion fertility. Reprod.
565 Domest. Anim. 38: 305-311.
566 Cooper, T. G. 2005. Cytoplasmic droplets: The good, the bad or just confusing? Hum. Reprod. 20: 9-11.
567 Cooper, T. G., Yeung, C. H. 2003. Acquisition of volume regulatory response of sperm upon maturation in the epididymis and the role of the
568 cytoplasmic droplet No. 61. p 28-38. Wiley Subscription Services, Inc., A Wiley Company.
569 Courtens, J. L., Bolet, G., Theau-Clement, M. 1994. Effect of acrosome defects and sperm chromatin decondensation on fertility and litter size in
570 the rabbit - preliminary electron-microscopic study. Reprod. Nutr. Dev. 34: 427-437.
571 del Coz, J.J., Díez, J., Bahamonde, A. 2009. Learning Nondeterministic Classifiers. J. Mach. Learn. Res., 10: 2273-2293
572 Estany J., Baselga, M., Blasco, A., Camacho, J. 1989. Mixed model methodology for the estimation of genetic response to selection in litter size
573 of rabbits. Livest. Prod. Sci. 21: 67-75.
574 Fan, R., Chang, K., Hsieh, C., Wang, X., Lin, C. 2008. LIBLINEAR: A library for large linear classification, J. Mach. Learn. Res. 9: 1871–1874.
575 Farrell, B. P., Foote, R. H., Simkin, M. E., Clegg, E. D., Wall, R. J. 1993. Relationship of semen quality, number of sperm inseminated, and fertility
576 in rabbits. J. Androl. 14: 464-471.

577 Flowers, W. L. 1997. Management of boars for efficient semen production. J. Reprod. Fertil. 52: 67-78.
578 Gadea, J. 2005. Sperm factors related to in vitro and in vivo porcine fertility. Theriogenology 63: 431-444.
579 Gadea, J., Sellés, E., Marco, M. A. 2004. The predictive value of porcine seminal parameters on fertility outcome under commercial conditions.
580 Reprod Domest Anim. 39:303-308.
581 Garcia-Tomás, M., Sanchez, J., Rafel, O., Ramon, J., Piles, M. 2006a. Reproductive performance of crossbred and purebred male rabbits. Livest.
582 Sci. 104: 233-243.
583 Garcia-Tomás, M., Sanchez, J., Rafel, O., Ramon, J., Piles, M. 2006b. Variability, repeatability and phenotypic relationships of several
584 characteristics of production and semen quality in rabbit. Anim. Reprod. Sci. 93, 88-100.
585 Garcia-Tomas, M., Tusell, L., López-Béjar, M., Ramon, J., Rafel, O., Piles, M. 2008. Influence of environmental temperature and relative humidity
586 on quantitative and qualitative semen traits in rabbits. Proc. 9th World Rabbit Congr. (Verona) 72, 111.
587 García-Tomás, M., Sánchez, J., Piles, M. 2009. Post-natal sexual development of testis and epididymis in the rabbit: Variability and relationships
588 among macroscopic and microscopic markers. Anim. Reprod. Sci. 110: 347-355.
589 Gómez, E. A., Rafel, O., Ramon, J. 2002a. The Caldes strain. Rabbit genetic resources in mediterranean countries Options méditerranéennes, serie
590 B: Etudes et recherches. p 189-198.

591 Gómez, E. A., Rafel, O., Ramon, J. 2002b. The Prat strain. Rabbit genetic resources in mediterranean countries Options Méditerranéennes, serie
592 B: Etudes et Recherches. p 199-208.
593 Guyon, I., Weston, J., Barnhill, S., Vapnik, V. 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–
594 422.
595 Hanley J. A., McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143, 2936.
596 Hulet, C. V., Ercanbrack, S. K. 1962. A fertility index for rams. J. Anim Sci. 21: 489-493.
597 Januskauskas A, Johannisson, A., Söderquist, L., Rodriguez-Martinez, H. 2000. Assessment of sperm characteristics post-thaw and response to
598 calcium ionophore in relation to fertility in Swedish dairy AI bulls. Theriogenology. 53:859-75.
599 Killian, G. J., Chapman, D. A., Rogowski, L. A. 1993. Fertility-associated proteins in holstein bull seminal plasma. Biol. Reprod. 49: 1202-1207.
600 Lin, C. J., Weng, R. C., Keerthi, S. S. 2008. Trust region Newton method for logistic regression, J. Mach. Learn. Res. 9:627–650.
601 More O'Ferrall, G. J., Meacham, T. N. 1968. Relationship between ph , other semen traits and fertility in rabbits. 6th Cong. Intern. Reprod. Anim.
602 Insem. Artif. Paris 2: 1279-1281.
603 Pérez-Sánchez, F., Tablado, L., Soler, C. 1997. Sperm morphological abnormalities appearing in the male rabbit reproductive tract. Theriogenology
604 47: 893-901.

605 Roca, J., Martinez, S., Vazquez, J. M., Lucas, X., Parrilla, I. Martinez, E. A. 2000. Viability and fertility of rabbit spermatozoa diluted in tris-buffer
606 extenders and stored at 15 degrees C. Anim. Reprod. Sci. 64: 103-112.
607 Saacke RG, White, J. M. 1972. Semen quality tests and their relationship to fertility. Proc 4th Tech. Conf. NAAB; 22-27.
608 Smola, A., Scholkopf, B. 2004. A tutorial on Support Vector Machines. Statistics and computing, 14(3): 199-222. Springer.
609 Tusell, L., Legarra, A., Garcia-Tomas, M., Rafel, O., Ramon, J., Piles M. 2011. Different ways to model biological relationships between fertility
610 and the ph of the semen in rabbits J. Anim Sci. 89: 1294-1303.
611 Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer, NY.
612 Viudes-de-Castro, M. P., Vicente, J. S. 1997. Effect of sperm count on the fertility and prolificity rates of meat rabbits. Anim. Reprod. Sci. 46:
613 313-319.
614 Waberski, D., Meding, S., Dirksen, G., Weitze, K. F., Leiding, C., Hahn, R. 1994. Fertility of long-term-stored boar semen: Influence of extender
615 (androhep and kiev), storage time and plasma droplets in the semen. Anim. Reprod. Sci. 36: 145-151.
616 Wasserman, L.A. 2006. All of nonparametric statistics. New York: Springer. ISBN 10 0387251456
617 Whitfield, C. H., Parkinson, T. J. 1995. Assessment of the fertilizing potential of frozen bovine spermatozoa by in vitro induction of acrosome
618 reactions with calcium ionophore (a23187). Theriogenology 44: 413-422.

Yeung, C.-H., Wagenfeld, A., Nieschlag, E., Cooper, T. G. 2000. The cause of infertility of male c-ros tyrosine kinase receptor knockout mice. Biol. Reprod. 63: 612-618.

FIGURE CAPTIONS

Figure 1. Boxplots of the seminal traits used to predict male fertility data. Refer to Table 1 for the description of the seminal traits.

Figure 2. Average and standard error of age of the male at collection (Age), semen volume (Vol), individual motility of the sperm (IM, from 0 to 5: 0 to 10, < 10 to 25, < 25 to 50, < 50 to 70, < 70 to 90, or < 90 to 100, respectively, of the motile spermatozoa showing progressive movement), pH of the semen, percentage of spermatozoa with normal apical ridge (NAR_0), percentage of spermatozoa with presence of cytoplasmic droplet (TD), percentage of acrosome reacted spermatozoa after the induction of the acrosome reaction (AR) and percentage of reacted spermatozoa during the process of acrosome reaction induction (DAR) for each fertility class (from 1 to 5: 0 to 20, > 20 to 40, > 40 to 60, > 60 to 80 and > 80 to 100%, respectively).

Table 1. Description of the features used to predict male fertility

Variable	Description
Age	Age of the male at ejaculate collection (mo)
Conc	Sperm concentration of the ejaculate per male and day ($x10^6$spermatozoa/ mL)

Vol	Total volume of the ejaculate per male and day (mL)
IM	Individual sperm motility (subjective scale 1-5)
pH	-log hydrogen-ion concentration of the semen (units of pH)
VI_0	% of viable spermatozoa at 0 h (fresh semen)
NAR_0	% of spermatozoa with normal apical ridge
HAP	% of spermatozoa with morphological abnormality of head
NAP	% of spermatozoa with morphological abnormality of neck-midpiece
TAP	% of spermatozoa with morphological abnormality of tail
PD	% of spermatozoa with presence of proximal cytoplasmic droplet
DD	% of spermatozoa with presence of distal cytoplasmic droplet
TD	% of spermatozoa with presence of cytoplasmic droplet
VI_{24}	% of viable spermatozoa at 24 h (after the storage period of the doses)
NAR_{24}	% of spermatozoa with normal apical ridge at 24 h (after the storage period of the doses)
AR	% of acrosome reacted spermatozoa after the induction of the acrosome reaction
DVI	Differences in % in sperm viability between 0h and 24 h after the storage period (VI_0- VI_{24})
DNAR	Differences in % in sperm with normal apical ridge between 0h and 24 h after the storage period (NAR_0- NAR_{24})
DAR	% of reacted spermatozoa during the process of acrosome reaction induction **[**AR-(100-NAR_{24})]
DAR_0	% of total reacted spermatozoa from 0h to the end of the process of acrosome reaction induction [AR-(100-NAR_0)]

DC	Dose concentration (x10^6spermatozoa/ mL)
Dilu	Dilution rate

Table 2. Intervals of fertility rate used for semen classification

Class	Interval of fertility rate	Number of cases	Percentage of total
1. Very low	[0.0, 0.2)	200	26.60
2. Low	[0.2, 0.4)	68	9.04
3. Intermediate	[0.4, 0.6)	112	14.89
4. High	[0.6, 0.8)	191	25.40
5. Very high	[0.8,1.0)	180	24.07
	Total	752	100.00

641

642

643

644 **Table 3**. Predictive ability obtained with the different classification procedures. Results are the mean and the standard deviation obtained using a
645 5-fold cross-validation repeated 2 times. Best performances for each one of the statistics are marked in bold face. F_1 is the harmonic average of
646 Precision and Recall.

Method	**Error rate**	**Linear loss**	**F_1**	**N. of classes**
MEAN	0.85 ± 0.0012	1.36 ± 0.0014	0.15 ± 0.0012	1 ± 0
MODE	0.73 ± 0.0003	2.11 ± 0.0013	0.27 ± 0.0003	1 ± 0
LR[1]	0.76 ± 0.0109	1.09 ± 0.0241	0.25 ± 0.0109	1 ± 0
OLR[2]	0.60 ± 0.0099	1.15 ± 0.0309	0.40 ± 0.0099	1 ± 0
SVR[3]	0.74 ± 0.0106	1.09 ± 0.0202	0.26 ± 0.0106	1 ± 0
SVOR[4]	0.67 ± 0.0127	**1.05** ± 0.0328	0.33 ± 0.0127	1 ± 0

NDOR[5]	**0.34** ± 0.0237	1.14 ± 0.0318	**0.45** ± 0.0072	2.04 ± 0.0978

[1] LR = Linear regression

[2] OLR = Ordinal Logistic Regression

[3] SVR = Support Vector Regression

[4] SVOR = Support Vector Ordinal Regression

[5] NDOR = Nondeterministic Ordinal Regression.

Table 4. Predictive ability obtained with the Mode and nondeterministic ordinal regression (NDOR). Results are the mean and the standard deviation obtained using a 5-fold cross validation repeated 2 times. Best performances for each one of the statistics are marked in bold face. MODE X is the method MODE predicting X classes.

Method	**Error rate**	**Linear loss**	**F_1**	**N. of classes**

MODE 1	0.73 ± 0.0003	2.11 ± 0.0013	0.27 ± 0.0003	1 ± 0
MODE 2	0.51 ± 0.0005	1.63 ± 0.0010	0.33 ± 0.0004	2 ± 0
MODE 3	0.36 ± 0.0009	1.37 ± 0.0011	0.32 ± 0.0004	3 ± 0
MODE 4	**0.24** ± 0.0008	1.50 ± 0.0009	0.30 ± 0.0003	4 ± 0
MODE 5	0 ± 0	1.36 ± 0.0014	0.33 ± 0	5 ± 0
NDOR	0.34 ± 0.0237	**1.14** ± 0.0318	**0.45** ± 0.0072	2.04 ± 0.0978

Table 5. Features kept in each one of the procedures used to predict male fertility are marked with X. Linear regression (LR), Ordinal Logistic Regression (OLR), Support Vector Regression (SVR), Support Vector Ordinal Regression (SVOR), Nondeterministic Ordinal Regression (NDOR). Results are the mean and the standard deviation obtained using a 5-fold cross validation repeated 2 times

Method[1]	Age[1]	Conc[1]	Vol[1]	IM[1]	pH[1]	VI_0[1]	NAR_0[1]	HAP[1]	NAP[1]	TAP[1]	PD[1]	DD[1]	TD[1]	VI_{24}[1]	NAR_2[1]	AR[1]	DVI[1]	DNAR[1]	DAR[1]	DAR_0[1]	DC[1]	Dilu[1]	Predictive ability with relevant seminal traits
LR	X		X	X	X		X			X			X			X	X				X	X	Error rate: 0.76 ± 0.0118 Linear Loss: 1.09 ± 0.0216 F_1:0.24 ± 0.0118 N. classes: 1 ± 0
OLR																							Error rate: 0.59 ± 0.0133 Linear Loss: 1.11 ± 0.0355 F_1:0.41 ± 0.0133 N. classes: 1 ± 0
SVR	X	X	X	X	X		X				X		X	X	X	X		X	X	X	X		Error rate: 0.74 ± 0.0070 Linear Loss: 1.10 ± 0.0164 F_1:0.26 ± 0.0070 N. classes: 1 ± 0
SVOR	X	X	X	X	X		X	X	X		X		X	X	X	X	X	X	X	X	X		Error rate: 0.66 ± 0.0092 Linear Loss: 1.04 ± 0.0289 F_1:0.34 ± 0.0092 N. classes: 1 ± 0
NDOR					X								X						X				Error rate: 0.35 ± 0.0271 Linear Loss: 1.20 ± 0.0343 F_1:0.44 ± 0.0076 N. classes: 2.09 ± 0.1410

[1]Refer to Table 1 for the description of the features used in to predict male fertility.