Predicting sperm production of young dairy bulls using collection history and management factors

Allison Quick[1], James Meronek[2], Kyle Amburn[2], Kevin Rozeboom[2], and Kent Weigel[1]
[1] Department of Animal & Dairy Sciences, University of Wisconsin-Madison, Madison, WI.
[2] ABS Global Inc., DeForest, WI.

ANIMAL & DAIRY SCIENCES
University of Wisconsin-Madison

INTRODUCTION

- Previous research has established the significant role of management and environmental factors in sperm production of older, progeny-tested bulls (1,2,3)
- Previous research used linear models to identify key factors in total sperm (TS) production (1,2,3,4,5)
- To the author's knowledge, there is no published research specifically focused on TS forecasting of genomic young sires

OBJECTIVES

- Develop prediction models for daily, weekly, and monthly TS production from collection history, health status, and management factors and compare their performance
- Assess the ability of these prediction models to forecast future TS production, as well as differences in prediction accuracy by seasonality or age of bull

MATERIALS & METHODS

- Ejaculate data, bull demographics, and health records of Holstein and Jersey bulls from 2015 to 2019 were obtained from a commercial AI company.
- Data were cleaned and aggregated into daily, weekly, and monthly collection records of bulls 10-28 months of age

Table 1: Dataset summary of number of records and bulls in daily, weekly, and monthly TS production records.

	Datasets		
	Daily	Weekly	Monthly
Records (n)	43,918	23,404	5,127
Holstein	39,172	20,661	4,441
Jersey	4,746	2,743	686
Bulls (n)	1,037	1,003	664
Holstein	900	872	570
Jersey	137	131	94

- Potential explanatory variables included: age at collection, first successful collection, and arrival to AI stud, collection frequency, collection interval (days since previous collection), breed of bull, barn where housed and collected, year-season of collection, scrotal circumference (SC) at 10 to 11 months of age, health events, TS on the three most recent collection dates (lags)

Table 2: Descriptive statistics (minimum, mean, and maximum) of TS per week, number of ejaculates per week, TS per month, collections per month, arrival age (months), and collection age (months).

	Min	Mean	Max
TS per Week (billion cells)	0.1	15.1	82.2
Ejaculates per Week	1	3.5	10
TS per Month (billion cells)	0.1	49.9	252.5
Collections per Month	1	5.9	15
Arrival Age (mo)	1	6.2	15
Collection Age (mo)	10	17.7	28

- **Model Selection:**
  - Datasets were randomly split into 75% training set, 25% testing set
  - Six models were tested on the three datasets using five-fold cross validation (R 3.5.1, caret): linear regression (LM), random forest (RF), Bayesian regularized neural network (BRNN), model tree (M5P), extreme learning machine (ELM), multi-level perceptron (MLP)

MATERIALS & METHODS

- **Age-based Analyses:**
  - Based on model selection results, LM (baseline), RF, and BRNN were tested
  - Data were split by age, in two different ways:
    - Additive: all cumulative prior data were used as training set
    - Fixed: only records from the most recent 3 months were used as training set
  - Each training set had 4 testing sets, containing the next 4 months of TS records
- **Date-based Analysis:**
  - LM (baseline) and RF were tested
  - **Monthly:** Monthly TS was forecasted for each month, beginning in 2017
    - Training sets set up similarly to age: Additive and two fixed (1 year windows and 4 month windows)
    - Testing sets: subsequent 4 months
  - **Weekly:** Weekly TS was forecasted for each month, beginning in 2017
    - Training sets: Additive, fixed 1 year, fixed 1 month
    - Testing: subsequent 4 weeks
- Health data did not improve model predictions and were excluded
- All models were trained using five-fold cross validation on training sets, and model performance was based on test set root mean square error (RMSE) and correlation (r) of predicted and actual TS

RESULTS

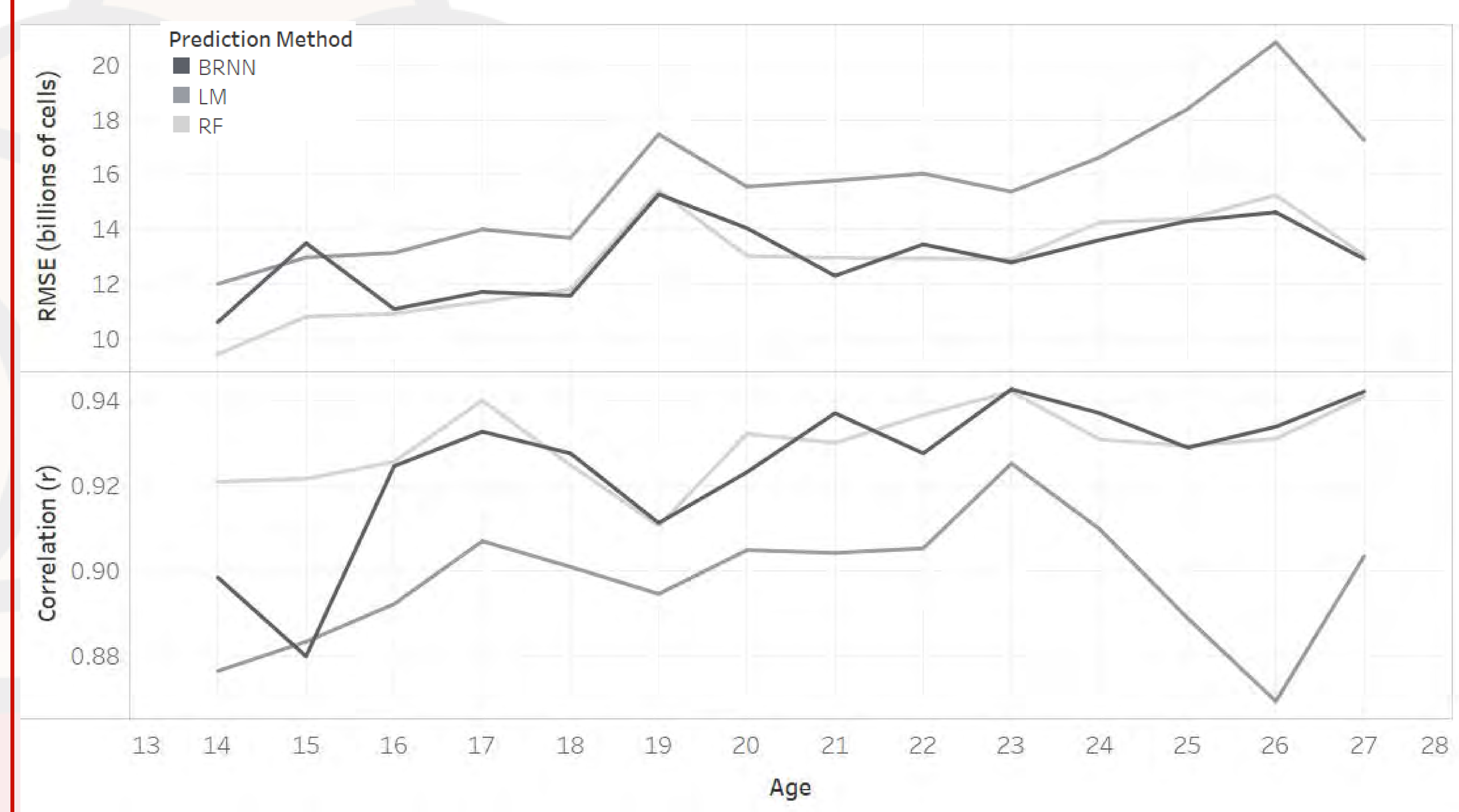

Figure 1. Comparison of prediction accuracy for monthly TS production LM, RF, and BRNN on additive training sets defined by age of bull, and with lag variables in the prediction model. Shown is the root mean squared error (RMSE; billions of cells) in the testing set one month out (top) and correlation (r) between actual and predicted TS production records of the same bulls one month into the future (bottom)

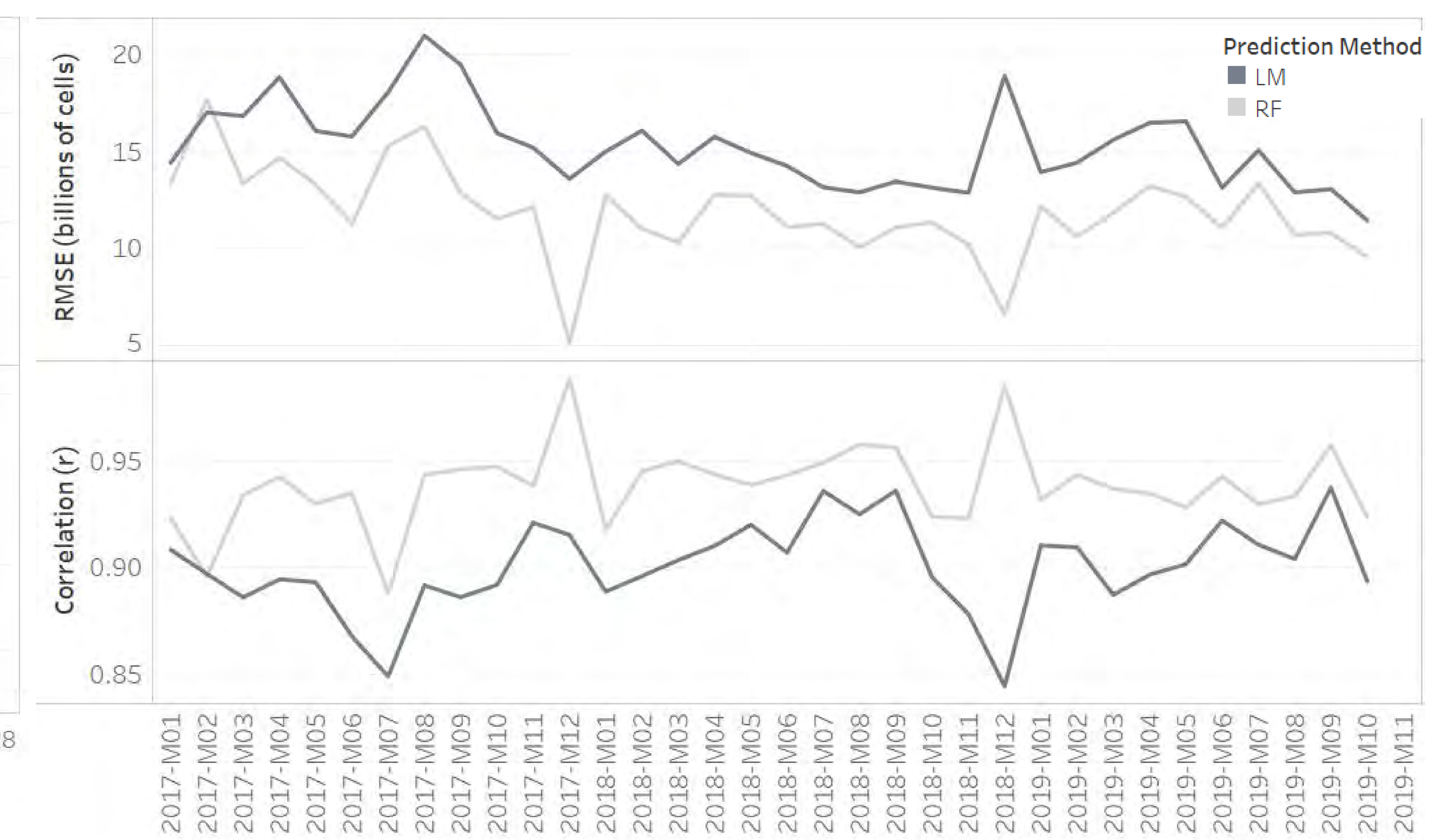

Figure 2. Comparison of predictive ability for monthly TS production using LM and RF using additive training sets defined by calendar date, using lag variables in the prediction model. Shown is the root mean squared error (RMSE; billions of cells) in the testing set one month out (top) and correlation (r) between actual and predicted TS production records of the same bulls one month into the future (bottom)

RESULTS

Table 3. Prediction accuracy of the best performing model within each method for daily, weekly, or monthly TS production using randomly constructed training and testing sets

	Daily TS		Weekly TS		Monthly TS	
Method	RMSE	r	RMSE	r	RMSE	r
LM	3.3	0.73	5.1	0.82	16.9	0.87
RF	3.0	0.77	4.7	0.84	14.5	0.90
BRNN	3.0	0.77	4.8	0.84	15.6	0.91
M5P	3.1	0.76	4.9	0.84	15.7	0.91
MLP	3.6	0.65	7.7	0.73	28.7	0.69
ELM	3.4	0.71	5.7	0.78	18.2	0.84

Table 4. Prediction accuracy for monthly TS production using training and testing sets defined by age of bull, and using additive (cumulative) or fixed (most recent 3 months) training windows, with lag variables

		Age + 1 month		Age + 2 months		Age + 3 months		Age + 4 months	
Model	Training Set	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)
LM	Additive	15.7 (2.4)	0.90 (0.02)	16.4 (2.0)	0.89 (0.02)	17.0 (1.8)	0.89 (0.02)	17.7 (1.6)	0.89 (0.02)
RF	Additive	12.7 (1.8)	0.93 (0.01)	13.4 (1.4)	0.93 (0.01)	14.0 (1.3)	0.92 (0.01)	14.5 (1.2)	0.92 (0.02)
BRNN	Additive	13.0 (1.4)	0.93 (0.02)	13.7 (1.0)	0.92 (0.02)	14.6 (1.6)	0.92 (0.03)	15.3 (1.7)	0.91 (0.03)
LM	Fixed 3 mo	15.1 (2.0)	0.90 (0.01)	15.7 (1.6)	0.90 (0.02)	16.3 (1.5)	0.90 (0.02)	17.0 (1.3)	0.89 (0.02)
RF	Fixed 3 mo	12.9 (1.8)	0.93 (0.01)	13.5 (1.5)	0.93 (0.01)	14.2 (1.4)	0.92 (0.01)	14.7 (1.2)	0.92 (0.01)
BRNN	Fixed 3 mo	13.5 (1.9)	0.92 (0.02)	14.5 (1.9)	0.91 (0.02)	15.5 (2.4)	0.91 (0.03)	16.3 (3.4)	0.91 (0.03)

Table 5. Prediction accuracy for monthly total sperm (TS) production using training and testing sets defined by calendar date, and using additive (cumulative) or fixed (most recent 1 year or 4 months) training windows, with lag variables

		Date + 1 month		Date + 2 months		Date + 3 months		Date + 4 months	
Model	Training Set	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)
LM	Additive	15.3 (2.2)	0.90 (0.02)	15.4 (2.0)	0.90 (0.02)	15.5 (2.1)	0.90 (0.02)	15.6 (2.1)	0.90 (0.02)
RF	Additive	11.9 (2.4)	0.94 (0.02)	12.0 (2.5)	0.94 (0.02)	11.8 (2.3)	0.94 (0.02)	11.8 (2.4)	0.94 (0.02)
LM	Fixed 1 yr	15.3 (2.2)	0.90 (0.02)	15.6 (2.0)	0.90 (0.02)	15.8 (2.1)	0.90 (0.02)	15.9 (2.0)	0.89 (0.02)
RF	Fixed 1 yr	12.1 (2.5)	0.94 (0.02)	12.2 (2.5)	0.94 (0.02)	11.9 (2.2)	0.94 (0.02)	11.9 (2.3)	0.94 (0.02)
LM	Fixed 4 mo	15.6 (2.5)	0.90 (0.03)	15.8 (2.6)	0.89 (0.02)	15.9 (2.6)	0.89 (0.02)	16.3 (2.9)	0.89 (0.03)
RF	Fixed 4 mo	13.1 (2.2)	0.93 (0.02)	13.1 (2.0)	0.93 (0.02)	13.0 (1.9)	0.93 (0.01)	13.1 (1.7)	0.93 (0.01)

Table 6. Prediction accuracy for weekly total sperm (TS) production using training and testing sets defined by calendar date, and using additive (cumulative) or fixed (most recent 1 year or 1 month) training windows, with lag variables

		Date + 1 week		Date + 2 weeks		Date + 3 weeks		Date + 4 weeks	
Model	Training Set	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)	RMSE (SD)	r (SD)
LM	Additive	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)
LM	Fixed 1 yr	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)	4.8 (0.6)	0.84 (0.04)
RF	Fixed 1 yr	4.4 (0.6)	0.86 (0.04)	4.4 (0.6)	0.86 (0.04)	4.4 (0.6)	0.86 (0.04)	4.5 (0.6)	0.86 (0.04)
LM	Fixed 1 mo	4.9 (0.6)	0.84 (0.04)	4.9 (0.6)	0.84 (0.04)	4.9 (0.6)	0.84 (0.04)	4.9 (0.7)	0.83 (0.04)
RF	Fixed 1 mo	4.7 (0.6)	0.85 (0.04)	4.7 (0.6)	0.85 (0.04)	4.7 (0.6)	0.85 (0.04)	4.7 (0.6)	0.85 (0.04)

CONCLUSIONS

- Total sperm production can be predicted with >90% accuracy up to four months into the future when using RF or BRNN
- Fixed-length training sets can improve computational feasibility in exchange for a small reduction in predictive ability
- Management factors, such as age at collection, frequency of collection, breed, barn, year, season, and scrotal circumference can contribute to prediction accuracy; however, data regarding the incidence of health events did not improve predictions
- Weekly predictions of total sperm output can be used to enhance the operational efficiency of semen processing, scheduling, and inventory control
- Monthly predictions can be used to forecast product availability for specific markets and manage the collection schedules of bulls of various ages
- Future studies should identify management, environmental, and genetic factors that can further improve prediction accuracy for (groups of) bulls whose semen is destined for specific markets and develop decision support tools to incorporate this information into standard operating procedures and replacement and inventory management decisions

REFERENCES

Amann, R. P. 1976. Bull management to maximize semen output. Page 1 in Proc. 6th NAAB Tech. Conf., February 20 to 21, 1976.

Everett, R.W., B. Bean, and R.H. Foote. 1978. Sources of Variation of Semen Output. J. Dairy Sci. 61:90–95. doi:10.3168/jds.s0022-0302(78)83555-3.

Mathevon, M., M.M. Buhr, and J.C.M. Dekkers. 1998. Environmental, Management, and Genetic Factors Affecting Semen Production in Holstein Bulls. J. Dairy Sci. 81:3321-3330.

Murphy, E.M., A.K. Kelly, C. O'Meara, B. Eivers, P. Lonergan, and S. Fair. 2018. Influence of bull age, ejaculate number, and season of collection on semen production and sperm motility parameters in holstein friesian bulls in a commercial artificial insemination centre. J. Anim. Sci. 96:2408–2418. doi:10.1093/jas/sky130.

Taylor, J.F., B. Bean, L.C.E. Marshall, and J.J. Sullivan. 1985. Genetic and Environmental Components of Semen Production Traits of Artificial Insemination Holstein Bulls. J. Dairy Sci. 68:2703–2722. doi:10.3168/jds.S0022-0302(85)81155-3.

ACKNOWLEDGEMENTS

Thank you ABS Global Inc. for providing data and funding this research.