animal

Prediction of dry-cured ham weight loss and prospects of use in a pig breeding program

V. Bonfatti[†] and P. Carnier

Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università 16, Legnaro 35020, Italy

*Large ham weight losses (**WL**) in dry-curing are undesired as they lead to a loss of marketable product and penalise the quality of the dry-cured ham. The availability of early predictions of WL may ease the adaptation of the dry-curing process to the characteristics of the thighs and increase the effectiveness of selective breeding in enhancing WL. Aims of this study were (i) to develop Bayesian and Random Forests (**RFs**) regression models for the prediction of ham WL during dry-curing using on-site infrared spectra of raw ham subcutaneous fat, carcass and raw ham traits as predictors and (ii) to estimate genetic parameters for WL and their predictions (**P-WL**). Visible-near infrared spectra were collected on the transversal section of the subcutaneous fat of raw hams. Carcass traits were carcass weight, carcass backfat depth, lean meat content and weight of raw hams. Raw ham traits included measures of ham subcutaneous fat depth and linear scores for round shape, subcutaneous fat thickness and marbling of the visible muscles of the thigh. Measures of WL were available for 1672 hams. The best prediction accuracies were those of a Bayesian regression model including the average spectrum, carcass and raw ham traits, with R^2 values in validation of 0.46, 0.55 and 0.62, for WL at end of salting (23 days), resting (90 days) and curing (12 months), respectively. When WL at salting was used as an additional predictor of total WL, the R^2 in validation was 0.67. Bayesian regressions were more accurate than RFs models in predicting all the investigated traits. Restricted maximum likelihood (REML) estimates of genetic parameters for WL and P-WL at the end of curing were estimated through a bivariate animal model including 1672 measures of WL and 8819 P-WL records. Results evidenced that the traits are heritable ($h^2 \pm SE$ was 0.27 ± 0.04 for WL and 0.39 ± 0.04 for P-WL), and the additive genetic correlation is positive and high ($r_a = 0.88 \pm 0.03$). Prediction accuracy of ham WL is high enough to envisage a future use of prediction models in identifying batches of hams requiring an adaptation of the processing conditions to optimise results of the manufacturing process. The positive and high genetic correlation detected between WL and P-WL at the end of dry-curing, as well as the estimated heritability for P-WL, suggests that P-WL can be successfully used as an indicator trait of the measured WL in pig breeding programs.*

Implications

We developed models predicting the dry-cured ham weight loss by means of information that can be acquired at the slaughter plant from carcasses and raw hams. The prediction accuracy is such to envisage a future use of the models in early identification of batches of hams requiring an adaptation of the processing conditions to optimise results of the manufacturing process. Due to the high genetic correlation between the predicted and measured weight loss, the proposed models may serve as a phenotyping tool in pig breeding programs for enhancing the technological quality of raw hams for dry-curing.

[†] E-mail: valentina.bonfatti@unipd.it

Introduction

Dry-cured ham marked with the European 'protected designation of origin' (**PDO**) label is the most valuable product of the pig industry in Italy, being the value of raw hams nearly 30% of the total carcass market value (Centro Ricerche Produzioni Animali, 2010). Protected designation of origin dry-cured hams require a curing period of at least 10 months and only salt is allowed as a preservative. Quality of raw hams and their suitability for curing are key features as the production process is unable to correct inadequate characteristics of the raw meat (Bosi and Russo, 2004). Dehydration during dry-curing is a fundamental process, hindering the development of anomalous fermentations and ensuring typical sensorial properties. Large ham weight

losses (**WL**) during dry-curing, resulting from poor quality of the processed meat, lead to a loss of marketable product, penalise the quality of the dry-cured ham and, from an economic point of view, are undesired.

Since the early nineties, the Italian pig breeders association (ANAS, Rome, Italy) addressed the selection goal of the Large White, Landrace and Duroc populations to the enhancement of WL. Phenotyping for the total WL at the end of dry-curing (**WLC**), which requires completion of the entire dry-curing process, is, however, troublesome because of penalising effects on the length of the generation interval and on the magnitude of the achievable selective response. As a consequence, such selection is performed indirectly by exploiting the genetic correlation ($r_g = 0.65$; Buttazzoni *et al.*, 1993) between WLC and the ham weight loss after 7 days of salting. Individual measures of weight loss at 7 days, as required in selective breeding, are not easy to obtain either as they require availability of facilities for ham processing and traceability of individual hams at the processing plant.

Early and accurate predictions of WLC ease the adaptation of the dry-curing process to the characteristics of the thighs and increase the effectiveness of selective breeding in the enhancement of WLC. Besides geographical area of origin, genetic line, age and weight at slaughter and manufacturing practices, dry-cured ham PDO procedural guidelines dictate requirements for quality of raw ham subcutaneous fat (Consorzio del Prosciutto di Parma, 1992; Consorzio del Prosciutto di San Daniele, 2007). For regular and large-scale screening for fat quality compliance, near infrared spectroscopy is an interesting alternative to conventional methods as it enables fast, simple and concurrent assessment of contents of many chemical compounds. This technology has also been successfully used to predict different technological and sensory attributes of dry-cured hams (Collell *et al.*, 2011; Prevolnik *et al.*, 2014; Prieto *et al.*, 2017). Hence, infrared spectra might play a role in early prediction of other relevant characteristics of raw hams, including their suitability for dry-curing. The aims of this study were (i) to develop Bayesian and Random Forest (**RF**) regression models for the prediction of ham WL during dry-curing using on-site infrared spectra of raw ham subcutaneous fat, carcass and raw ham traits as predictors, (ii) to assess the prediction accuracy of the investigated models and (iii) to estimate genetic parameters for WL and their predictions (**P-WL**).

Material and methods

Animals
This study involved 1672 crossbred pigs, which were offspring of 50 boars of the Goland C21 sire line (Gorzagri, Fonzaso, Italy) and 144 Large White-derived crossbred sows. All piglets were born and reared at the C21 sib-testing program farm (Todi, Italy) under consistent feeding conditions as detailed by Rostellato *et al.* (2015). Pigs were slaughtered in groups of 70 animals each (24 slaughter groups, from

December 2015 to July 2017), after CO_2 stunning, at the same slaughterhouse (Montorsi, Magreta, Italy) at 9 months of age and an average body weight of 165 ± 14 kg. Age at slaughter was constrained to a minimum of 9 months to comply with guidelines of Parma dry-cured ham production (Consorzio del Prosciutto di Parma, 1992). After slaughter, all carcasses were weighted, and hams were removed from both halves and weighted (weight before trimming). Hams were then dry cured for 12 months following the Parma ham PDO specification (Consorzio del Prosciutto di Parma, 1992).

Carcass and raw ham traits
Measures of weight and backfat depth of carcass and weight of raw hams were obtained for each pig. The Fat-O-Meater optical probe (Carometec, Soeborg, Denmark) was used to assess carcass backfat and loin depth. These measures were taken in compliance with the European legislation (2014/38/EU: Commission Implementing Decision of 24 January 2014 authorising methods for grading pig carcasses in Italy).

All left thighs were scored by a trained expert, using a linear grading system, for round shape (from $0 =$ low roundness to $4 =$ high roundness), subcutaneous fat depth (from $-4 =$ low depth to $4 =$ high depth) and marbling of the visible muscles of the thigh (from $0 =$ low to $4 =$ high). Ham subcutaneous fat depth was measured in the proximity of *m. semimembranosus* and *m. quadriceps femoris* using a gauge and a portable ultrasound system (Aloka SSD 500 equipped with UST-5512 7.5 MHz linear transducer probe, Hitachi Medical Systems S.p.A., Milan, Italy), respectively.

Spectra acquisition
On-site visible-near infrared spectra were collected from raw hams after trimming (24 h after slaughter). For each ham, a total of five spectra were acquired on the transversal section of subcutaneous fat at 5 fixed sites: (1) cranio-medial, (2) caudo-medial, (3) caudo-lateral, (4) lateral and (5) cranio-lateral. Spectra were collected using a post-dispersive diode array scanning monochromator spectrometer LabSpec® 5000 (ASD Inc., Boulder, CO, USA) working in reflectance mode over the spectral range between 350 and 2500 nm (2151 spectral variables), to which a 1.5 m fiber-optic contact probe with a quartz-halogen source was attached. Data acquisition was carried out using the Indico™ Pro software (ASD Inc., Boulder, CO, USA). Spectra reflectance (R) was converted into absorbance (A) according to the equation $A = \log(1/R)$.

Samples with standardised Mahalanobis distance > 3 were classified as outliers and eliminated, leaving 1619 samples for subsequent analyses. Spectra were transformed with a first derivative mathematical treatment. Spectral variables were standardised to a null mean and unit variance before use in prediction.

Ham weight losses in dry-curing
Hams were dry-cured for 368 ± 4 days. The three major steps occurring during ham processing (salting, resting and curing)

are described in the Parma ham PDO specification (Consorzio del Prosciutto di Parma, 1992). The initial weight of the hams was measured after removal of rind and excess of subcutaneous fat (trimming). Then, salt was added to hams, which were stored for 6 or 7 days at 1°C to 4°C and 80% of humidity. Salt residues were then removed, and hams were sprinkled again with tiny amounts of salt and stored back in cold rooms for additional 15 to 18 days. The salting phase lasted 23 days. In this phase, the water dripping out of the ham dissolves the salt, which diffuses towards the interior of the muscle (Čandek-Potokar and Škrlep, 2012).

After removal of salt residues, hams were stored in resting rooms for 70 days at 1°C to 5°C and 75% of humidity (resting phase). During resting, the salt equilibrates inside the entire ham (Čandek-Potokar and Škrlep, 2012).

After resting, hams were transferred to the curing phase. In the early stage of curing, hams were hung on frames in properly ventilated rooms (pre-maturation rooms), ensuring a gradual and continuing drying. The exposed surface of hams was then softened with a mixture of minced lard and salt to prevent the over-dehydration of exposed muscular tissues. Hams were finally moved to 'cellars', cooler and less ventilated than pre-maturation rooms, where they remained until the end of curing. In this phase, the water loss occurs due to evaporation from the surface while muscle proteins undergo intensive proteolysis conferring dry-cured ham its characteristic texture and flavor (Čandek-Potokar and Škrlep, 2012).

Measures of weight loss (%) at 23 days (end of salting, **WLS**), 90 days (at resting, **WLR**), 12 months (i.e., at the end of dry-curing, WLC) and weight loss from 23 to 90 days (**WLSR**), 23 days to 12 months (**WLSC**) and 90 days to 12 months (**WLRC**) were calculated.

Explanatory variables included in prediction models
Variables included in prediction models are summarised in Table 1. Although the slaughter group effect has been proved to be one of the most important sources of variation in WL occurring during dry-curing (Sturaro, 2004), it was not included in prediction models in order to make the prediction equation generalisable and extendable to slaughter groups different from those considered in this study.

The slaughter batch includes all the possible sources of variation arising from the transportation of pigs to the slaughterhouse, the slaughtering process as well as the environmental conditions at the slaughterhouse. The slaughter batch captures also the variation in weight loss arising from effects common to pigs belonging to the same fattening batch (pigs born in a given period are slaughtered also in

Table 1 *Predictors included in regression models for the prediction of weight losses in dry-curing of hams from heavy pigs*

Model[1]	Algorithm	Collection site in raw ham subcutaneous fat[2]					Average spectrum[3]	Carcass traits[4]	Raw ham traits[5]	Weight loss after salting[6]
		1	2	3	4	5				
1	BR	X								
2	BR		X							
3	BR			X						
4	BR				X					
5	BR					X				
6	BR						X			
7	BR	X	X	X	X	X				
8	BR						X	X		
9	BR						X		X	
10	BR/RF						X	X	X	
11	BR/RF						X	X	X	X
12	BR	X	X	X	X	X		X	X	
13	BR/RF	X	X	X	X	X		X	X	X
14	BR							X		
15	BR								X	
16	BR							X	X	
17	BR							X	X	X
18	BR									X

BR = Bayesian regression, RF = random forests.
[1]Models 11, 13, 17 and 18 were used only for the prediction of weight loss at the end of dry-curing.
[2]A total of five spectra were acquired on the transversal section of subcutaneous fat at five fixed sites: 1 = cranio-medial, 2 = caudo-medial, 3 = caudo-lateral, 4 = lateral, 5 = cranio-lateral point.
[3]Average of the spectra acquired at sites 1, 2, 3, 4 and 5.
[4]Carcass traits: carcass weight, backfat depth, lean meat content and carcass raw ham percentage.
[5]Raw ham traits: round shape score, marbling score, subcutaneous fat score and subcutaneous fat depth measured nearby *m. semimembranosus* and nearby *m. quadriceps femoris.*
[6]Weight loss after salting is the weight loss of the ham measured at the end of salting (23 days) as a percentage of the raw ham initial weight after trimming.

the same batch). However, inclusion of slaughter batch effects in the statistical models under study is unfeasible because such models must serve as prediction tools of ham weight loss for future samples of pigs, and the nature of slaughter batch effects is random. For future slaughter batches, the quantitative effects on the traits of concern are unknown when the prediction is made. Hence, inclusion of slaughter batch effects in the investigated models would lead to the development of tools not applicable in prediction.

As the relationship between ham WL and carcass or ham traits is linear (Sturaro, 2004), carcass and ham traits were included in models as continuous variables. All the investigated traits and the explanatory variables included in models, with the only exception of sex, were standardised to null mean and unit variance before the regression analysis.

Prediction models for ham weight losses
Partial regression coefficients (i.e., the estimated effects of predictors) were estimated using Bayesian parametric shrinkage and variable selection procedures and Markov chain Monte Carlo methods performing numerical integration through the Gibbs sampler. Such modelling procedures demonstrated superior ability to cope with noise in comparison with partial least squared regression (Bonfatti et al., 2017a) because they decrease the estimator variance to guarantee the stability of the estimates and perform variable selection excluding uninformative predictors (de los Campos et al., 2013).

The general model used in this study was a parametric regression model, known as Bayes B in genomic data analysis, implemented in the BGLR package (de los Campos and Pérez-Rodriguez, 2014) of the R software (R Development Core Team, 2013). In such model, the prior density assigned to effects of spectral variables was a finite mixture prior of a point of mass at zero and a scaled-t slab (Habier et al., 2011). The hyperparameters of such prior density (i.e., π, the prior proportion of non-zero effects and S_β, the scale parameter) were treated as unknown to which prior densities (a beta and a gamma density for π and S_β, respectively) were assigned by using built-in rules in BGLR yielding proper, but weakly informative, priors. For sex, carcass and ham traits and WLS effects, the prior density was a flat prior. Models and algorithms implemented in BGLR are comprehensively explained in Pérez and de los Campos (2014).

A unique Gibbs chain of 120 000 iterations was run for each Bayesian analysis. Samples were saved every 250 iterations. The length of the burn-in period was determined in a preliminary analysis by visual inspection of the chain and was set to 20 000 samples.

Assessment and comparison of model prediction performance
Data available for each trait after outlier elimination were randomly assigned to 10 non-overlapping subsets. Prediction equations were developed using nine subsets at a time and validated on the remaining subset. Prediction performance was assessed by computing the R^2 between the predicted

and observed values in the validation subset (**R^2v**) and the RMSE of prediction in validation (**RMSEv**). The statistical significance of the difference in the average R^2v and RMSEv between models was assessed by paired t tests performed using the R software (Bonfatti et al., 2017a).

Random forest algorithm
The sets of predictors ensuring the best prediction performance in Bayesian regression analyses were considered also in a machine learning approach, the RF algorithm (Breiman, 2001), implemented in the randomForestSRC package (Ishwaran and Kogalur, 2018) of the R software.

RF is a learning method that creates an ensemble (a 'forest') of base learners (decision trees) able to predict better than any of its members. Further improvement in prediction performance is ensured by randomisation in the learning process which occurs at two levels: (1) at the root node of a tree (i.e., when a tree starts to grow) by drawing a bootstrapped sample from the training data and (2) at the splitting of a node, which yields two smaller daughter nodes, by sampling m variables from a set of predictors. At each node of a tree, the objects are split using the best splitting rule, which involves a single predictor.

Fine tuning of the RF algorithm was performed by comparing the error in the prediction of WLC for different values for the number of trees in the forest (1000, 10 000 or 50 000), terminal node size (3, 5 or 10) and number m of sampled variables at node splitting ($n/3$, $n/2$, $n/1.2$, where n is the total number of predictors), which resulted in 27 RF regression analyses. It was observed that 1000 trees were enough to stabilise the prediction errors, a terminal node size of five slightly decreased the prediction error when compared with larger node sizes and $m = n/1.2$ was better than $m = n/3$ (default value, as suggested by Liaw and Wiener (2002) for regressions). Hence, such parameter values were used in the RF regression analysis of all traits.

Genetic parameters of measures and predictions of weight loss at the end of curing
An additional dataset, including records of carcass traits, raw ham traits and spectral variables of 7437 crossbred pigs, was extracted from the historical database of the genetic evaluation program of the Goland C21 sire line. After removing records with missing information on carcass or raw ham traits and spectra that were considered outliers, the dataset included 7200 records. Predictions of WLC (**P-WLC**) were obtained by applying Model 10 (see Table 1) to these additional records.

Using all the available information, (co)variance components for WLC and P-WLC were estimated in a bivariate animal model REML analysis using VCE software (version 6.0; Groeneveld et al., 2010). The genetic analysis considered a total of 1672 WLC and of 8819 P-WLC records. Of these, 1619 records had a value for both WLC and P-WLC. For both traits, the model included the fixed effects of sex (female or castrated male) and slaughter group (175 groups) and the random additive genetic effects of animals. Additive

relationships between animals were traced back for as many generations as possible. Additive relationships were computed on the basis of at least 10 generations of known ancestors for the paternal side. Only the sire, the maternal grandsire and the granddam were known for the dams of the crossbred pigs. Sires and dams of pigs with a record for WLC, P-WLC or both were unrelated. The final pedigree included 15 730 animals.

Results and discussion

Data

Descriptive statistics for the investigated traits are reported in Table 2. Average weight at slaughter and initial weight of hams were 165 ± 14 and 13.7 ± 1 kg, respectively, in compliance with the product specification of PDO dry-cured hams, which requires weight at slaughter to be at least 146 kg and raw hams to weight at least 10 kg, in order to ensure optimal

Table 2 *Descriptive statistics for heavy pig carcass, ham and dry-curing traits* (n = 1672)

Trait[1]	Mean	SD	Minimum	Maximum
Carcass traits				
Weight (kg)	140.1	11.8	102.8	181.4
Backfat depth (mm)	26.0	4.6	14	46
Lean meat content (%)	49.8	2.6	40.3	58.3
Raw ham percentage (%)	20.1	0.9	16.9	26.3
Raw ham traits				
Round shape score	1.68	0.80	0	4
Marbling score	1.53	0.81	0	4
Subcutaneous fat depth score	0.24	1.74	-4	4
Subcutaneous fat depth (mm)				
Nearby *m. semimembranosus*	18.2	5.6	8	41
Nearby *m. quadriceps femoris*	6.6	1.1	3.8	10.3
Dry-curing traits				
Initial ham weight (kg)	13.7	1.0	10.5	17.1
Final ham weight (kg)	9.9	0.9	6.9	12.8
Ham weight loss (%)				
At the end of salting	3.6	0.6	1.5	6.9
At the end of resting	15.6	1.3	11.3	20.5
At the end of curing	27.8	2.4	21.6	38.2
From salting to the end of resting	12.5	1.1	8.4	15.9
From resting to the end of curing	14.5	2.0	8.8	23.0
From salting to the end of curing	25.1	2.2	19.6	34.7

[1]Backfat depth: measured using the Fat-O-Meater optical probe; carcass lean meat content: estimated as $y = 65.9993 + 0.4619 \cdot x_1 - 0.0048 \cdot x_2$, where x_1 is the Fat-O-Meater measure of backfat depth (including skin; mm), and x_2 is the Fat-O-Meater measure of loin depth (mm); raw ham percentage: percentage ratio of raw (untrimmed) ham weight to carcass weight; round shape score: 0 = low, 4 = high; marbling score: marbling of visible thigh muscles, 0 = low, 4 = high; subcutaneous fat depth score: -4 = low, 4 = high. Initial ham weight: weight of the raw ham after trimming.

curing conditions. Age at slaughter in our sample was 278 ± 1 days, greater than the minimum age of 270 days required by the product specification to ensure optimal body tissue composition (Consorzio del Prosciutto di Parma, 1992).

Dry-curing lasted on average 368 days and ranged from 354 to 379 days. At the end of the process, thighs lost from 22% to 38% of their initial weight. The rate of weight loss was higher in the first phases, with an average loss of 21 g/day during salting and resting, whereas in the curing phase, the rate of loss decreased to 6 g/day in agreement with values reported in previous studies (Russo *et al.*, 1992; Gallo *et al.*, 1999; Sturaro, 2004). The rate of weight loss is greater during salting when the osmotic and hygroscopic effects of salt, which covers the entire thigh, remove water more easily from tissues. Salting is also the phase of processing where the greatest variability in losses is observed: WLS had a CV of 17% unlike WLR and WLC, which, even with average values of 15% and 27%, have a lower coefficient of variability.

Prediction accuracy of models using only the visible-near infrared spectra information

Accuracy of models for the prediction of ham WL is reported in Table 3. Within the investigated traits, WLS was generally predicted with lower accuracy than WLR and WLC, which was the trait predicted with the highest accuracy. Models 1 to 5, which perform prediction using only the spectral variable set acquired at one of the five sites at a time, had the lowest accuracy (R^2v from 0.29 to 0.31 for WLS, from 0.28 to 0.36 for WLR and from 0.33 to 0.39 for WLC). The use of spectra collected at different sites resulted in similar accuracy for WLS and WLRC, whereas the other traits were significantly better predicted using the spectra at sites 4 and 5.

When the spectral variable sets were averaged across the five sites (Model 6), R^2v increased approximately by 10 percentage points in comparison with that of Models 1 to 5. The concurrent use of the five spectral variable sets (10 755 variables; Model 7) did not improve the prediction accuracy of WL in comparison with that of Model 6. Irrespective of the trait predicted, Model 7 had a significantly lower R^2v and a significantly higher RMSEv than Model 6.

The uniformity in the R^2v of Models 1 to 5, as well as results of Model 7, indicates that each of the five spectral variable sets contains some valuable information for the prediction of WL, but their concurrent use does not provide any additional information. This can either indicate a non-specific contribution of the different sites to the overall prediction of WL or a low repeatability of the reflectance measure at some sites, which may generate noise whose effects are greatly reduced when averaging the spectra.

The most influential spectral variables for the prediction of ham WL, identified on the basis of solutions of Model 6, are reported in Figure 1. The spectral regions markedly associated with variations in WL were located below 1000 nm, which corresponds primarily to the visible spectra region. This result was consistent for different traits, with the only

Table 3 *Parameters of prediction performance in validation (R^2v and RMSEv) of Bayesian regression models for weight losses in dry-curing of hams from heavy pigs[1,2,3]*

Model	WLS R^2v	WLS RMSEv	WLR R^2v	WLR RMSEv	WLC R^2v	WLC RMSEv	WLSR R^2v	WLSR RMSEv	WLRC R^2v	WLRC RMSEv	WLSC R^2v	WLSC RMSEv
1	0.31[b]	0.83[b]	0.32[bc]	0.83[ab]	0.35[b]	0.80[b]	0.24[ab]	0.87[ab]	0.30[b]	0.83[b]	0.33[bc]	0.82[bc]
2	0.29[b]	0.84[b]	0.26[a]	0.86[a]	0.33[b]	0.82[b]	0.21[a]	0.89[a]	0.31[b]	0.83[b]	0.31[b]	0.83[b]
3	0.29[b]	0.84[b]	0.28[ab]	0.85[a]	0.35[b]	0.81[b]	0.22[a]	0.88[a]	0.32[b]	0.83[b]	0.33[b]	0.82[b]
4	0.30[b]	0.84[b]	0.36[cd]	0.80[bc]	0.41[cd]	0.77[cd]	0.29[bc]	0.84[bc]	0.34[b]	0.82[b]	0.39[cde]	0.79[cd]
5	0.29[b]	0.85[b]	0.36[cd]	0.80[b]	0.39[bc]	0.79[bc]	0.29[bc]	0.84[bc]	0.30[b]	0.84[b]	0.36[bcd]	0.80[bcd]
6	0.42[d]	0.76[d]	0.46[f]	0.74[d]	0.52[f]	0.69[e]	0.38[d]	0.79[d]	0.46[d]	0.73[d]	0.49[f]	0.72[e]
7	0.39[c]	0.78[c]	0.42[e]	0.77[c]	0.46[e]	0.74[d]	0.33[c]	0.82[c]	0.39[c]	0.78[c]	0.43[e]	0.76[d]
8	0.44[ef]	0.75[e]	0.51[g]	0.70[e]	0.55[g]	0.67[e]	0.43[e]	0.75[ef]	0.47[de]	0.73[d]	0.51[g]	0.70[f]
9	0.46[fg]	0.73[f]	0.54[h]	0.68[f]	0.61[i]	0.62[g]	0.46[f]	0.74[f]	0.54[f]	0.68[e]	0.58[i]	0.65[h]
10	0.46[g]	0.74[f]	0.55[h]	0.67[f]	0.62[i]	0.62[g]	0.47[g]	0.73[g]	0.53[f]	0.69[e]	0.58[i]	0.65[h]
11	–	–	–	–	0.67[k]	0.57[i]	–	–	–	–	–	–
12	0.44[e]	0.75[de]	0.52[g]	0.69[e]	0.58[h]	0.65[f]	0.41[e]	0.76[e]	0.48[e]	0.72[d]	0.54[h]	0.68[g]
13	–	–	–	–	0.65[j]	0.60[h]	–	–	–	–	–	–
14	0.19[a]	0.90[a]	0.27[a]	0.85[a]	0.25[a]	0.87[a]	0.21[a]	0.89[a]	0.13[a]	0.94[a]	0.22[a]	0.89[a]
15	0.29[b]	0.84[b]	0.39[de]	0.79[bc]	0.43[de]	0.75[cd]	0.29[bc]	0.84[bc]	0.30[b]	0.84[b]	0.39[de]	0.78[d]
16	0.30[b]	0.84[b]	0.40[de]	0.77[c]	0.45[de]	0.75[cd]	0.30[c]	0.84[bc]	0.31[b]	0.84[b]	0.40[de]	0.77[d]
17	–	–	–	–	0.54[fg]	0.68[e]	–	–	–	–	–	–
18	–	–	–	–	0.38[bc]	0.79[bc]	–	–	–	–	–	–

WLS = weight loss at the end of salting; WLR = weight loss at the end of resting; WLC = weight loss at the end of curing; WLSR = weight loss during resting; WLRC = weight loss during curing; WLSC = weight loss during resting and curing; R^2v = R^2 in validation; RMSEv = RMSE in validation.
[1]Models are detailed in Table 1, different letters indicate significantly different performance parameter estimates at $P < 0.05$.
[2]All weight losses are percentages of raw ham initial weight after trimming.
[3]SD of R^2v and RMSEv values ranged from 0.01 to 0.09.

exception of WLS, albeit the quantitative influence of the informative spectral regions was variable across traits.

The weight loss after salting was mostly influenced by the wavelength regions in proximity of 979 nm (informative also for WLR and WLSR) and from 1711 to 1720 nm (informative for all traits). Variations in weight loss occurring in later stages of processing were mostly affected by other spectral variables (in proximity of 575 nm, from 775 to 790 nm and at 894, 933 and 1170 nm). These wavelengths correspond for a large part to regions of absorption of $-CH_2$ bonds. This is an indication that the information provided by the spectral variables is to be ascribed mostly to the fatty acid composition of subcutaneous fat. Interestingly, the spectral variables corresponding to 894, 933 and 1170 nm were specifically related to traits measured during the curing phase (WLC, WLRC and WLSC), indicating that the relationship between subcutaneous fat composition on WL is variable in the different phases of ham processing. To our knowledge, an association between fatty acid composition and ham curing losses has never been reported, but it must be pointed out that, likely, this association is indirect and it is due to the relationships between total fat content and fatty acid composition and between total fat content and ham weight loss. Increased fat deposition in pigs, which decreases ham WL, is associated with increased proportions of SFA in fat (Čandek-Potokar and Škrlep, 2012). As a consequence, the association between the relative content of PUFA in ham subcutaneous fat and ham weight loss is expected to be positive. Indeed, for our data, the estimated Pearson's correlations between ham weight loss and PUFA content in subcutaneous fat, between carcass backfat depth and ham weight loss and between carcass backfat depth and PUFA were 0.56, −0.42 and −0.53 ($P < 0.001$; data not presented in tables), respectively. Similar results have been observed when such correlations were computed by using ham subcutaneous fat depth in place of carcass backfat depth.

Prediction accuracy of models using infrared spectra and additional variables
Performance of models using infrared spectra and additional variables is reported in Table 3. When the carcass traits were added to the predictive models including the averaged spectral variable set, the R^2v increased by 2 to 5 percentage points (Model 8; $P < 0.05$) for all traits with the exception of WLRC.

Subcutaneous fat plays a key role in the maturing process by moderating excessive dryness of the underlying muscular parts, influencing the amount of salt absorbed and preventing modifications penalising product quality. Leaner hams are expected to have greater salt content because of the increased weight loss (Čandek-Potokar and Škrlep, 2012). It is, therefore, very important that subcutaneous fat is consistent, thick and uniformly distributed (Sturaro, 2004).

Carcass weight, backfat depth, lean meat content and average weight of the raw trimmed hams are traits correlated with carcass fatness. Increased carcass and ham weights at constant age are associated with a greater percentage of fat tissues and lower water content, which result in reduced ham weight curing losses (Bosi and Russo, 2004; Sturaro, 2004).
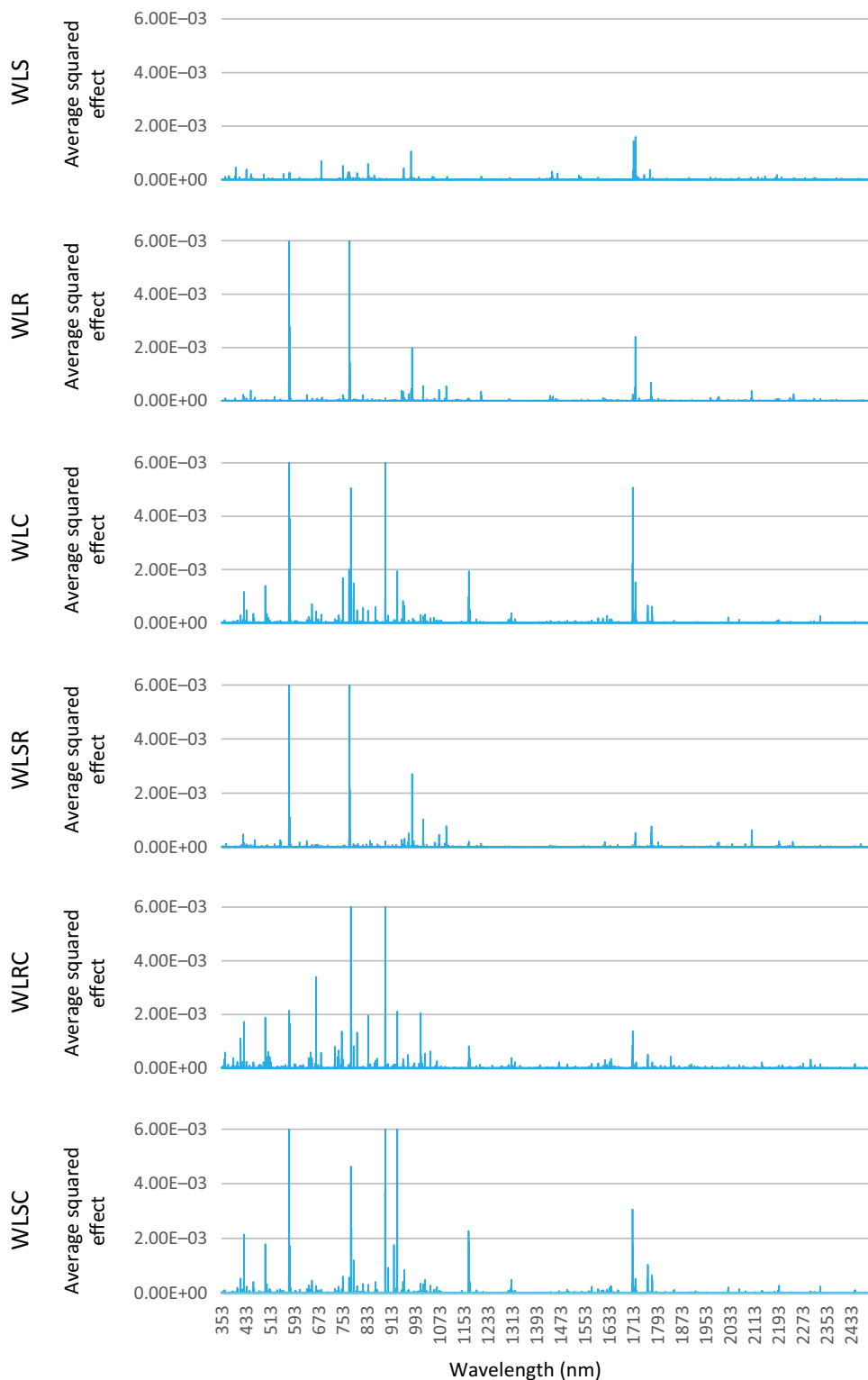
Figure 1 (colour online) Squared effect of the standardised visible-near infrared spectral variables measured in ham subcutaneous fat for the prediction of dry-curing weight losses of hams from heavy pigs. Effects were estimated with Bayesian regression Model 6 analysing 10 training sets and were averaged across training sets; all weight losses are expressed as percentages of raw ham initial weight after trimming. WLS = weight loss at the end of salting; WLR = weight loss at the end of resting; WLC = weight loss at the end of curing; WLSR = weight loss during resting; WLRC = weight loss during curing; WLSC = weight loss during resting and curing.

In addition, the muscular fraction is less exposed to the dehydration effect of salt as a consequence of greater adipose tissue coverage, which hampers exchanges between muscle and external environment (Bosi and Russo, 2004).

Model 9, which included the average spectrum and raw ham traits as predictors, was the model, together with Model 10, providing the highest accuracy for WLS, WLR, WLC, WLRC and WLSC. Only for WLSR, Model 10 exhibited

a slightly better prediction performance than Model 9 ($+1\%$, $P < 0.05$), but such difference was negligible from a practical perspective. Hence, when models included the spectrum and raw ham evaluation traits as predictors, the inclusion of carcass traits had trivial influence on model predictive ability, as many raw ham traits were measures or scores of the adiposity of the thighs.

The estimated squared effects of carcass and raw ham traits on ham WL (Model 10) are reported in Figure 2. The predictor with the largest effect on WLS was the marbling score, which is consistent with findings of Bosi and Russo (2004), and subcutaneous fat thickness measured nearby *m. quadriceps femoris*, which is the site of the ham where fat depth is at its lowest. In particular, these two traits were relevant predictors of the weight loss across all stages of dry-curing and were associated with the variation in all the investigated traits.

During resting, the effect of ham marbling on ham weight loss (WLSR) was less important than in the phase of salting, whereas ham subcutaneous fat depth and weight of the trimmed ham became the variables with the largest effects. Also the weight and the lean meat content of the carcass started to exert effects on WLSR. During the curing phase, the weight of the trimmed ham did not affect WLRC, whereas all other traits exhibited relevant influences, in particular ham marbling, and fat depth at *mm. semimembranosus* and *quadriceps femoris*. Due to variations in their relative importance for different processing stages, all predictors were associated with variations in WLC, with only ham roundness, carcass weight and carcass backfat thickness having negligible effects. Sex had a marked effect on WLR and WLSR only (data not reported), so it was particularly important when predicting the WL occurring in the resting phase. In a previous study (Sturaro, 2004), females had higher ham WL than castrated males. This effect can be attributed to high adipogenesis and low carcass water contents of castrated males relative to females.

The best prediction model, which included the spectra, carcass and ham traits (Model 10), had R^2v of 0.46, 0.55 and 0.62, for WLS, WLR and WLC, respectively and of 0.47, 0.53 and 0.58 for WLSR, WLRC and WLSC, respectively (Table 3).

Use of weight loss at the end of salting as an early predictor of total curing loss

We investigated the added value contributed by WLS in the prediction of WLC resulting from a model using spectra, carcass and ham traits. This model (Model 11, Table 3) predicted WLC with an R^2v of 0.67 (i.e., 5 percentage points more than Model 10).

Russo *et al.* (1991) investigated relationships between dry-curing weight loss, carcass traits and data recorded in early stages of processing (weight loss at refrigeration and at first salting). The weight loss at first salting, which can be measured after a few days since the arrival of the thigh at the ham processing factory, was the most effective predictor of dry-curing losses.

Phenotyping for WLC requires the completion of the entire dry-curing process, with penalising effects on the length of the generation interval and on the magnitude of the achievable selective response. As a consequence, the weight loss at first salting, which is genetically correlated to WLC (Buttazzoni *et al.*, 1993), has been serving, since the early nineties, as an indicator trait in the breeding programs operated by the Italian pig breeders association (ANAS) to address selection for decreasing dry-curing losses in the Large White, Landrace and Duroc populations. In our study, the increase in R^2v of models due to the use of WLS as a predictor of WLC translated into an increase in the correlation between the measured WLC and the predicted WLC of 3 percentage points (from 79% to 82%). This can be considered only a marginal improvement and it indicates that, when spectral information on subcutaneous fat, carcass traits and ham traits are available, there is no need to perform routine recordings of WLS, which requires individual traceability of hams at the ham factory.

Value of the spectral information in prediction models

Despite the overall moderate accuracy of models, it is worth noting that existing predictions of ham weight loss use weight loss at first salting and carcass traits as predictors. When WLS was used as the sole predictor of WLC, the R^2v was 0.38 (Model 18; Table 1). Hence, the R^2v was 39% lower than the R^2v provided by Model 10 and 27% lower than the one for the model including the spectra only (Model 6). The prediction based only on carcass traits (Model 14) had the lowest accuracy, with a R^2v for WLC of 25%.

Addition of ham traits to Model 14 (Model 16) increased the R^2v by 80%. Predictions exploiting the spectral information exhibited enhanced accuracies in comparison with models based on carcass traits, ham traits, WLS or combinations thereof. The increase in R^2v when predicting WLC ranged from 24 (Model 11 *v.* 17) to 120% (Model 8 *v.* 14).

Hence, ham traits and spectral information of subcutaneous fat are effective predictors of weight loss in dry-curing. The spectral information captures variation in fat quality characteristics of hams and contributes additional predictive value when compared with carcass and raw ham traits, which instead are descriptors of the amount of fat coverage and adiposity.

Prediction accuracy obtained by the random forest algorithm

Models 10, 11 and 13 provided the highest accuracies in the prediction of WLC. Behaviour of predictors used in these models was also investigated in regression analysis with RF. Results are reported in Table 4. All RF regression models had lower accuracies than Bayesian models. Generally, the explained variance obtained using RF was from 7 to 14 percentage points lower than the R^2v obtained using Bayesian regression. As a RF model is a non-linear predictor, it might be outperformed by other methods if the relationship
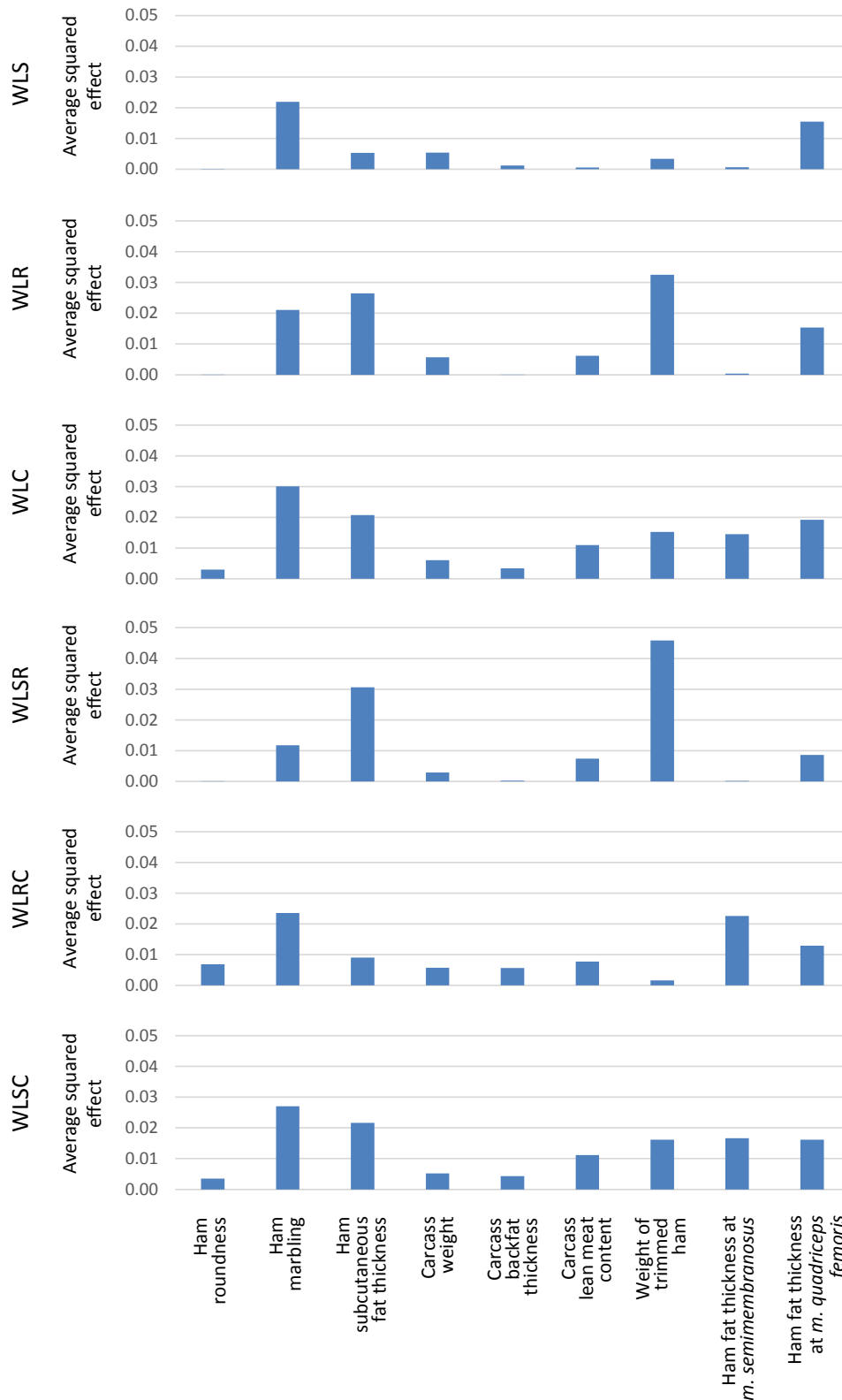
Figure 2 (colour online) Squared effect of the standardised carcass and ham variables for the prediction of dry-curing weight losses of hams from heavy pigs. Effects were estimated with Bayesian regression Model 10 analysing 10 training sets and were averaged across training sets; all weight losses are expressed as percentages of raw ham initial weight after trimming. WLS = weight loss at the end of salting; WLR = weight loss at the end of resting; WLC = weight loss at the end of curing; WLSR = weight loss during resting; WLRC = weight loss during curing; WLSC = weight loss during resting and curing.

between the response and the predictors is linear. Spectral variation is a linear response to the variation in chemical composition of subcutaneous fat, and all additional predictors used

in our study are linearly correlated to carcass or ham adiposity (Sturaro, 2004). The maximum proportion of variance explained by RF models was 57% (WLC; Model 11), which

Table 4 *Parameters of prediction performance in validation (R^2v and RMSEv) of random forest regressions for weight losses in dry-curing of hams from heavy pigs[1,2]*

Model	WLS		WLR		WLC		WLSR		WLRC		WLSC	
	R^2v	RMSEv	R^2v	RMSEv	R^2v	RMSEv	R^2v	RMSEv	R^2v	RMSEv	R^2v	RMSEv
10	0.39	0.78	0.45	0.74	0.49	0.72	0.37	0.85	0.44	0.76	0.46	0.74
11	–	–	–	–	0.57	0.66	–	–	–	–	–	–
13	–	–	–	–	0.54	0.68	–	–	–	–	–	–

WLS = weight loss at the end of salting; WLR = weight loss at the end of resting; WLC = weight loss at the end of curing; WLSR = weight loss during resting; WLRC = weight loss during curing; WLSC = weight loss during resting and curing; R^2v = R^2 in validation; RMSEv = RMSE in validation.
[1]All weight losses are expressed as percentages of raw ham initial weight after trimming
[2]Standard deviation of R^2v values ranged from 0.00 and 0.01; SD of RMSEv ranged from 0.01 to 0.08.

is much lower than the R^2v (67%) obtained with Bayesian regression.

Effectiveness of the predictions of ham weight losses for processing

Our best results indicate that the models are not expected to provide accurate predictions at the individual level, but they could potentially be used to discriminate between low and high values or to predict aggregate values (i.e., the average expected weight loss of a single batch of hams) with reasonable accuracy. Because the prediction error variance of the mean of a group of n samples is $\sqrt{n}$ times smaller than the prediction error variance of an individual sample (Heuer *et al.*, 2001), these predictions are expected to be accurate enough to be used as indicators for batches of production. In addition, it has to be noted that models were developed using pigs of a single genetic line raised under standardised rearing conditions. Such circumstances, which differ from ordinary productive conditions, affect unfavourably the variability in the training set used in model building and penalise the resulting prediction accuracy.

Prospects of use of ham weight losses predictions in a pig breeding program

Variance components and genetic parameters estimates obtained for WLC and P-WLC are reported in Table 5. The estimated h^2 for WLC was 0.27 ± 0.04, in agreement with a previous study performed on the same trait (Sturaro, 2004), whereas the estimate for P-WLC was 0.39 ± 0.03.

Table 5 *Estimated genetic parameters for measures of ham weight loss at the end of curing (WLC) and predictions of WLC (P-WLC) of hams from heavy pigs*

Parameter	Trait	Estimate	Standard error
Additive genetic variance (σ_a^2)	WLC	1.24	0.21
	P-WLC	0.96	0.08
Heritability (h^2)	WLC	0.27	0.04
	P-WLC	0.39	0.04
Additive genetic correlation (r_g)	WLC, P-WLC	0.88	0.03

In the literature, h^2 of infrared predictions has been reported to be both lower and higher than that of the measured trait, as observed in an extensive study comparing genetic parameters of actual and infrared-predicted traits (Bonfatti *et al.*, 2017b).

Given the unfeasibility of obtaining large-scale phenotypes for WLC, the use of indicator traits that are genetically correlated with the trait of concern is fundamental for animal breeding. In this scenario, P-WLC might be used as an indicator trait of WLC in pig selective initiatives, also as an alternative to WL at first salting. This is especially true when considering that even poorly predicted traits may exhibit a large genetic correlation with the breeding goal trait, as reported by Bonfatti *et al.* (2017b).

In this study, the estimated genetic correlation (r_g) between WLC and P-WLC was high and positive ($r_g = 0.88 \pm 0.03$), indicating that P-WLC can be successfully used as an indicator trait of WLC in a pig breeding program. The estimated genetic correlation between WLC and P-WLC was also higher than the genetic correlation detected between WL at first salting and WLC (0.65 to 0.77; Buttazzoni *et al.*, 1993; Sturaro, 2004). This suggests that P-WLC might be, not only more convenient but also more effective than WL at first salting as an indicator trait of WLC. Estimates of h^2 for WL at first salting ranged from 0.30 to 0.61 (Buttazzoni *et al.*, 1993; Carnier *et al.*, 1999; Associazione Nazionale Allevatori Suini, 2016).

Summarising, we investigated models for the prediction of WL of dry-cured hams. The detected correlation between measured and predicted ham WL is large enough to envisage a future use of the investigated prediction models in the provision of technological information to the pork industry. Such information can facilitate early identification of ham batches requiring an adaptation of the processing conditions. In addition, results encourage a future use of prediction models to provide large-scale phenotypes for WLC in pig selective breeding initiatives.

(Verona, Italy) in operations at the slaughter plant is also gratefully acknowledged.

🆔 V. Bonfatti 0000-0003-3970-5764

References

Associazione Nazionale Allevatori Suini 2016. ANAS Notizie: La Selezione per il Prosciutto DOP. Una Sfida Tecnica per un Orizzonte Dilungo Periodo. Retrieved on 8 July 2019 from http://www.anas.it/giornale/201600002.PDF.

Bonfatti V, Tiezzi F, Miglior F and Carnier P 2017a. Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations. Journal of Dairy Science 100, 7306–7319.

Bonfatti V, Vicario D, Lugo A and Carnier P 2017b. Genetic parameters of measures and population-wide infrared predictions of 92 traits describing the fine composition and technological properties of milk in Italian Simmental cattle. Journal of Dairy Science 100, 5526–5540.

Bosi P and Russo V 2004. The production of the heavy pig for high quality processed products. Italian Journal of Animal Science 3, 309–321.

Breiman L 2001. Random forests. Machine Learning 45, 5–32.

Buttazzoni L, Gallo M, Baiocco C and Carchedi G 1993. La selezione per la qualità della carne suina destinata alla trasformazione. Rivista di Suinicoltura 4, 139–145.

Čandek-Potokar M and Škrlep M 2012. Factors in pig production that impact the quality of dry-cured ham: a review. Animal 6, 327–338.

Carnier P, Cassandro M, Knol E and Padoan D 1999. Genetic parameters for some carcass and fresh ham traits of crossbred Goland pigs. In Recent progress in animal science (eds. G Piva, G Bertoni, F Masoero, P Bani and L Calamari), pp. 221–223. Franco Angeli, Milan, Italy.

Centro Ricerche Produzioni Animali 2010. Suinicoltura italiana e costo di produzione. Brochure No. 1. Centro Ricerche Produzioni Animali, Reggio Emilia, Italy.

Collell C, Gou P, Arnau J and Comaposada J 2011. Non-destructive estimation of moisture, water activity and NaCl at ham surface during resting and drying using NIR spectroscopy. Food Chemistry, 129, 601–607.

Consorzio del Prosciutto di Parma 1992. Prosciutto di Parma (Parma ham) protected designation of origin. Retrieved on 11 Dec 2019 from https://www.prosciuttodiparma.com/wp-content/uploads/2019/07/Parma_Ham_Specifications_Disciplinare_Consolidato_Nov_13.pdf.

Consorzio del prosciutto di San Daniele 2007. Production specification for the protected designation of origin "Prosciutto di San Daniele". Retrieved on 17 May 2019 from https://www.prosciuttosandaniele.it/content/uploads/2018/05/Production-Specifications-SAN-DANIELE.pdf.

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD and Calus MPL 2013. Whole genome regression and prediction methods applied to plant and animal breeding. Genetics 193, 327–345.

de los Campos G and Perez-Rodriguez P 2014. BGLR: Bayesian generalized linear regression. Version 1.0.5. Retrieved on 7 May 2019 from https://cran.r-project.org/web/packages/BGLR/index.html.

Gallo L, Padoan D, Bondesan V and Penzo N 1999. Effect of some carcass and fresh ham traits on curing loss of hams from crossbred Goland pigs. In Recent progress in animal science (eds. G Piva, G Bertoni, F Masoero, P Bani and L Calamari), pp. 689–691. Franco Angeli, Milan, Italy.

Groeneveld E, Kovac M and Mielenz N 2010. VCE user's guide and reference manual v. 6.0. Institute of Farm Animal Genetics, Mariensee, Germany.

Habier D, Fernando R, Kizilkaya K and Garrick D 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12, 186–197.

Heuer C, Luinge HJ, Lutz ETG, Schukken YH, van der Maas JH, Wilmink H and Noordhuizen JPTM 2001. Determination of acetone in cow milk by Fourier transform infrared spectroscopy for the detection of subclinical ketosis. Journal of Dairy Science 84, 575–582.

Ishwaran H and Kogalur UB 2018. Random Forests for Survival, Regression, and Classification (RF-SRC), R package version 2.7.0. Retrieved on 7 May 2019 from https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf.

Liaw A and Wiener M 2002. Classification and regression by random forest. R News 2, 18–22.

Pérez P and de los Campos G 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198, 483–495.

Prevolnik M, Andronikov D, Žlender B, Font-i-Furnols M, Novič M, Škorjanc D and Čandek-Potokar M 2014. Classification of dry-cured hams according to the maturation time using near infrared spectra and artificial neural networks. Meat Science 96, 14–20.

Prieto N, Pawluczyk O, Dugan MER and Aalhus JL 2017. A review of the principles and applications of near-infrared spectroscopy to characterized meat, fat, and meat products. Applied Spectroscopy 7, 1403–1426.

R Development Core Team 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rostellato R, Sartori C, Bonfatti V, Chiarot G and Carnier P 2015. Direct and social genetic effects on body weight at 270 days and carcass and ham quality traits in heavy pigs. Journal of Animal Science 93, 1–10.

Russo V, Davoli R and Nanni Costa L 1992. Recenti acquisizioni nel miglioramento genetico della qualità della carne suina. Rivista di suinicoltura 4, 47–55.

Russo V, Nanni Costa L, Lo Fiego DP and De Grossi A 1991. Early estimation of seasoning loss in Parma ham production. In Proceedings of the 37th International Congress of Meat Science and Technology, 1–6 September 1991, Kulmbach, Germany, pp. 926–929.

Sturaro E 2004. Caratteristiche tecnologiche e qualitative di cosce suine fresche destinate alla trasformazione in prodotti tipici stagionati: aspetti genetici e relazioni con la qualità del prodotto finale. PhD thesis, University of Padua, Padua, Italy.