Animal Science 2004, **78:** 409-418 1357-7298/04/34840409$20·00

The effectiveness of a visual image analysis (VIA) system for monitoring the performance of growing/finishing pigs

R. P. White[1]†, C. P. Schofield[1], D. M. Green[2], D. J. Parsons[1] and C. T. Whittemore[2]

[1]*BBSRC Silsoe Research Institute, Wrest Park, Silsoe, Bedford MK45 4HS, UK*
[2]*University of Edinburgh School of Geosciences, Agriculture Building, West Mains Road, Edinburgh EH9 3JG, UK*

† E-mail : Rodger.White@bbsrc.ac.uk

Abstract

A visual image analysis (VIA) system provided continuous, automatic collection of size and shape data for a total of 116 pigs slaughtered serially from 25 to 115 kg live weight. Males and females of three types of pigs ('Meishan' type, 'Pietrain' type, and 'Landrace' type) were selected to provide variation in both composition and conformation (the three types being, respectively, 'fat', 'blocky', and 'lean'). Results below are presented in this order. Regression analysis was used to relate VIA size to platform weigher (FIRE) measurements of live weight. Residual maximum likelihood (REML) analysis showed that at the observed growth rate, a change in pig state could be detected by VIA after 8, 9, and 10 days respectively for the three types, and by the platform weigher system after 12, 4, and 13 days (in both cases with a confidence of 95%). Artificial neural network and canonical variates analysis were used to test the ability of VIA to distinguish between pig types and sexes. With cross validation, the canonical variates analysis correctly classified the three types in 72, 83, and 64% of observations, and the neural network in 81, 81, and 64% of observations. The VIA system is considered to be a valuable monitoring system which may play a rôle in the construction of integrated management systems (IMS).

Keywords: *canonical analysis, growth, integrated management systems, pigs, visual image analysis.*

Introduction

Real-time measurement of performance and timely management response to that information are essential elements of management systems that have the objective of integrating output performance simultaneously with product quality, animal welfare and environmental protection imperatives. Such systems are described as integrated management systems (IMS, Whittemore *et al.*, 2001).

Visual image analysis (VIA) is a novel system, which enables the real-time automated measurement of the plan view image of a pig by video camera (Marchant *et al.*, 1999; Schofield *et al.*, 1999). VIA offers substantial potential benefits to both science and industry in the measurement of pig size (Whittemore and Schofield, 2000). It is the ability of VIA to measure in real time that enables the closure of the control loop in integrated management systems.

Weigh scales can also measure in real time, but are not used in such a way in practical pig production units. When placed within pens, weigh scales are prone to mechanical error, due to the necessary physical connection between animal and machine, and between machine and a dirty environment. A video camera placed above a pen is less vulnerable. When the weigh scale is placed outside of the pen, frequent disruptive pig movements are required. In most applied science investigations pigs are weighed no more frequently than at the beginning and end of the period in question, which may be the full period of test. In commercial practice pigs are weighed infrequently, sometimes not at all, and often no more than at the end of a period of test. Thus not only does VIA offer the possibility of measurement where none was previously utilized, but by provision of a continuous data stream offers mitigation of weigh-scale error due to single-point (unreplicated) measurement, and variation due to diurnal and gut and bladder content effects.

The advantages of real-time performance measurement are numerous. These include (a) improved monitoring of pig welfare; (b) characterization of growth response to nutrient supply within the context of specific pig and farm types; (c) comparison (and thereby optimization) of pig performance against target; (d) determination of limits to rate and quality of pig growth; and (e) allowance for changes in production and market environment during the growth period of the pigs.

VIA is operated at the level of individually identified pigs within a pen of pigs. For the application of scientific research and for the development of breeding, feeding, management and welfare programmes this level would be appropriate for all pigs involved in any study. For the application of best commercial production practice, VIA-equipped pig pens would operate as sentinel monitors for contemporaneous peer groups. The loss of precision in this case is directly related to the extent of the scaling-up involved, and the variation in performance between pigs within pens, pens of pigs within houses, and houses of pigs within farms.

Proof of the principle that VIA can predict animal weight and measure change in animal size over time is given by Schofield *et al.* (1999), Marchant *et al.* (1999), Doeschl *et al.* (2003) and White *et al.* (2003). For the establishment of the methodology in respect to the functions outlined above, the following information now needs to be secured:

(a) In the continuing likely use (for the immediate future) of mechanically determined weight as a determinant of animal size, VIA can indicate true pig live mass at a given moment in the course of a test period with accuracy sufficient to purpose. This should be with no greater error than a mechanical weigh scale used in the practical circumstances as described above. Further, that if the prediction of live weight by VIA is prone to influence by breed variation in pig shape, that influence is quantifiable such as to maintain the integrity of the prediction.

(b) VIA can indicate a real change in animal state (as measured by size), with at least equal accuracy as a mechanical weigh scale, and in a sufficiently short time-scale for the determination of size on a continuous basis to be worthwhile.

(c) VIA can indicate a real change in animal state (as measured by size), in a sufficiently short time-scale for the information to be usefully acted upon by management systems. For this a realistic time scale of one week may be proposed.

(d) VIA can give as useful a descriptor of animal growth over time and in response to nutrient input as does live weight.

(e) The VIA system is able to recognize differences in shape and to state in which dimensions the differences occur.

This paper addresses the above points, examining the performance of the VIA system and comparing it with an automatic weighing system comprising a platform balance. Canonical variates analysis and artificial neural networks are used to test the ability of VIA to distinguish between pig types, and REML analysis is used to determine the confidence limits on changes in pig state through time, to determine the time periods required to observe significant changes in pig size.

Material and methods

A total of 116 pigs of three mixed genetic origins were serially slaughtered from 25 to 115 kg live weight for the purpose of testing the VIA and IMS systems (Green *et al.*, 2003; Fisher *et al.*, 2003; Whittemore *et al.*, 2003). Sows from the BOCM Pauls Ltd Barrhill herd (JSR Genepacker 90 type) were inseminated with semen from selected boars chosen to produce offspring that could represent three distinct pig shapes. The boar types were 50% Meishan synthetic (for 'fat' pigs), 100% Pietrain (for 'blocky' pigs) and 100% pure Landrace (for 'lean' pigs). These three offspring types are henceforth referred to as 'Meishan', 'Pietrain', and 'Landrace' type pigs. All pigs were individually identified by electronic ID tag.

Diet and housing for the trial pigs was as described by Green *et al.* (2003). In brief, all pigs were given *ad libitum* a cereal and soya-bean-based diet, with a calculated digestible energy content of 14·5 MJ/kg and a lysine content of 11·4 g/kg. Pigs were housed in fully slatted pens in groups of up to 20 animals.

Pigs were housed in six pens, two pens per type, without mixing of types. Each of the six pens was equipped with an electronic feeding station with integrated platform balance (FIRE ® Feeder, Osborne Europe Ltd, Newcastle upon Tyne, UK) and a separate VIA system, installed above the feeding station. FIRE feeder equipment is normally employed for research and development (rather than practical production) purposes. It uses patented Osborne ACCU-ARM ® technology, claimed to measure live weight to within 2% in practice. Eleven separate area and length measurements are taken by the VIA software while the pig is in view of the

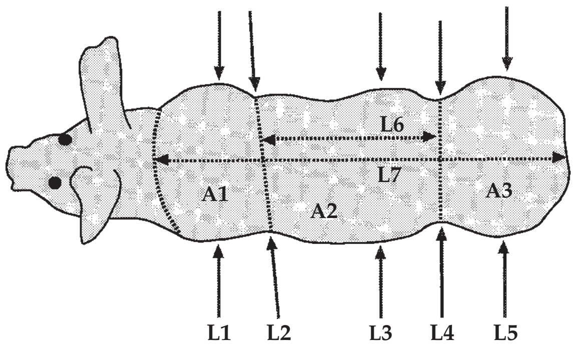

Figure 1 Length and area measurements produced by the VIA system.

imaging camera; these measurements are shown in Figure 1 and are described as follows:

L1 = width of shoulder at its widest point; L2 = width of body at the constriction between the shoulder and trunk; L3 = width of trunk at its widest point; L4 = width of body at the constriction between the trunk and pelvic limb; L5 = width of pelvic limb at its widest point; L6 = length of trunk between the L2 and L4 loci; L7 = total length of body, excluding head; A1 = plan area of shoulder, proximal to the L2 locus; A2 = plan area of trunk, between the L2 and L4 loci; A3 = plan area of pelvic limb, distal to the L4 locus; A4 = total plan area of body, excluding head (A1 + A2 + A3).

Below, all linear measurements are measured in cm, and all area measurements in cm^2. Of most importance is the total plan area of the upper surface of the pig, denoted A4, which previous studies found to be closely correlated to live weight (Schofield *et al.*, 1999; White *et al.*, 2003). Both VIA and FIRE systems provide data while the pig is present at the feeding station; daily medians were recorded by both systems to provide robust estimates of the size of the pig on a given day. These values were statistically analysed, as described below. Daily data did not allow analysis of diurnal variation.

Statistical analyses

Regression of VIA data against live weight. To determine the relationship between VIA A4 measurement and live weight as measured by the FIRE system, standard linear regression was used. All VIA records within a pig type were pooled and a simple linear model was used that did not include individual pig. Analyses for each pig type were carried out separately.

Regression of VIA data against time. For experiments conducted over time, the growth of a pig in terms of weight and VIA A4 is expected to be linear, and so the statistical model may be expected to be ordinary linear regression, where the dependence on time is through the slope of the regression line. However, this assumes that successive observations are independent, and so the correlation between them is zero. This basic assumption that observations are independent in time is clearly not true, and standard errors of estimates of slopes and intercepts determined in this way would be too small.

It is more realistic to assume that successive observations are not independent of each other, and so there should be a correlation between neighbouring observations, and it should be further assumed that the correlation decreases between observations taken further apart. An auto-regressive model is one in which the current observation depends on the previous set of observations; the order determines how far back that dependence extends, and so an order one implies dependence only on the previous measurement. An auto-regressive model of order one is equivalent to having

Table 1 *Experimental design of trial, showing the five slaughter batches, trial duration for each batch, and the numbers of animals used for the two sets of statistical analysis (where used, the number of males is indicated in brackets)*

	Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
Length of trial (days)					
'Meishan' type	13	34	48	69	104
'Pietrain' type	17	37	58	79	107
'Landrace' type	17	31	52	73	101
Number of animals in REML analysis					
'Meishan' type	-	-	-	5	12 (7)
'Pietrain' type	-	-	-	6	12 (7)
'Landrace' type	-	-	-	6	12 (6)
Number of animals in canonical variates and neural network analysis					
'Meishan' type	10 (5)	5	5	7	12 (7)
'Pietrain' type	10 (5)	5	5	6	12 (7)
'Landrace' type	11 (6)	5	5	6	12 (6)

a correlation structure depending on a power law with respect to distance or, in this case, time (Chatfield, 1996). The dependence on time for the pig's attribute (weight, VIA measurements, etc.) is expressed by the slope of the regression line and the correlation structure imposed on the observations. The correlation structure is assumed to be the same for pigs of a given pig type. This analysis is solved by weighted least squares. Compared with standard linear regression without the correlation structure, application of this model has two effects : the slope of the regression becomes slightly smaller and the standard error of the slope becomes larger. Because an extra parameter is estimated, the residual variance should be reduced if the parameter is useful.

The VIA data were analysed both by standard linear regression and by REML analysis with the correlation structure described above. The model chosen has a common slope *versus* time for all animals within a type, but separate intercepts for individual animals. This is in common with previously fitted models (Marchant *et al.*, 1999). To allow comparison between models, regressions are done for each pig type separately. The model is fitted using residual maximum likelihood (REML) in Genstat 5 Committee (1993). Because of problems with lack of convergence of the model algorithm, the model was fitted using weekly data throughout the growth period for which continuous data were available for all pigs.

For the REML analysis, the following statistical model is adopted for pigs of each type (Verbyla *et al.*, 1992).

$$\mathrm{VIA}_{ij} = \mu_i + bt_{ij} + \varepsilon_{ij}$$
$$\mathrm{Var}(\varepsilon_{ij}) = s^2$$
$$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{ij+1}) = s^2\rho^2$$
$$\mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{kj}) = 0 \qquad i \neq k.$$

Here, i indicates the individual pig, and t age; t_{ij} is age in days for an individual pig and VIA_{ij} any of the measurements. The model has a common slope, b, for all pigs of a given type, related to the age of the individual pig, but with separate intercepts (μ_i). In addition to the linear relationship with pig age for the VIA measurement, there is a covariance structure placed on residuals (ε_{ij}); successive residuals for individuals within a given pig type all have the same correlation, but there is no correlation between measurements on different individuals. (Note $\rho^2 = 0$ is equivalent to ordinary linear regression.)

For the standard error of a difference between a pair of measurements, the following equation was used for the VIA data:

$$s_{\delta t}^2 = 2s^2(1 + \gamma(1-\rho^2)).$$

The γ represents the scaled variance component due to the differences between the individual animals. In certain cases, this may be simplified, in particular when the pigs are all observed at each time point. For FIRE weights, the data were expanded to give a 'complete' set for the factors, which leads to a slightly different form of formula for calculating the standard error of a difference (without a γ term):

$$s_{\delta t}^2 = 2s^2(1-\rho^2).$$

As described earlier, a power law of the form $\rho_t = \rho_{(7)}{}^t$ was used to adjust the $\rho_{(7)}$ correlation coefficient obtained for the analysis of weekly data to other time differences, t.

Discriminating between pig types with VIA data. The canonical variates analysis and artificial neural network analysis for classifying records of VIA data according to pig type are similar in that they both look for combinations of variables that distinguish between groups. For both analyses, a total of 5634 records of the 11 VIA measurements across 116 pigs of three pig types and two sexes (Table 1) were available. These were standardized to remove scale effects by dividing the three partial areas (A1 to A3) by A4 and the seven lengths (L1 to L7) by $A4^{1/2}$. This gives a set of 10 dimensionless VIA measurements, since the scaling measurement has been factored out. Because pigs were serially slaughtered, the number of pigs reduced gradually through to the end of the trial.

For both analyses, individual records were assigned to either a 'training' set or a 'test' set. The training set was used to construct a model, while the test set was used to measure how good the model was at discriminating between the different groups. Two approaches are adopted here to partition the data into training and prediction sets. The first uses a random allocation of the set of records from the complete set, without reference to the identity of the pig, while the second uses either even or odd pig identity numbers. This method increased the independence of the two sets by preventing similar records for the same pig appearing in both sets. Where random records were used, 1000, 2000, or 3000 records from the complete set were chosen for the training set. The number of records in the even and odd identity training sets were respectively 2703 and 2940 records.

Canonical variates analysis (CVA) is a technique for calculating an optimal reduced dimensional space to visualize groupings of data (Mardia *et al.*, 1979). The linear combinations of variables define 'new' dimensions in a reduced dimensional space that best explains differences between the groups on the basis of maximizing the variance between the groups while minimizing the variance within the groups. The analysis is based on the between-groups and within-groups sums of squares and the cross-products matrix, so the rank (dimensionality) of the problem is the number of groups minus one. CVA can also be used to assign new objects to the *a priori* groups based on minimum (Mahalanobis) distance of the object in reduced dimensional space from the group centroids.

For artificial neural network analysis the data set was then partitioned into training and test sets according to the ID number parity : even for the training set, odd for the test set. The resulting test set contained 611 records of 'Meishan' type, 1134 of 'Pietrain' type, and 954 of 'Landrace' type pigs.

Three neural network configurations were investigated:
(a) Multi-layer perceptron (MLP) with one hidden layer;
(b) Multi-layer perceptron with two hidden layers;
(c) Radial basis function network (RBF).
These were constructed and trained using Trajan software (http : //www. trajan-software. demon. co. uk/).

The MLP is probably the most popular architecture in use. When used for classification its behaviour is a series of linear discriminants, dividing the space by hyperplanes. With a single hidden layer, the result in the output space may be visualized as a plateau with a convex hull. With two hidden layers, more complex outputs are possible, indeed it is theoretically sufficient to model any problem. The number of nodes required in each layer depends on the problem being modelled; a common heuristic is to use half the number of inputs (10) plus half the number of outputs (three), i.e. seven nodes. A network with too few nodes may be unable to classify the data; using too many can lead to overtraining (the equivalent of over fitting a conventional statistical model). The effect of the number of nodes was investigated by starting with four per layer and doubling this until the mean classification rate (proportion of type x classified as type x) stopped increasing.

In contrast, the RBF network partitions the space by hyperspheres, the result of a single unit is a multi-dimensional Gaussian. It is thus able to represent non-linear functions with a single hidden layer. The relative suitability of RBF and MLP networks depends on the data being modelled. RBFs generally require more nodes in the hidden layer than MLPs to produce a smooth output function. The same approach as for the MLP was used, starting with eight nodes.

Table 2 *Regression slopes* (b) *for models of VIA A4* versus *time* (cm^2*/day*) *and FIRE weight* (*kg/day*) *with standard errors in brackets for both standard regression and REML analysis of the three pig types*

	'Meishan' type	'Pietrain' type	'Landrace' type
VIA A4			
Regression	15·2 (0·27)	15·1 (0·26)	15·1 (0·33)
REML	14·8 (0·56)	14·7 (0·49)	14·8 (0·52)
FIRE weight			
Regression	0·87 (0·010)	0·86 (0·010)	0·81 (0·012)
REML	0·83 (0·024)	0·83 (0·024)	0·78 (0·027)

Each network was trained using the training set data, as described above, then applied to the test set. For the MLPs, the training, which optimizes the parameters of the model, used two optimization algorithms in sequence. First the standard back-propagation algorithm was applied for 100 steps to obtain an approximate solution. This was followed by a conjugate gradient method to improve the fit until the error function was seen to reach a lower plateau, which generally required fewer than 20 steps. The structure of RBFs means they can be trained rapidly using linear modelling methods. In each case the objective was to minimize the sum of squares of the classification errors.

Results

Regression of VIA data against live weight

The regression model gave the following equations to express live weight (W) as measured by FIRE weight platforms as a function of VIA A4 measurements here in m^2. Parameter standard errors are given in brackets.

'Meishan' type : W = –32·8 (0·63) + 516 (4·6) × A4.
'Pietrain' type : W = –38·5 (0·67) + 516 (3·9) × A4.
'Landrace' type : W = –29·8 (0·74) + 489 (4·7) × A4.

Table 3 *Residual variances* (s^2), *and weekly serial correlations* ($\rho^2_{(7)}$), *and scale factors* (γ) *for REML analysis of VIA A4* (cm^2) versus *time* (*day*) *for the three pig types*

	'Meishan' type	'Pietrain' type	'Landrace' type
$\rho^2_{(7)}$	0·962	0·836	0·898
s^2	614	1050	2620
γ	65·2	9·84	3·75

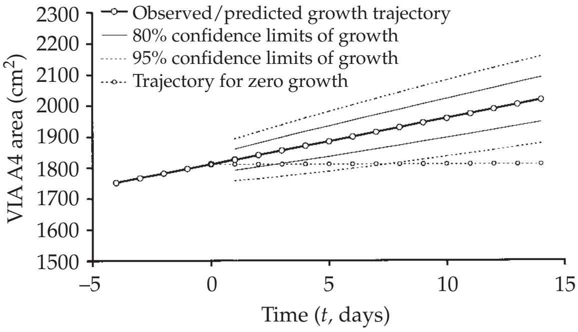

Figure 2 Observed past (and predicted future) growth trajectory, relative to day *t*, for VIA A4. 80 and 95% confidence intervals (one-sided, equivalent to 60 and 90% two-sided confidence intervals) of predicted growth difference are shown. A significant positive growth increment is observed where the confidence band does not intersect with the zero difference (horizontal) trajectory. Growth trajectory for 'Meishan' type pigs is shown.

The model had an R^2 of 0·905 and residual s.d. of 6·2 kg. A separate equation was required for each pig type, indicating differences between the types in the weight *versus* size relationship.

Regression of VIA data against time

REML analysis and standard linear regression provided growth rates for FIRE weight and VIA A4 data, together with their standard errors, shown in Table 2. Growth rates of A4 were found to be similar and not significantly different between pig types. It is noted that the REML analysis provides a lower estimate of slope than standard regression. For FIRE weight, the 'Landrace' type pigs were found to have a significantly lower growth rate than the other two types ($P < 0{\cdot}05$), whose growth rates did not differ. Associated with the REML analyses are the estimates of residual variances (s^2), weekly serial correlations ($\rho^2_{(7)}$), and scale factors (γ), shown for A4 data in Table 3 and for FIRE data in Table 4. From the results shown in Tables 2, 3 and 4, the standard error of the difference in FIRE weight or VIA A4 between any 2 days along the slope can be calculated. Unlike standard linear regression, where the standard error of the differences is a constant, in the REML analysis, the standard error increases with increasing separation, but at an ever-decreasing rate. An

Table 4 *Residual variances* (s^2), *and weekly serial correlations* ($\rho^2_{(7)}$) *for REML analysis of FIRE weight* (*kg*) versus *time* (*day*) *for the three pig types*

	'Meishan' type	'Pietrain' type	'Landrace' type
$\rho^2_{(7)}$	0·9806	0·9408	0·9831
s^2	511·4	57·97	568·3

Table 5 *Number of days elapsed until expected change in pig size becomes significant, for FIRE and A4 data, assuming confidence limits of 80% and 95%* (*one-sided*) (*expectation for the three pig types is shown*)

		'Meishan' type	'Pietrain' type	'Landrace' type
FIRE weight	80%	3	2	4
	95%	12	4	13
A4	80%	3	4	5
	95%	8	9	10

example set of predictions, with their 80 and 95% confidence intervals is shown for the 'Meishan' type pigs in Figure 2. Note that strictly the confidence bands shown on this figure pertain not to the forward predictions of A4 (or FIRE weight), but to the difference between each A4 forward prediction and the value of A4 with which the prediction is compared.

Where the difference between two measurements of A4 exceeds the confidence limits of that difference, the difference is significantly different from zero. The VIA system can therefore detect a change in the status of the pig over such a time period. The time period required to detect such a change in A4 or FIRE weight for the three pig types is shown in Table 5. Similar time periods of the order of a week are necessary to detect changes in both FIRE weight and VIA A4, depending on the level of statistical significance examined. The residual variance (s^2) cannot be used as an absolute measure of 'goodness of fit' but can be used comparatively. Comparing the residual variances in Table 4 for FIRE data, the statistical model fitted the 'Pietrain' type data better than the other two types. As a consequence, the interval for detecting a significant change in weight is much smaller for the 'Pietrain' type.

Table 6 *Confusion matrix for type only* (*as percentages*) *for neural network analyses using training set of even ID animals and test set of odd ID animals*

Predicted type	Observed type		
	'Meishan'	'Pietrain'	'Landrace'
RBF-one hidden layer			
'Meishan'	81	9	24
'Pietrain'	6	81	12
'Landrace'	12	10	64
MLP-one hidden layer			
'Meishan'	78	8	21
'Pietrain'	8	82	15
'Landrace'	15	10	63

Table 7 *Confusion matrix for type only (as percentages) for canonical variates analysis using training set of even ID animals and test set of odd ID animals*

Predicted type	Observed type		
	'Meishan'	'Pietrain'	'Landrace'
'Meishan'	72·5	4·2	17·1
'Pietrain'	7·7	83·0	18·7
'Landrace'	19·8	12·8	64·3

For comparison with the above, in the standard regression analysis, without serial correlation structure, the corresponding variance of the difference between any pair of measurements can be calculated as $2s^2$, where s^2 is the residual variance. This is because observations are assumed to be independent and have the same variance. The standard errors of the differences for A4 here are 119, 108, and 105 cm^2 for 'Meishan', 'Pietrain', and 'Landrace' pigs respectively, giving times to detect a significant difference of 13, 12, and 12 days respectively (using a one-sided test, 95% confidence). Similarly, for FIRE weights, the standard errors of the differences were 4·91, 4·68, and 4·24 kg, respectively, giving times to detection of a statistically significant difference of 10, 9, and 9 days.

Discriminating between pig types with VIA data
The best versions of the three designs of neural network gave similar classification rates. Hence, only the results from the best-fitting network, the RBF, plus, for comparison, the single-hidden layer MLP, are presented (Table 6). Results presented below pertain to the RBF network, and are presented as percentages in a 'confusion matrix' showing how the pigs of each type have been classified by the network. A perfect classifier would have scores of 100% down the leading diagonal. The scores shown are derived from the test data set only; the scores for the training set were close to those for the test set in every case, showing that it is a representative sample of the whole data set. Both the MLPs performed best with 16 nodes per layer, although with two layers the improvement from 8 to 16 was marginal. The best RBF had 64 nodes. The mean classification rates of the one-layer MLP, two-layer MLP and RBF networks were 74%, 74% and 75% respectively.

The neural network (Table 6) and canonical variates analysis (Table 7) behave similarly in the classification of pig type from VIA records. Classification by both was reliable, with about 75% VIA records correctly classified (compared with 33% by chance alone). Misclassifications for 'Meishan' type pigs were biased towards 'Landrace' type in

Table 8 *Confusion matrix for sex (female (F) and male (M)) and type (as percentages) for canonical variates analysis using training set of even ID animals and test set of odd ID animals (correctly classified observations are located across the diagonal)*

Predicted type	Observed type					
	F 'Meishan'	M 'Meishan'	F 'Pietrain'	M 'Pietrain'	F 'Landrace'	M 'Landrace'
F 'Meishan'	69·5	31·3	1·7	2·4	13·6	5·4
M 'Meishan'	8·6	35·0	6·7	1·1	9·9	14·0
F 'Pietrain'	6·9	1·8	46·5	32·1	3·9	8·0
M 'Pietrain'	1·3	5·1	31·3	47·9	3·9	20·2
F 'Landrace'	10·4	18·9	2·6	5·1	47·7	13·4
M 'Landrace'	3·3	7·8	11·3	11·5	21·0	39·0

Table 9 *Confusion matrix for sex (female (F) and male (M)) and type (as percentages) for canonical variates analysis using 3000 record training set and remaining records as test set*

Predicted type	Observed type					
	F 'Meishan'	M 'Meishan'	F 'Pietrain'	M 'Pietrain'	F 'Landrace'	M 'Landrace'
F 'Meishan'	71·5	20·8	3·3	3·0	13·9	5·4
M 'Meishan'	11·7	39·5	5·4	1·2	10·5	12·4
F 'Pietrain'	3·3	5·9	48·9	28·0	3·6	9·4
M 'Pietrain'	3·3	10·4	24·6	54·3	5·7	13·1
F 'Landrace'	7·1	17·2	2·8	3·4	41·7	15·8
M 'Landrace'	3·0	6·2	15·1	10·1	24·7	44·0

both analyses. 'Pietrain' type pigs were misclassified almost uniformly by the neural network, though bias was towards the 'Landrace' type in the canonical variates analysis. Conversely, 'Landrace' type pigs were misclassified almost uniformly by the canonical variates analysis, though neural networks had a clear bias towards misclassification as 'Meishan'. In general, misclassifications between 'Pietrain' and 'Meishan' types were comparatively rare, with misclassifications in other directions considerably more common. The highest rate of misclassification was found in the neural network, where 24% of 'Landrace' type records were misclassified as 'Meishan'.

Table 8 shows the results of canonical variates analysis where the pigs are divided into six groups according to both type and sex. Overall classification of pigs according to type is achieved with similar success to the analysis without sex, shown in Table 7. However, classification according to sex is achieved with much less success, with correct classification in about 64% of VIA records (compared with 50% by chance). Only in two cases does the correct classification rate differ markedly from 50%: with 'Meishan' type females (87%) and 'Landrace' type males (73%). Classification according to sex had a conflicting effect on classification according to type. Fewer male 'Meishan' type VIA records were correctly classified as 'Meishan' than female (66 *versus* 78%), and similarly fewer male than female 'Landrace' type VIA records were correctly classified as 'Landrace' (52 *versus* 69%).

Alternatively to dividing the individual pigs into training and test sets, the individual VIA records can be divided into training and test sets. This does not account for pig ID, therefore records from every pig are likely to be found in the training set, and the test set does not contain novel pigs. Results from this analysis are shown in Table 9 using a 3000 record training set, classifying according to both type and sex. The classification rates from this analysis are similar to those obtained in Table 8 with similar overall correct classification rates for both type and sex. Similarly, the number of records used for the training set (1000, 2000 or 3000) had little effect on classification rates. As with the comparison of classification rates of the test and training sets in the neural network analysis, this would suggest that the pigs used for the training set used in Table 8 were representative of the whole population.

For these data, the between-groups sums of squares and cross products matrix is similar to the within-groups sums of squares and cross products matrix because the trace (sum of diagonal elements or

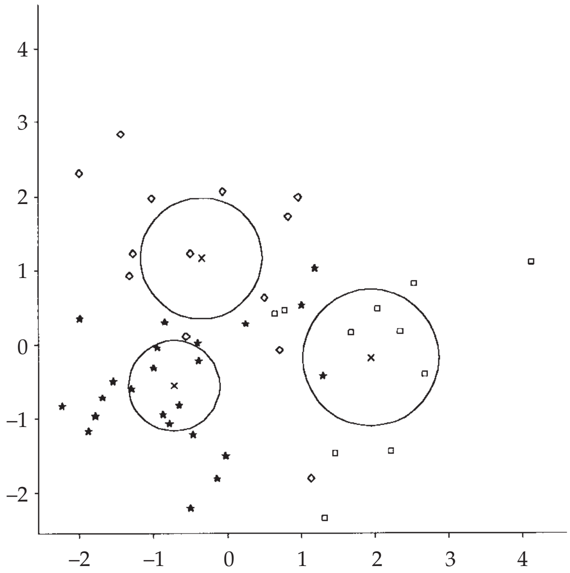

Figure 3 Plot of first two canonical variates for normalized VIA data (first on x-axis). The sizes of the confidence circles for this canonical variates analysis depend on the number of cases selected for each strain: 21 'Meishan' (★), 10 'Pietrain' (□) and 14 'Landrace' type (◇). The 45 cases were selected at random from the set of pigs used as the training set in Table 7. The centroids of the groups are marked by ×.

eigenvalues of the matrix) is approximately 1·4. This implies that the overall scatter between groups is similar to the scatter within groups; this is much more apparent when looking at the scatter plots in the canonical variates sub-space for the individual records. It accounts for the poorer classifications for the 'Landrace' and 'Meishan' types : all pigs are pig-shaped.

For the three types, the first canonical variate accounts for 70% of the trace and the second for the remaining 30%. This is the case regardless of the method of selection of the training set. From Figure 3, it is clear that the first canonical variate discriminates between the 'Pietrain' type and the 'Meishan' and 'Landrace' types, and that the second canonical variate further distinguishes between the 'Meishan' and 'Landrace' types.

Discussion

Regression of VIA data against live weight

The regression equations presented indicate that different calibration equations are required for different pig types in order to obtain the best estimate of live weight using VIA A4 area. Some small, but systematic error is therefore expected for individual animals, depending upon the exact genotype and posture of the individual; if a pig consistently presents itself to the camera in an unusual posture, this can be problematic and produce biased measurements of weight. The quoted residual s.d. is fairly high at 6·2 kg. However, A4 area has here been correlated against FIRE, which is itself not measured without error. For comparison, the residual standard error of the equation describing FIRE weight as a function of time (Table 2) was about 3 kg. These results suggest VIA A4 measurement to be an accurate reflector of pig live weight.

Regression of VIA data against time

It is apparent that VIA may be used to determine the rate of growth in pigs with an accuracy equal to that of a weigh scale. Though no difference was found in area growth rates by regression and REML analysis, the 'Landrace' type pigs were found to have a lower growth rate in terms of weight than the other two types. This finding is similar to the findings of Green *et al.* (2003), whose authors used the same three types of pigs (though the quantity of data selected for study and the statistical models used varied).

The elapsed days between an initial determination of size or weight and a meaningful identification of subsequent gain appears to be about 10 days in the case of both VIA and weigh scale measurements at 95% confidence, and less than half of that if confidence is reduced to 80%. This indicates the VIA system to be equally as accurate as a weigh scale for purposes of the management of growing pigs.

Discriminating between pig types with VIA data

The preferred neural network configuration (RBF with 64 units) found VIA to classify pigs correctly into the three types with an overall classification rate of 75%. Further examination of the neural network results showed that, while most pigs are misclassified occasionally (on a few days), some are systematically misclassified. For practical use, where data on each pig will be gathered for many days, occasional errors are not a problem. Where a pig is consistently misclassified as one of the other types, it is probable that its shape is more typical of that type, which again provides useful information. For example, one 'Landrace' type male was classified to types 'Landrace', 'Meishan' and 'Pietrain' 22, 30 and 32 times respectively. In a *post hoc* study it is not possible to examine in detail the reasons for this, though there seems to be an age effect : in the first 15 days there were ten correct classifications.

'Pietrain' type pigs were consistently found to be more rarely misclassified. Likewise the other two types were rarely misclassified as 'Pietrain' type. This most likely results from the greater similarity of the genotypes of the 'Meishan' and 'Landrace' pigs; each comprised mostly of alleles from Landrace/ Large White type pigs, whereas the 'Pietrain' type pigs were 50% pure Pietrain. Chemical composition, tissue composition, and carcass conformation of these pigs have been studied by Fisher *et al.* (2003) and Whittemore *et al.* (2003). Fisher *et al.* found the largest relative difference in carcass shape to be shoulder width, which was largest in the case of the 'Pietrain' type pigs, with the 'Meishan' and 'Landrace' type pigs more similar. The 'Pietrain' type pigs were also considerably less variable in chemical composition, with variation in fatness much less pronounced. Lipid masses at given weights in the other two types were similar, but considerably higher and more variable than those of the 'Pietrain' type.

It is not unexpected that the VIA system largely failed to distinguish between sexes of pig. These pigs were monitored only as far as 115 kg live weight, by which weight sex differences would be small. Pigs of the Meishan breed are early maturing; interestingly the highest rate of correct classification of sex was found in the case of the 'Meishan' type females. It is expected that higher rates of correct classification of sex would be found with older animals, where sex differences develop.

The three types of pigs described above were selected with the view to generating a population of pigs with variation in composition and conformation, for the purpose of exercising the VIA system and measuring its potential. Nevertheless, these types are within themselves highly variable; so it is to be expected that the ability of the VIA system to identify pigs of particular *proportions*, rather than pigs of a particular *type*, would be considerably higher than the rates of correct classification outlined above.

Conclusions

The above analysis has shown that visual image analysis (VIA) can serve as an accurate means of reflecting pig live mass. Further, VIA can track a change in pig size over time periods that are sufficiently short to allow commercial use. This time period is similar to that from platform weigh stations. In addition, artificial neural network and canonical variates models can distinguish between types of pigs that differ in shape on the basis of VIA data with a high rate of correct classification. Such VIA systems will therefore prove valuable in the development of integrated management systems to manage pig growth, welfare, and nutrition.

Acknowledgements

This work is part of the UK DEFRA (formerly MAFF) LINK Sustainable Livestock Production Programme LK 0614 Integrated Management Systems for Pig Nutrition Control and Pollution Reduction. The authors acknowledge the support of the sponsors : UK Department for Environment, Food and Rural Affairs (formerly Ministry of Agriculture, Fisheries and Food), UK Meat and Livestock Commission; and the industrial partners BOCM PAULS Ltd, Pig Improvement Company (UK) Ltd and Osborne (Europe) Ltd.

References

Chatfield, C. 1996. *The analysis of time series: an introduction*. Chapman and Hall, London.

Doeschl, A. B., Whittemore, C. T., Green, D. M., Fisher, A. V. and Schofield, C. P. 2003**.** Use of visual image analysis for the description of pig growth in size and shape. *Proceedings of the British Society of Animal Science, 2003*, p. 21 (abstr.).

Fisher, A. V., Green, D. M., Whittemore, C. T., Wood, J. D. and Schofield, C. P. 2003. Growth of carcass components and its relations with conformation in pigs of three types. *Meat Science* **65:** 639-650.

Genstat 5 Committee. 1993. *Genstat 5 release 3 reference manual*. Oxford University Press, Oxford.

Green, D. M., Brotherstone, S., Schofield, C. P. and Whittemore, C. T. 2003. Food intake and live growth performance of pigs measured automatically and continuously from 25 to 115 kg live weight. *Journal of the Science of Food and Agriculture* **83:** 1150-1155.

Marchant, J. A., Schofield, C. P. and White, R. P. 1999. Pig growth and conformation monitoring using image analysis. *Animal Science* **68:** 141-150.

Mardia, K. V., Kent, J. T. and Bibby, J. M. 1979. *Multivariate analysis.* Academic Press, London.

Schofield, C. P., Marchant, J. A., White, R. P., Brandl, N. and Wilson, M. 1999. Monitoring pig growth using a prototype imaging system. *Journal of Agricultural Engineering Research* **72:** 205-210.

Verbyla, A. P. and Cullis, B. R. 1992. The analysis of multistratum and spatially correlated repeated measures data. *Biometrics* **48:** 1015-1032.

White, R. P., Parsons, D. J., Schofield, C. P., Green, D. M. and Whittemore, C. T. 2003. Use of visual image analysis for the management of pig growth in size and shape. *Proceedings of the British Society of Animal Science, 2003*, p. 101 (abstr.).

Whittemore, C. T., Green, D. M. and Schofield, C. P. 2001. Nutrition management of growing pigs. In *Integrated management systems for livestock* (ed. C. M. Wathes, A. R. Frost, F. Gordon and J. D. Wood) *British Society of Animal Science occasional publication no. 28*, pp. 89-95.

Whittemore, C. T., Green, D. M., Wood, J. D., Fisher, A. V. and Schofield, C. P. 2003. Physical and chemical composition of the carcass of three different types of pigs grown from 25 to 115 kg live weight. *Animal Science* **77:** 235-245.

Whittemore, C. T. and Schofield, C. P. 2000. A case for size and shape scaling for understanding nutrient use in breeding sows and growing pigs. *Livestock Production Science* **65:** 203-208.

(*Received* 10 *July* 2003—*Accepted* 3 *December* 2003)