Variogram investigation of covariance shape within longitudinal data with possible use of a krigeage technique as an interpolation tool: Sheep growth data as an example

A. Chalh[†] and M. El Gazzah

Université de Tunis El Manar, Faculté des Sciences de Tunis, Département de Biologie,
Campus Universitaire El Manar, Le Belvédère, 2092, Tunis, Tunisia

Most quantitative traits considered in livestock evolve over time and several continuous functions have been proposed to model this change. For individual records (longitudinal data), it is evident that measures taken at close dates are generally more related than these further apart in time. Since milk production involves several parities, the covariance structure within this trait has been analysed by time series methodology. However, the covariance structure within traits that are not repeated during life, such as those linked to growth, has not yet been formally modelled by considering time lags as is done in time series analysis. We propose an adaptation of the variogram concept to shape this structure; which gives the possibility of kriging missing data at any particular time. A new parameter, the halftime variogram, has been proposed to characterise the growing potential of a given population. The weight records of a Barbarine male lamb population were used to illustrate the methodology. The variogram covering the whole growth process in this population could be modelled by a logistic equation. To estimate the missing data from birth to 105 days of age, a simple linear interpolation was sufficient since kriging on a linear model basis gives a relatively more accurate estimation than kriging on a logistic model basis. Nevertheless, when both known records around the missing data are distant, a krigeage on the basis of the logistic model provides a more accurate estimation.

Keywords: covariogram; growth modelling; interpolation; krigeage; sheep; variogram

---

†Corresponding author: Abdellah Chalh; E-mail: A.Chalh@fst.rnu.tn

Introduction

When repeated measurements of a trait are made, a data structure that is commonly called longitudinal data is generated. In livestock, most traits of economic importance have this property, such as milk and body weight. To estimate missing data, an improvement in accuracy may be achieved by accounting for covariances between records.

For example, to model the lactation curve in dairy livestock, the most widely used function is the Wood's function (Wood 1967; Pollot and Gootwine 2000; Chang *et al.* 2001). The Wilmink function (Wilmink 1987) is also in common use (Flores, Kinghorn and van der Werf 2013). Further mathematical models were proposed to overcome some specific problems like absence of a lactation peak (Morant and Gnanasakthy 1989; Cappio-Borlino, Pulina and Rossi 1995). However, some limitations of this approach have been surmounted by applying the mixed linear modelling concept directly on test day records to adjust curve parameters for environmental factors influencing individual lactation (Fernández, Sánchez and Garcés 2002; Macciotta *et al.* 2005; Leclerc *et al.* 2008).

Mixed linear model methodology was used to model directly the covariance structure among repeated measures (Littell, Pendergast and Natarajan 2000). It is known that measures closer in time are generally more correlated than measures further apart in time (Wang and Goonewardene 2004). This time-dependent relationship is usually modelled by a lag-dependant covariance function in the form of a first order autoregressive procedure. Since each animal may have production records during several parities, the variances and covariances among milk production data have been accounted for by time series analysis models (Macciotta, Cappio-Borlino and Pulina 2000) which were initially developed for econometric problems (Chatfield 1996). Nevertheless, to be able to fit such models, a sufficiently large data set is required. This condition seems to exclude the use of time series methodology for the analysis of data derived from dairy livestock recording schemes (five or six records). Fortunately, repeated parities have enabled the construction of series with an index variable that does not have an immediate reference to time. Records are ordered by lactation and, within lactation, by distance from birth (Macciotta *et al.* 2000). The autocorrelation function represented in a lag-plot form highlighted a periodic component of the variability and an autoregressive component of the first order that allowed modelling of variability by an autoregressive moving average model (Littell, Henry and Ammermann 1998).

Similar to milk production, the simplest mathematical method to analyse weight records is to fit a continuous time function which is differentiable in the whole interval of time covering the growth period. It is biologically known that an animal follows a predefined process as it grows towards its maturity weight (Owens, Dubeski and Hanson 1993). Therefore, a family of non-linear growth equations, like Brody, Richards, Von Bertalanffy, Gompertz, Exponential and Logistic, has been proposed in order to model the shape of the growth curve in livestock (Gous *et al.* 1999; Lambe *et al.* 2006; Abegaz, Van Wyk and Olivier 2010; da Silva *et al.* 2012). In developing these equations, it is taken into account that weight tends to an asymptotic value with time (maturity weight), that growth rate has a maximum at an intermediate time (a null second derivative), and that the relative growth rate decreases as

weight increases towards maturity (Lewis and Brotherstone 2002). The Gompertz model is an easy function that meets these requirements with only three parameters, each one having a biological sense. It is the most common function in use for different species (Kratochvilovà *et al.* 2002; Lewis and Brotherstone 2002; Akbas *et al.* 2006; Lambe *et al.* 2006). Recently, da Silva *et al.* (2012) fitted five models to the growth curve of Santa Inês breed of sheep. They reported that the Richards model presented some problems to achieve convergence. Regarding their results, the Logistic fit seems to be the best in describing the growth pattern of the breed studied. Furthermore, Pittroff *et al.* (2008) identified the Logistic function as the best model to fit to the growth curve of two sheep breeds. For Lambe *et al.* (2006), both Gompertz and Richards models produced the best fit and were not significantly different from each other.

For growth data, the covariance function has been assessed implicitly by using random regression methodology implemented with polynomial and/or nonlinear models (Lewis and Brotherstone 2002; Alvarez *et al.* 2008; Gbangboche *et al.* 2008; Cai, Wu and Dekkers 2011). It has been shown that phenotypic and genetic correlations between the estimated parameters of these models are high and generally negative (Kratochvilovà *et al.* 2002; Lambe *et al.* 2006; Koivula *et al.* 2008).

Unfortunately, the covariance structure within longitudinal data, mainly those that are not repeated during the lifetime (like milk), is not yet formally described by the time series paradigm. However, a number of techniques exist which are applied in the linear geostatistics field that may be adapted to longitudinal data to provide a detailed analysis of the covariance behaviour in time. The variogram concept, as well as the krigeage technique were formalised by Matheron (1971) in his theory of regionalised variables from previous studies of Krige (1951 and 1966). These techniques are widely applied in many fields like geology (Burgess and Webster 1980; Cressie 1986), hydrology (Kitandis 1983), petrology (Jaquet 1989), forestry (Uuttera *et al.* 1998), meteorology (Marzban and Sandgathe 2009), agronomy (Nansen 2012), epidemiology (Yates *et al.* 1986; Carrat and Valleron 1992; Lai 2004) and also in molecular biology (Diaz *et al.* 1997). The main problem that will be encountered in trying to apply these techniques to most longitudinal data remains the fact that both the mean and the variance should be considered as stationary. Taking weight data as an example, the more the average weight value increases over time more the variance value increases. In the present work, we propose an adaptation of the variogram concept to enable analysis of autocorrelation within longitudinal data. A new interpolation approach, in a krigeage form, is also proposed to estimate missing data at any desired time or at a standard age. An application to actual sheep growth data is used as an illustration.

Materials and Methods

Statistical formulation

In geostatistics, it is generally considered that: "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Using this law, let us consider $Z(t)$ a continuous variable evolving in time, so that: $Z(t+h)$ is greater than $Z(t)$; h is a defined period (which is equivalent to a "lag" in the time series methodology). The classical form of the variogram $\gamma(h)$ (or semi-variogram) would be as follows:

$$\gamma(h) = \frac{\sum_{i=1}^{N(h)} (Z(t_i + h) - Z(t_i))^2}{2N(h)}. \qquad (1)$$

With $N(h)$: the number of observations separated by h.

Theoretically, the variogram is defined by:

$$\gamma(h) = \frac{1}{2} \text{var}[Z(t+h) - Z(t)].$$

That can be developed to be:

$$\gamma(h) = \frac{1}{2} E([Z(t+h) - Z(t)] - E[Z(t+h) - Z(t)])^2. \qquad (2)$$

If we consider that: $E[Z(t+h)] = E[Z(t)]$, the first order moment stationarity hypothesis that is always adopted in geostatistics, the equation (2) leads to:

$$\gamma(h) = \frac{1}{2} \text{var}(Z(t+h))$$
$$+ \frac{1}{2} \text{var}(Z(t)) - \text{cov}(Z(t+h); Z(t)).$$

Let us assume a second order stationarity, so that the variances depend only on h [i.e., $\text{var}(Z(t+h)) = \text{var}(Z(t))$]. The last equation can be rewritten as:

$$\gamma(h) = \text{var}(Z(t)) - C(h). \qquad (3)$$

$C(h)$ is called the covariogram of Z which is a measure of a process resulting from a regionalised variable that deploys in time. This variable is simply a function $Z(t, t+h)$. Hence, for a given random function, it can admit a covariance function only if it has a finite variance that corresponds to the variogram's sill (generally denoted C), which is the limit of the variogram when h tends towards infinity [i.e., $\lim_{h \to \infty} \gamma(h) = \text{var}(Z(t)) = C$].

From the experimental data, one can represent an experimental variogram by using equation (1). This will allow the determination of: the sill (C), the range (denoted R) that corresponds to the extent of a period beyond which the measured weights become "unrelated" (R should correspond to the maturity age) and a nugget effect (denoted C_0) if any, that represents a micro-scale variation in measurement error [i.e., $\lim_{h \to 0} \gamma(h) = C_0$], and finally the theoretical model to be adopted for computing $\gamma(h)$. From equation (3) it is possible to easily compute $C(h)$ for kriging.

If the variable mean μ is unknown, it can be estimated using the ordinary krigeage, simultaneously with the unknown coefficients λ_i of a $Z(t_i)$ set. Thus, let us consider k known values of Z measured around an unknown quantity at a point t. An estimated value of $Z(t)$ may be obtained by:

$$Z(t) = \sum_{i=1}^{k} \lambda_i Z(t_i). \qquad (4)$$

With the constraint: $\sum_{i=1}^{k} \lambda_i = 1$,

a variance of the estimation σ^2 may be obtained by computing:

$$\sigma^2 = \text{var}[Z(t)] - \sum_{i=1}^{k} \lambda_i \text{cov}[Z(t), Z(t_i)] - \mu$$

$$= C - \sum_{i=1}^{k} \lambda_i C(h_i) - \mu. \qquad (5)$$

The basic idea of the ordinary krigeage is to minimise the variance of estimation which should be unbiased since the mean must also be estimated. Therefore, a lagrangian $L(\lambda, \mu)$ is formed enabling a minimisation under the constraint $\sum_{i=1}^{k} \lambda_i = 1$. Partial derivatives versus all λ_i and μ are put equal to zero. Thus, a linear system of

$k+1$ equations for $k+1$ unknown terms (k terms λ_i and μ) is built. In matrix form, this system is as follows:

$$
\begin{bmatrix}
C & C[Z(t_1),Z(t_2)] & C[Z(t_1),Z(t_3)] & . & . & C[Z(t_1),Z(t_{k-1})] & C[Z(t_1),Z(t_k)] & 1 \\
. & C & C[Z(t_2),Z(t_3)] & . & . & C[Z(t_2),Z(t_{k-1})] & C[Z(t_2),Z(t_k)] & 1 \\
. & . & C & . & . & C[Z(t_3),Z(t_{k-1})] & C[Z(t_3),Z(t_k)] & 1 \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & \textit{Symmetric} & . & . & . & C & C[Z(t_{k-1}),Z(t_k)] & 1 \\
. & . & . & . & . & . & C & 1 \\
1 & 1 & 1 & . & . & 1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ . \\ . \\ \lambda_{k-1} \\ \lambda_k \\ \mu
\end{bmatrix}
=
\begin{bmatrix}
C[Z(t_1),Z(t)] \\ C[Z(t_2),Z(t)] \\ C[Z(t_3),Z(t)] \\ . \\ . \\ C[Z(t_{k-1}),Z(t)] \\ C[Z(t_k),Z(t)] \\ 1
\end{bmatrix}
$$

$$\mathbf{A\lambda = b.}$$

And,

$$\sigma^2 = C - \lambda^{\mathrm{T}} b.$$

After solving the above system, an estimation of Z at a time t is given by equation (4).

When the variable mean m is known, one can use a simple krigeage to estimate the adjustment coefficients λ_i and the obtained system is similar to that above but by withdrawing the last equation (i.e., the equation relative to the constraint). If the variogram is without nugget this system is quite comparable to the Yule-Walker linear system (except for the right hand side vector) that is used in autoregression methodology in time series analysis. The estimation of Z at a time t is instead obtained as:

$$Z(t) = m + \sum_{i=1}^{k} \lambda_i \left(Z(t_i) - m \right).$$

Adaptation to longitudinal data
Consider for example the case of animal growth data and consider that we have a population of newborn individuals which will grow under some conditions and fix a time interval **T** for taking successive individual weights as shown in Figure 1.

This sequential weighing process is widely used for controlling or measuring growth data in almost all livestock.

The main issue that prevents the application of a variogram analysis and the krigeage technique to such data remains the stationarity hypothesis. Indeed, neither the first moment nor the second can be considered stationary. The mean should increase as a result of the growth process and the amount of variance should follow under a scale effect. However, one can transform all weights (i.e., W1, W2, W3 …etc) to standard scores (i.e., Z1, Z2, Z3 …etc) so that all new variables will share zero as mean and one as variance. Therefore, equation (3) becomes:

$$\gamma(h) = 1 - C(h).$$

One may notice that there are two variogram senses. One, computed by considering successive weights separated by a fixed lag (i.e.: 1T or 2T, etc.) and defined as a translational variogram which is similar to 1 minus the corresponding Pearson correlations (or correlogram). However, when h takes all possible increments of T, the computed descriptive variogram allows the modelling of the structure of the covariance within the longitudinal data. Note that for $h=1T$, the corresponding covariogram value is simply the average of Pearson correlations between all records discarded by this interval. Following the shape of the experimental variogram a
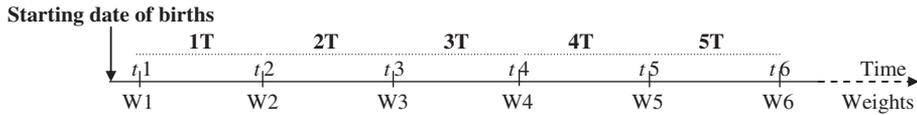
Starting date of births

Figure 1. System of weight control.

theoretical model can be adopted to compute a covariogram enabling the construction of a set of krigeage equations. Note that, if a linear model was retained, the krigeage technique would lead to coefficients matching a linear interpolation. Otherwise, it is preferable to fit an asymptotic model whose upper asymptote (i.e., the variance C) is ≤ 1. Whatever the chosen model, there is a range of time (**a**) that corresponds to the ordinate 0.5 where both variogram and covariogram intersect. This will be called a halftime variogram. Regarding the simplicity of its calculation, this new parameter may be considered as a straightforward tool to estimate and/or to compare the growth potential of populations. Thus, for a given population, a small halftime variogram should indicate a great growing potential, and vice-versa.

After solving the krigeage system, the coefficients obtained λ_i are used to estimate the missing data at the considered time. Unfortunately, the variance of the estimation (i.e., σ^2), cannot be computed by using equation (5) since the adjustment coefficients are determined on the basis of standard scores not on the original data.

Application to actual growth data of a fat-tailed Barbarine lamb population
Growth data were recorded on 242 Barbarine male lambs born in single litters during the fall of 2009 in the central region of Tunisia. A sequential weighing scheme (Figure 1) was used to collect six individual weights at six dates separated by twenty-one days (i.e., T=21 days). Lambs were weighed in the morning at each time point. Data collection started ten days after the first lambing and continued until six weights were recorded for lambs that were born late.

Results

The descriptive parameters of the recorded growth data, as well as their standard scores are reported in Table 1.

Experimental variogram
By applying equation (1) both experimental variograms are obtained:

1. A translational variogram (on one T) is shown in Figure 2.

Table 1. Distribution parameters of the six weights Wt and their standardised scores Zt

	Weights					
	W1	W2	W3	W4	W5	W6
Mean (kg)	3.02±0.05	6.73±0.08	9.96±0.09	13.40±0.13	16.19±0.18	19.78±0.25
CV (in %)	25.70	19.24	19.69	20.85	20.99	19.86
Scores	Z1	Z2	Z3	Z4	Z5	Z6
Means	−4.98E-07	−3.20E-07	−2.55E-05	1.52E-06	3.39E-07	7.61E-08
Variances	0.99	1.00	1.00	1.00	1.00	0.99

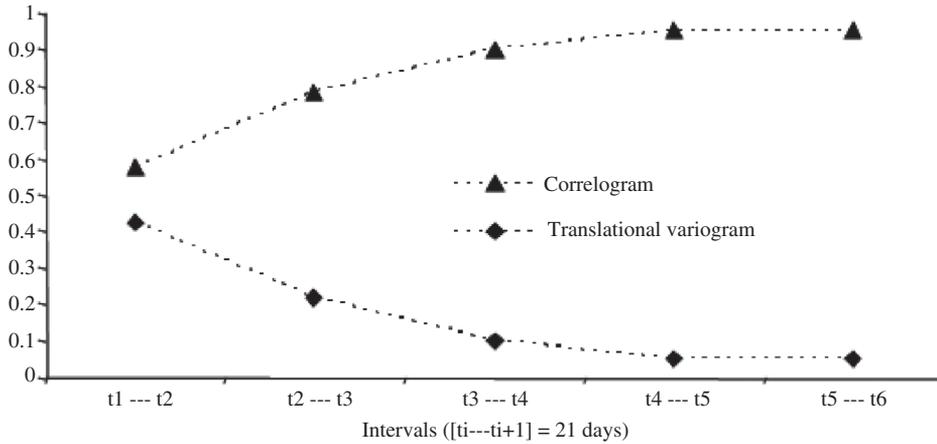*Figure 2. Translational variogram (γ) of Z scores on a period T=21 days and the corresponding correlogam (i.e., 1−γ).*

Note that as the age progresses, the correlations between closer weight records increase as shown by the correlogram (i.e., 1 minus variogram).

2. A descriptive variogram is reported in Table 2.

Variogram modelling

When $h=0$, two situations were distinguished in the model elaboration. The first is to consider that $\gamma(h)$ is equal to zero so no nugget effect is imposed on the model. The second is to consider that there exists a non-zero nugget effect (i.e., $C_0 \neq 0$) in the variogram model. To fit both model types, CurveExpert Software (Hyams 2010) was used by specifying the curve finder application. The first two models retained for each situation cited above are reported in Table 3. The coefficient of determination (R²) values were computed as: $\sum(\hat{y}_i - \overline{y})^2 / \sum(y_i - \overline{y})^2$ and models were selected and classified by considering the χ² test of goodness of fit.

Two models were retained from this analysis: a linear fit to consider the conventional practice of determining weights at standard ages and the Logistic model because its upper asymptote (i.e., $C=0.968$) is lower than 1 and therefore it may be adopted as the variogram sill to compute the covariogram. As standardised records should have 1 as variance whatever the considered time, for the linear model, which is not asymptotic, the covariogram is computed as 1 minus the variogram. The observed variogram, as well as the fitted curves are represented in Figures 3 and 4. The nugget value indicated on the last figure is computed by considering $t=0$ in the expression of the Logistic model of the variogram [i.e., $C_0 = 0.968/(1+9.91)=0.089$].

The halftime variograms (**a**) for both models are obtained as follows:

Table 2. Experimental variogram γ for different time lags h (T=21 days)

h	$\gamma(h)$
1T	0.16
2T	0.30
3T	0.44
4T	0.63
5T	0.76

Table 3. Variogram models retained on the basis of their goodness of fit criterion (R^2); two best models were retained by type of fit (with and without nugget effect)

Type of fit	Model name	Equation modelling variogram $\gamma(h)$	R^2
Without nugget ($C_0=0$) [i.e., at: $h=0, \gamma(h)=0$]	Polynomial	$-1.88\text{E-}08\ t^4 + 3.71\text{E-}06\ t^3 - 2.22\text{E-}04\ t^2 + 1.15\text{E-}02\ t - 8.19\text{E-}04$	0.9999
	Linear	$7.60\text{E-}03\ t + 3.94\text{E-}04$	0.9994
With nugget ($C_0 \neq 0$) (i.e., at: $h=0, \gamma(h) \neq 0$)	Logistic	$\dfrac{0.968}{1+9.91\,\text{Exp}(-0.036\ t)}$	0.9995
	Gompertz	$1.297\text{Exp}\big[-\text{Exp}(1.08-0.0172\ t)\big]$	0.9994

Nugget represents a micro-scale variation in measurement error ($\lim\limits_{h \to 0}\gamma(h)=C_0$).

For a linear fit ($y=\alpha x+\beta$), **a** is calculated as : $(1-2\beta)/2\alpha$.

For a logistic fit ($y=\alpha/[1+\beta\text{Exp}(-\nu x)]$), **a** is calculated as: $[\ln(\beta)-\ln(\alpha)]/\nu$.

These two values appear on their corresponding figures.

Kriging data
By way of illustration, W3 was removed and was estimated by ordinary krigeage following three plan types as shown in Table 4. Plans A and B assume that W3 is missing within an equivalent interval of 2T, plan C assumes that this record is missing within an equivalent interval of 3T while plans D, E and F assume that W3 is missing within an equivalent interval of 4T. To estimate W3 following plan A, the adjustment coefficients are obtained by solving the following systems:
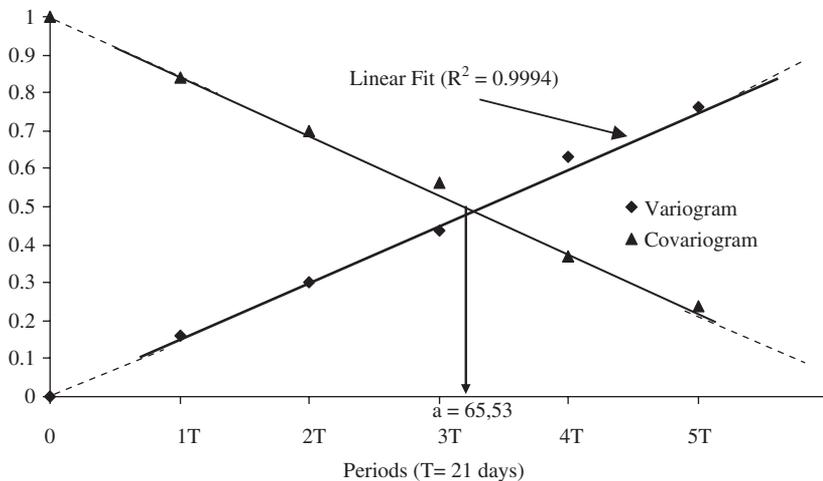

*Figure 3. Descriptive variogram γ(**h**) of Z scores modelled following a linear fit and its corresponding covariogram (i.e.: 1−γ(**h**)). (a) reported on the figure corresponds to the halftime variogram.*
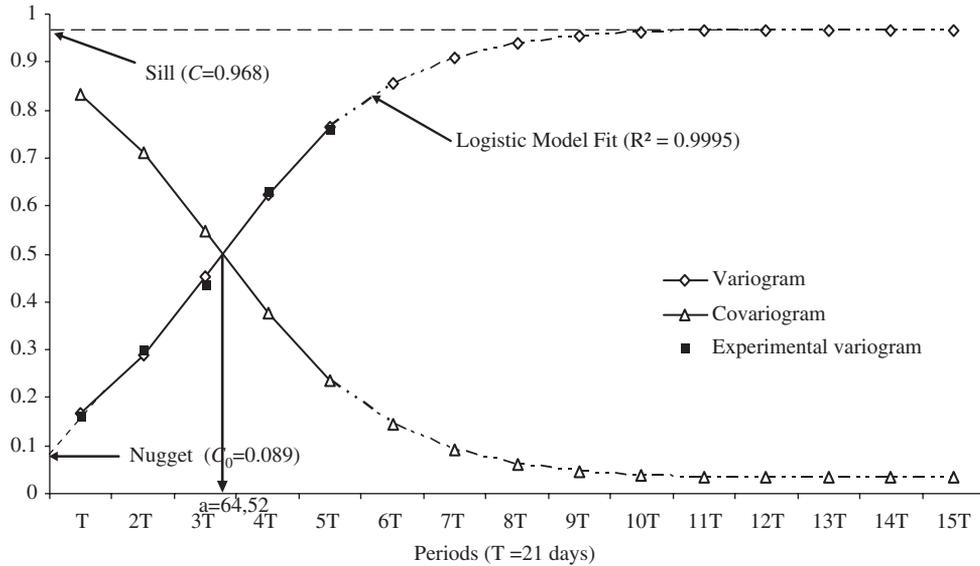
*Figure 4. Descriptive variogram γ(**h**) of Z scores modelled following a Logistic equation and its corresponding covariogram (i.e.: 0.968–γ(**h**)). (a) Reported on the figure corresponds to the halftime variogram.*

From a linear fit (without nugget effect)

$$
\begin{bmatrix}
1 & 0.85 & 0.54 & 0.39 & 0.24 & 1 \\
 & 1 & 0.69 & 0.54 & 0.39 & 1 \\
 & & 1 & 0.85 & 0.69 & 1 \\
 & Symmetric & & 1 & 0.85 & 1 \\
 & & & & 1 & 1 \\
 & & & & & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \mu
\end{bmatrix}
=
\begin{bmatrix}
0.69 \\ 0.85 \\ 0.85 \\ 0.69 \\ 0.54 \\ 1
\end{bmatrix}
$$

The solutions obtained are:

$$\begin{bmatrix} 0.0013 & 0.4987 & 0.4987 & 0.0013 & 3.4E{-}6 & -5.2E{-}7 \end{bmatrix}^{T}$$

and individual weights can be estimated by:
$\hat{W}3 = 0.0013\ W1 + 0.4987\ W2 + 0.4987\ W4 + 0.0013\ W5 + 3.4E\text{-}6\ W6$.

From a Logistic fit (with nugget effect)

$$
\begin{bmatrix}
0.97 & 0.71 & 0.43 & 0.26 & 0.12 & 1 \\
 & 0.97 & 0.59 & 0.43 & 0.26 & 1 \\
 & & 0.97 & 0.71 & 0.59 & 1 \\
 & Symmetric & & 0.97 & 0.71 & 1 \\
 & & & & 0.97 & 1 \\
 & & & & & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \mu
\end{bmatrix}
=
\begin{bmatrix}
0.59 \\ 0.71 \\ 0.71 \\ 0.59 \\ 0.43 \\ 1
\end{bmatrix}
$$

The solutions obtained are:

$$\begin{bmatrix} 0.171 & 0.329 & 0.328 & 0.169 & 0.003 & 0.0005 \end{bmatrix}^{T}$$

and individual weights can be estimated by:

$\hat{W}3 = 0.171\ W1 + 0.329\ W2 + 0.328\ W4 + 0.169\ W5 + 0.003\ W6$.

As indicated before, the solutions obtained by kriging on a linear model are equivalent to a simple linear interpolation (i.e., $\hat{W}3 = 0.5W2 + 0.5\ W4$). Nevertheless, with a logistic model, W1 and W5 are also involved in the adjustment of the estimated weight (i.e., $\hat{W}3$) in the plan A framework.

Linear model vs. Logistic model kriging

To assess the goodness of fit by kriging on both variogram models under the different plans of missing records around W3 (Table 4), the *n* observed weights at time *t*3 (W3) were compared to their corresponding estimated weights ($\hat{W}3$) by using the following criteria:

Table 4. Sampling scheme for missing records around W3

Plans	Weight records				
	W1	W2	W4	W5	W6
A	+	+	+	+	+
B		+	+		
C	+	+		+	+
D	+			+	
E	+	+			+
F	+			+	+

+ = Indicates that the corresponding record is known.

The mean absolute percent error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|W3_i - \hat{W}3_i\right|}{W3_i}\right)*100.$$

The mean absolute deviation (MAD):

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}\left|W3_i - \hat{W}3_i\right|.$$

The residual mean square (RMS):

$$\text{RMS} = \frac{1}{n}\sum_{i=1}^{n}\left(W3_i - \hat{W}3_i\right)^2.$$

The root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(W3_i - \hat{W}3_i\right)^2}{n}}.$$

The percentage of squared bias (PSB) proposed by Ali and Schaeffer (1987):

$$\text{PSB} = \frac{\sum_{i=1}^{n}\left(W3_i - \hat{W}3_i\right)^2}{\sum_{i=1}^{n}\left(W3_i\right)^2}\times 100.$$

The coefficient of determination calculated as the squared linear correlation between the observed and the estimated weights.

These statistics for the different scenarios of missing records around W3 are reported in Table 5. It can be seen that both interpolation approaches were quite comparable with a slight superiority in the accuracy term for a linear interpolation. As the data were collected during the first months of lamb life, which match a linear phase of growth, the results suggest that a simple linear interpolation can be used to estimate the missing weights at any desired time within this period. This technique is actually in common use for determining the weight at standard ages. However, when both known data points around the missing data are distant in the time-series data, a krigeage on the basis of the logistic model provided a more accurate estimation (i.e., plan F).

Discussion

Almost all studies of growth traits in livestock have observed that data recorded closer in time are usually more correlated than those recorded further apart in time (Portolano *et al.* 2002; Gowane *et al.* 2011). The descriptive variogram investigation as presented in this work clearly confirms those results, mainly from a phenotypic viewpoint. Furthermore, it has been shown, by the means of a translational variogram, that during the period of lamb growth studied, the correlation between weights recorded close together is higher at the end of this period than at the beginning. This is probably due to an enhanced maternal effect (i.e., suckling) characterising the earlier phase of lamb growth (María, Boldman and van Vleck 1993). Initially the weight gain depends on the milk production of the dam and the lamb potential for growth; thereafter it becomes progressively dependent on the potential of growth under a specific environment (herd management) as the lamb becomes independent of its dam.

Table 5. Goodness of fit criteria for estimation of W3 data by considering different plans of missing records around them. [Mean absolute percent error (MAPE), mean absolute deviation (MAD), residual mean square (RMS), root mean squared error (RMSE), percentage of squared bias (PSB) and coefficient of determination (R^2)]

Plans [1]	A		B		C		D		E		F	
MAPE	5.55 [2]	*6.07* [2]	5.55	*5.55*	6.39	*6.72*	8.93	*8.93*	6.34	*6.83*	8.91	*8.28*
MAD	5.11	*5.84*	5.11	*5.11*	6.14	*6.51*	8.67	*8.67*	5.96	*6.68*	8.67	*8.02*
RMS	55.01	*61.31*	55.01	*55.01*	68.95	*75.03*	125.6	*125.6*	68.11	*78.18*	125.2	*111.3*
RMSE	7.42	*7.83*	7.42	*7.42*	8.30	*8.66*	11.21	*11.21*	8.25	*8.84*	11.20	*10.55*
PSB	0.54	*0.60*	0.54	*0.54*	0.67	*0.73*	1.22	*1.22*	0.66	*0.76*	1.22	*1.08*
R^2	0.93	*0.88*	0.93	*0.93*	0.83	*0.81*	0.90	*0.90*	0.79	*0.75*	0.90	*0.92*

[1]Plans are A: all other records are known, B: W1, W5 and W6 are missing, C: only W4 is missing, D: W2 and W4 are missing, E: W4 and W5 are missing and F: W2 and W4 are missing.
[2]Values in underlined italic style correspond to the Krigeage on a logistic model and those in normal style correspond to the Krigeage on a linear model.

Unlike previous studies which have attempted to model growth by continuous time functions (Gous *et al.* 1999; Kratochvilovà *et al.* 2002; Lewis and Brotherstone 2002; Lambe *et al.* 2006; Abegaz *et al.* 2010; da Silva *et al.* 2012), in the present work the standardised weight variogram trend over time gaps was assessed. This has generated a clear vision of the covariance structure within the growth data. The Logistic model seems to be more suitable to fit a variogram of this function in the Barbarine breed of sheep throughout its weight gain period examined. This result agrees with Pittroff *et al.* (2008) and da Silva *et al.* (2012) who confirmed that the Logistic equation is the most appropriate model of the growth curve in sheep. The Gompertz fit proposed by other authors like Lewis and Brotherstone (2002) and Lambe *et al.* (2006) was not retained here for modelling the variogram since its upper asymptotic value exceeded 1 (the variance of a standard Gaussian distribution).

Through the deduced covariogram, it has been possible to elaborate the krigeage equations system. This allowed the possible estimation of any missing record. In the example presented in this work, only six weights for each animal covering only the first part of the growth process were used. During this phase, growth over time is essentially linear in sheep (Bush and Lewis 1977) and the estimation of missing data may be based only on a simple linear interpolation since both variogram models give comparable results. Nevertheless, it is possible to obtain more accurate estimation with the krigeage on the Logistic model basis if the known records are very distant from the missing record.

For further investigations of the entire variogram shape, there should be more records covering the whole growth period. For example, twenty records on average were used by Lewis and Brotherstone (2002). It is expected that an interpolation based on a Logistic model would be more appropriate for estimating data between the earlier phase of growth and mature weight since the nonlinearity of the growth curve within this inter-phase period is most pronounced (i.e., the growth rate is more variable).

It is possible to apply the methodology developed in this work to study the covariance structure within other longitudinal data like milk production. For this particular trait, it would be interesting to know if the covariance within a lactation can be also modelled by a known equation (Wood, Wilmink, etc.) and how much it can be affected by environmental

components. A comparison between the covariance structures of different parities should likewise be of great interest.

The inflection point for a Logistic model of the form: $y(t)=[a/(1+be^{-ct})]$ is given by: $\ln(b)/c$ (Winsor 1932). From our variogram function, this value is 63.71 days which is similar to 64.52 or 65.53 days that we determined in this work as halftime variograms. Therefore, this new parameter may be used as a straightforward tool to characterise the population growth potential. Pittroff *et al.* (2008) discussed the coincidence of the inflexion point and the onset of puberty in ewes which is coupled with an increase in proportion of fat gain (i.e., the Brody law). Since male and female lambs have different inflection points (Goliomytis *et al.* 2006; Ulutas *et al.* 2010), female Barbarine lambs are expected to have a different value of halftime variogram. A comparison between local breeds on the basis of this parameter could also be envisaged since age at inflection point may vary between breeds (Akbas *et al.* 2006).

If the first three weights W1, W2 and W3 are known, the estimation of other missing data (i.e., W4, W5 or W6) may be achieved only if we consider that $\sum_{i=1}^{k} \lambda_i > 1$. If the average weight at $t4$, $t5$ or $t6$ is known, one can iterate computations by incrementing the value assigned to this constraint in the right hand side of the system until reaching the defined mean in the estimated data vector. It would also be possible to make estimations of the missing data through a multiple regression performed on the independent variables: W1, W2 and W3.

Conclusion

The variogram methodology has been adapted to perform a detailed description of the (co)variance structure within longitudinal data. The modelling of this structure has allowed the use of the krigeage technique as a means of estimating missing data. With this method all known records are involved in order to give a more accurate estimation.

References

Abegaz, S., Van Wyk, J.B. and Olivier, J.J. 2010. Estimation of genetic and phenotypic parameters of growth curve and their relationship with early growth and productivity in Horro sheep. *Archiv Tierzucht* **53:** 85–94.

Akbas, Y., Alçiçek, A., Önenç, A. and Güngör, M. 2006. Growth curve analysis for body weight and dry matter intake in Friesian, Limousin x Friesian and Piemontese x Friesian cattle. *Archiv Tierzucht* **49:** 329–339.

Ali, T.E. and Schaeffer, R. 1987. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science* **67:** 637–644.

Alvarez, R.J., Joy, M., Villalba, D. and Sanz, A. 2008. Growth analysis in light lambs raised under different management systems. *Small Ruminant Research* **79:** 188–191.

Burgess, T.M. and Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties. 1. The semi-variogram and punctual kriging. *Journal of Soil Science* **31**: 315–331.

Bush, L.F. and Lewis, J.K. 1977. Growth patterns of range-grazed Rambouillet lambs. *Journal of Animal Science* **45:** 953–960.

Cai, W., Wu, H. and Dekkers, J.C.M. 2011. Longitudinal analysis of body weight and feed intake in selection lines for residual feed intake in pigs. *Asian-Australian Journal of Animal Science* **24:** 17–27.

Cappio-Borlino, A., Pulina, G. and Rossi, G. 1995. A non-linear modification of Wood's equation fitted to lactation curves of Sardinian dairy ewes. *Small Ruminant Research* **18:** 75–79.

Carrat, F. and Valleron, A.J. 1992. Epidemiologic mapping using the "kriging" method: application to an

influenza-like illness epidemic in France. *American Journal of Epidemiology* **135:** 1293–1300.

Chang, Y., Rekaya, R., Gianola, D. and Thomas, D.L. 2001. Genetic variation of lactation curves in dairy sheep. A Bayesian analysis of Wood's function. *Livestock Production Science* **71:** 241–251.

Chatfield, C. 1996. The Analysis of Time Series: An Introduction. 5th edition, Chapman & Hall, London, UK, 283 pages.

Cressie, N. 1986. Kriging nonstationary data. *Journal of the American Statistical Association* **395:** 625–634.

Da Silva, L.S.A., Fraga, B.A., de Lima da Silva, F., Beelen, P.M.G., de Oliveira Silva, F.M., Tonhati, H. and da Costa Barros, C. 2012. Growth curve in Santa Inês sheep. *Small Ruminant Research* **105:** 182–185.

Diaz, G., Zucca, A., Setzu, M.D. and Cappai, C. 1997. Chromatin pattern by variogram analysis. *Microscopy Research and Technique* **39:** 305–311.

Fernández, C., Sánchez, A. and Garcés, C. 2002. Modeling the lactation curve for test-day milk yield in Murciano-Granadina goats. *Small Ruminant Research* **46:** 29–41.

Flores, E.B., Kinghorn, B.P. and van der Werf, J. 2013. Predicting lactation yields in dairy buffaloes by interpolation and multiple trait prediction. *Livestock Science* **151:** 97–107.

Gbangboche, A.B., Glele-Kakai, R., Salifou, S., Albuquerque, L.G. and Leroy, P.L. 2008. Comparison of non-linear growth models to describe the growth curve in West African Dwarf sheep. *Animal* **2:** 1003–1012.

Goliomytis, M., Orfanos, S., Panopoulou, E. and Rogdakis, E. 2006. Growth curves for body weight and carcass components, and carcass composition of the Karagouniko sheep, from birth to 720 days of age. *Small Ruminant Research* **66:** 222–229.

Gous, R.M., Moran E.T.J.R., Stilborn, H.R., Bradford, G.D. and Emmans, G.C. 1999. Evaluation of the parameters needed to describe the overall growth, the chemical growth, and the growth of feathers and breast muscles of broilers. *Poultry Science* **78:** 812–821.

Gowane, G.R., Chopra, A., Prakash, V., and Arora, A.L. 2011. Estimates of (co)variance components and genetic parameters for growth traits in Sirohi goat. *Tropical Animal Health and Production* **43:** 189–198.

Hyams, D.G. 2010. CurveExpert Software (Basic Version 1.40). The University of Tennessee, Chattanooga, TN, USA Available online: http://www.curveexpert.net/ [Accessed 19 August 2014].

Jaquet, O. 1989. Factorial kriging analysis applied to geological data from petroleum exploration. *Mathematical Geosciences* **21:** 683–691.

Kitandis, P.K. 1983. Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research* **19:** 909–921.

Koivula, M., Sevon-Aimonen, M.L., Strandén, I., Matilainen, K., Serenius, T., Stalder, K.J. and Mäntysaari, E.A. 2008. Genetic (co)variances and breeding value estimation of Gompertz growth curve parameters in Finnish Yorkshire boars gilts and barrows. *Journal of Animal Breeding and Genetics* **125:** 168–175.

Kratochvilovà, M., Hyànkoà, L., Knizetovà, H., Fiedler, J. and Urban, F. 2002. Growth curve analysis in cattle from early maturity and mature body size viewpoints. *Czech Journal of Animal Science* **47:** 125–132.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52:** 119–139.

Krige, D.G. 1966. Two-dimensional weighted moving average trend surfaces for ore-evaluation. *Journal of Southern African Institute of Mining and Metallurgy* **66:** 13–38.

Lai, D. 2004. Geostatistical analysis of Chinese cancer mortality: Variogram, kriging and beyond. *Journal of Data Science* **2:** 177–193.

Lambe, N.R., Navajas, E.A., Simm, G. and Bünger, L. 2006. A genetic investigation of various growth models to describe growth of lambs of two contrasting breeds. *Journal of Animal Science* **84:** 2642–2654.

Leclerc, H., Duclos, D., Barbat, A., Druet, T. and Ducrocq, V. 2008. Environmental effects on lactation curves included in a test-day model genetic evaluation. *Animal* **2:** 344–353.

Lewis, R.M. and Brotherstone, S. 2002. A genetic evaluation of growth in sheep using random regression techniques. *Animal Science* **74:** 63–70.

Littell, R.C., Henry P.R. and Ammermann, C. 1998. Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science* **76:** 1216–1231.

Littell, R.C., Pendergast, J. and Natarajan, R. 2000. Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* **19:** 1793–1819.

Macciotta, N.P.P., Cappio-Borlino, A. and Pulina, G. 2000. Time series autoregressive integrated moving average modelling of test-day milk yields of dairy ewes. *Journal of Dairy Science* **83:** 1094–1103.

Macciotta, N.P.P., Fresi, P., Usai, G. and Cappio-Borlino, A. 2005. Lactation curves of Sarda breed goats estimated with test-day models. *Journal of Dairy Research* **72:** 470–475.

María, G.A., Boldman, K.G. and van Vleck, L.D. 1993. Estimates of variances due to direct and

maternal effects for growth traits of Romanov sheep. *Journal of Animal Science* **71**: 845–849.

Marzban, C. and Sandgathe, S. 2009. Verification with variograms. *Weather and Forecasting* **24**: 1102–1120.

Matheron, G. 1971. "The Theory of Regionalised Variables and its Applications". (Translated from: La Théorie des Variables Régionalisées et ses Applications, Note Geostatistique No 117). Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, Ecole des Mines de Paris, Paris, 212 pages.

Morant, S.V. and Gnanasakthy, A. 1989. A new approach to the mathematical formulation of lactation curves. *Animal Production* **49**: 151–162.

Nansen, C. 2012. Use of variogram parameters in analysis of hyperspectral imaging data acquired from dual-stressed crop leaves. *Remote Sensing* **4**: 180–193.

Owens, F.N., Dubeski, P. and Hanson, C.F. 1993. Factors that alter the growth and development of ruminants. *Journal of Animal Science* **71**: 3138–3150.

Pollot, G.E. and Gootwine, E. 2000. Appropriate mathematical models for describing the complete lactation of dairy sheep. *Animal Science* **71**: 197–207.

Pittroff, W., Dahm, F., Blanc, F., Keisler, D. and Cartwright, T.C. 2008. Onset of puberty and the inflection point of the growth curve in sheep – Brody's Law revisited. *The Journal of Agricultural Science* **146**: 239–250.

Portolano, B., Todaro, M., Finocchiaro, R. and van Kaam, J.H.B.C.M. 2002. Estimation of the genetic and phenotypic variance of several growth traits of the Sicilian Girgentana goat. *Small Ruminant Research* **45**: 247–253.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46**: 234–240.

Uuttera, J., Maltamo, M., Kurki, S. and Mykrä, S. 1998. Differences in forest structure and landscape patterns between ownership groups in central Finland. *Boreal Environment Research* **3**: 191–200.

Ulutas, Z., Sezer, M., Aksoy, Y., Sirin, E., Sen, U., Kuran, M. and Akbas, Y. 2010. The effect of birth types on growth curve parameters of Karayaka lamb. *Journal of Animal and Veterinary Advances* **9**: 1384–1388.

Wang, Z. and Goonewardene, L.A. 2004. The use of MIXED models in the analysis of animal experiments with repeated measures data. *Canadian Journal of Animal Science* **84**: 1–11.

Wilmink, J. 1987. Comparison of different methods of predicting 305-day milk yield using means calculated from within-herd lactation curves. *Livestock Production Science* **17**: 1–17.

Winsor, C.P. 1932. The Gombertz curve as a growth curve. *Proceedings of the National Academy of Sciences* **18**: 1–8.

Wood, P.D.P. 1967. Algebraic model of the lactation curve in cattle. *Nature* **216**: 164–165.

Yates, M.V., Yates, S.R., Warrick, A.W. and Gerba, C.P. 1986. Use of geostatistics to predict virus decay for determination of septic tank setback distances. *Applied Environmental Microbiology* **52**: 479–483.