Gappy Data Reconstruction and Applications in Archaeology

R. Stephan[1], K. Carlberg[2]

[1] Department of Classics, Stanford University
[2] Department of Aeronautics and Astronautics, Stanford University
{rstephan, carlberg}@stanford.edu

Abstract
This paper applies the Gappy proper orthogonal decomposition method, a recently-developed quantitative methodology for reconstructing unknown data, to archaeological problems and highlights the benefits of the method for quantitative analysis within the field. There are three main advantages of the method over polynomial regression, which is most commonly used for missing data problems. First of all, the method can be applied to problems where there are more attributes (variables) than complete samples (samples without missing data). Secondly, the method generates principal components (i.e. eigensamples) as a byproduct of its prediction algorithm; these principal components are useful for characterizing data sets and quantitatively assessing their dominant trends. Thirdly, the method employs fewer ad hoc *modelling choices compared with standard regression, as it uses empirically-derived basis functions (the principal components) to make predictions. The Gappy proper orthogonal decomposition method is applied to two case studies on Roman housing focusing on the sites of Pompeii and Herculaneum. These case studies highlight the merits of the method and provide a new quantitative assessment of housing characteristics of the area, a subject which has been primarily analyzed impressionistically. The study concludes by offering extensions and future directions for applying the methodology.*

Keywords: Pompeii, housing, quantitative methods, statistical analysis, proper orthogonal decomposition.

1. Introduction

Archaeology is, by definition, the study of the past through fragmentary material remains. When scholars generalize about the past, they essentially complete the fragments, stating implicitly or explicitly that their argument would hold true if the material record were indeed complete. While quantitative methods are becoming more commonplace in archaeological research today, generalizations and categorizations are still often made impressionistically and anecdotally without any quantitative assessment of the dominant trends involved.

In this paper, a new quantitative methodology known as Gappy proper orthogonal decomposition (POD) is introduced to address these shortcomings. This method, which has received recent attention in the machine learning and numerical analysis communities, fills in missing data by first computing statistical trends in the data, and then using these trends to fill in the corresponding gaps. Thus, this multidisciplinary study aims to leverage this new engineering tool to develop fresh insights into archaeological problems.

Archaeological research into housing practices is important because it provides information about the economic well-being and cultural values of a population. Studies on Roman housing, however, have exemplified the impressionistic and non-quantitative methodologies described above. Scholars have identified numerous housing styles across the Roman empire primarily by choosing several archetypal examples with little or no accompanying statistical analysis (BARTON (ED.) 1996, ELLIS 2000, MCKAY 1975).

This paper uses Gappy POD to address these shortcomings of traditional scholarship on Roman housing. Two case studies are undertaken in which the number of attributes (i.e. variables) and complete samples (i.e. samples with no missing attributes) differ greatly. First, the principal components of the data set are statistically determined for the well-preserved towns of Pompeii and Herculaneum. The principal components are then used to quantitatively identify the dominant characteristics of the attributes. Finally, these trends are used to reconstruct the gaps in partially excavated or poorly preserved houses. Pompeii and Herculaneum were specifically chosen for this study because the completeness of their data enables comparing the predicted values to the actual values.

The paper proceeds by first discussing the Gappy POD requirements and methodology. It then highlights benefits of the method over standard regression techniques. Next, it uses housing attributes at Pompeii and Herculaneum as case studies to show the efficacy in using Gappy POD to reconstruct fragmentary data. Finally, alternative potential archaeological applications for Gappy POD are discussed as possible avenues for further research.

2. Gappy data problem formulation

We begin by mathematically formulating the problem at hand. Essentially, the problem is to estimate unknown or uncomputed entries (i.e. "gaps") in vector-valued data.

The following terms will be used in the subsequent discussion.

Definition 1. A **sample** A is represented by an n-vector, that is, a vector with n entries:

$$x(A) = \begin{bmatrix} x_1(A) \\ \vdots \\ x_n(A) \end{bmatrix}.$$

Definition 2. An **attribute** is an entry of the vector used to represent a sample. For example, if sample A is a house in Pompeii, then attribute 1 of sample A, denoted by $x_1(A)$, may correspond to the number of rooms in the house, and $x_2(A)$ may be the area of the house, etc.

Definition 3. A **complete sample** is a sample where all attributes are known. That is, sample A is complete if $x_i(A)$ is known for all $i = 1, \ldots, n$.

Definition 4. A **gappy sample** is a sample where some attributes are unknown or missing. That is, sample A is gappy if $x_i(A)$ is unknown for some i.

The objective of gappy data reconstruction is to estimate the unknown attributes of gappy samples with a quantitative measure of uncertainty attached to the estimate.

While several methods exist in the literature to accomplish this task, one recently-developed method is particularly well-suited for the problem: Gappy POD. The next section describes this method.

3. Gappy POD

Gappy POD is a method for solving the gappy data problem that was introduced for the purpose of facial image reconstruction (EVERSON et al.1995). It has also been successfully used for aerodynamic flow field reconstruction (BUI-THANH et al. 2004, VENTURI et al. 2004, WILLCOX 2006), optimization (ROBINSON et al. 2006) and efficient numerical simulation of nonlinear systems (BOS et. al. 2004, CARLBERG et al. 2010).

The method requires both a collection of complete samples and some gappy samples whose unknown entries are to be estimated. The steps of the method are: 1) compute a set of "eigensamples" that best represent the complete samples, 2) reconstruct the gappy samples by least squares regression in one discrete-valued variable using these eigensamples as basis functions.

3.1. Eigensample computation

To build the eigensamples, m complete samples, denoted by A_j for $j = 1, \ldots, m$, are first "rescaled" according to the expression

$$\bar{x}_i(A_j) = \frac{x_i(A_j) - \mu_i}{s_i}.$$

Here, $\mu_i = \frac{1}{m} \sum_{j=1}^{m} x_i(A_j)$ is the sample mean and $s_i^2 = \frac{1}{m} \sum_{j=1}^{m} (x_i(A_j) - \mu_i)^2$ is the sample variance, which serves as a scaling factor. This rescaling essentially puts all attributes, regardless of units and relative magnitudes, on the same playing field. The rescaled samples are then assembled into the **complete sample matrix**:

$$W = [\bar{x}(A_1) \ \cdots \ \bar{x}(A_m)].$$

The method next computes the singular value decomposition (GOLUB and VAN LOAN, 1996), a common matrix decomposition, of this matrix:

$$W = U\Sigma V^T.$$

Here, $U = [u^1 \ \cdots u^m]$ is the matrix containing the left singular vectors u^i; Σ is a matrix with all zeros except for its diagonal containing the singular values σ_i, for $i = 1, \ldots, m$; and $V = [v^1 \ \cdots v^m]$ is the matrix containing the right singular vectors v^i.

Definition 5. The i^{th} **eigensample** is u^i, the i^{th} left singular vector of the complete sample matrix. In the literature, it is often referred to as the i^{th} principal component of the data set. It constitutes the i^{th} best representation of the (scaled) complete samples in the following sense:

$$u^i = \arg\min_{u \in \mathcal{S}^i} \sum_{j=1}^{m} \frac{(u, \bar{x}_j)}{\|u\|}$$

where

$$\mathcal{S}^i = \{u \mid \|u\| = 1, (u, u^k) = 0, \ \forall k < i\},$$

and $\|\cdot\|$ and $(\cdot,\cdot)$ are the Euclidean norm and inner product, respectively.

Often, the first $p < m$ eigensamples are selected to represent the set. This approach to computing dominant modes is (equivalently) known as proper orthogonal decomposition, principal component analysis, and the Karhunen-Loève transform in the literature.

3.2. Least-squares reconstruction

Once the eigensamples have been computed, the Gappy POD method uses them to estimate the missing entries of the gappy samples. This is accomplished by solving a least squares regression problem. Denoting by $\mathcal{N}(B) \subset \{1, 2, \ldots, n\}$ with $|\mathcal{N}(B)| = l$ the l known entries of gappy sample B, the least squares problem computes coefficients a_j, $j = 1, \ldots, p$ that satisfy the minimization problem

$$\underset{\{a_1,\ldots,a_p\}}{\text{minimize}} \sum_{i \in \mathcal{N}(B)} \left(\bar{x}_i(B) - \sum_{j=1}^{p} u_i^j a_j \right)^2 .$$

The expected value (mean) of the missing attributes $x_i(B)$, $i \in \{1, 2, \ldots, n\} \backslash \mathcal{N}(B)$ can then be computed as

$$E\left[x_i(B)\right] = \mu_i + s_i^2 \sum_{j=1}^{p} u_i^j a_j .$$

The variance of the estimate is

$$\sigma_i^2 = \frac{s_i^2}{l} \sum_{k=1}^{l} \left(\bar{x}_k(B) - \sum_{j=1}^{p} u_k^j a_j \right)^2 .$$

Since least-squares regression assumes a Gaussian distribution, the expected value and standard deviation can be used to quantify the uncertainty in the gappy prediction; that is, a confidence interval can be constructed.

Note that the least-squares minimization problem above is of the standard form $\text{minimize} \| Xa - y \|$, where $X \in \mathbb{R}^{l \times p}$ is simply the $\mathcal{N}(B)$ rows of $\left[u^1 \ \cdots \ u^p \right]$, which correspond to the known attributes of the gappy sample. This is depicted in Figure 1.

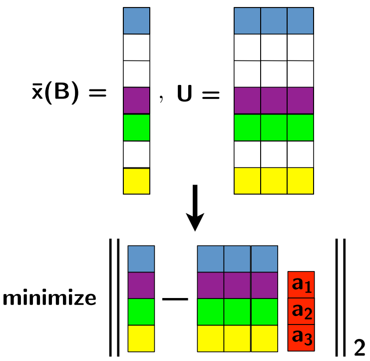

Figure 1: *Least-squares reconstruction example with* $n = 7$, $l = 4$, $p = 3$, *and* $\mathcal{N}(B) = \{1, 4, 5, 7\}$.

4. Gappy POD v. standard regression

This method carries several advantages over standard regression methods in the context of gappy data reconstruction. By "standard regression," we refer to the commonly-used supervised learning method which treats the known attributes as inputs and the unknown attributes as outputs; the approach uses the complete samples to construct a mapping from the inputs to the outputs using canonical mathematical basis functions such as polynomials.

First, Gappy POD can handle problems characterized by more known gappy attributes than full samples ($p > n$). Such problems arise in archaeology when there are a small number of well-preserved sites, features, or artifacts, which could yield a large amount of information. In the ancient Roman world this is exemplified in the Campania region of Italy. The extraordinarily well-preserved sites of Pompeii and Herculaneum have been repeatedly used as prototypes for Roman urban layout and housing. In both of these cases the number of attributes measured, regarding either the urban fabric or domestic architecture, can easily exceed the number of complete samples.

To see this advantage of Gappy POD, consider the associated least-squares problem, which has a unique solution if and only if the associated $(l \times p)$ matrix is "skinny" (i.e. $l \geq p$) and has full column rank. The first of these conditions is satisfied if there are more known attributes than complete samples ($l \geq m$), since $m \geq p$ in this case. The second condition is difficult to ensure in the general case, as it depends on which attributes $\mathcal{N}(B)$ are known. Thus, it is likely that the Gappy POD method will generate a unique solution when there are more known gappy attributes than full samples. On the other hand, if (say) first-order polynomials are chosen as basis functions with standard regression, the associated least squares matrix is of dimension $(n \times p + 1)$. The matrix will only be "skinny" if $n \geq p + 1$, which is not possible for problems with $p > n$. In order to compute a unique solution, some basis functions must be removed; this decision is inherently *ad hoc.*

Secondly, Gappy POD requires fewer arbitrary choices than standard regression. In standard regression, the basis functions are chosen (arbitrarily) to be "typical" mathematical functions. This choice is made *a priori*, and there is generally no reason to believe that the choice reflects the actual trends in the data (why, for example, are polynomials a better choice than radial basis functions?). On the other hand, Gappy POD employs the empirically-derived eigensamples to make predictions. Not only are these functions not arbitrarily-chosen, they are *optimal* in terms of representing the complete samples.

Finally, the eigensamples used by the method are interesting in their own right: They can be used to mathematically characterize the population. For example, assume that the samples correspond to excavated houses in Pompeii, and the attributes are the number of rooms and square footage. In this case, the first eigensample describes (roughly) the ratio of rooms to total area that is most likely to be encountered in this region.

Clearly, there are many problems in archaeology that fall under the umbrella of data reconstruction. The Gappy POD method is very well-suited to tackle many of these problems. The next section describes one such problem and presents some preliminary results.

5. Case studies

Pompeii and Herculaneum form the basis of almost every traditional account of housing in Roman Italy (BARTON (ED.) 1996, CLARKE 1991, ELLIS 2000). The "Romanness" of these houses is identified through their constituent parts: the atrium, the peristyle, the cubiculum, the triclinium, etc. Ellis defines the typical Roman house as, "one with a large richly decorated reception room opening onto a central colonnaded courtyard or peristyle (ELLIS 2000, 10)." These accounts largely lack statistical analysis of the composition and location of the Pompeian houses.

One good attempt at a quantitative study of houses at Pompeii and Herculaneum was done by Wallace-Hadrill (WALLACE-HADRILL 1994). His study focused primarily on the social structure of the houses, showing how layout and decoration gave cues as to the public / private, grand / humble nature of particular rooms. Wallace-Hadrill also quantified certain aspects of the houses he surveyed. His statistical analysis included the size of the houses, the number of rooms they contained, and the number of rooms decorated. Wallace-Hadrill even went on to correlate house size with decoration, yet his correlations never moved beyond two variables.

This paper considers two case studies with different parameters. Both are concerned with Roman housing in Campania and use houses from Pompeii and Herculaneum to build the data sets. The first case study measures and assesses the relationship between four attributes; it draws on over 120 complete samples. The second case study increases the number of measured attributes to ten, but only uses seven complete samples to build the model. Both scenarios are common occurrences within archaeological research, but the latter is rarely undertaken due to the dimensionality limitations of standard regression.

5.1. Case study 1: Pompeii and Herculaneum

This case study uses Wallace-Hadrill's three categories of quantified data, along with a new fourth variable of 'distance to the forum.' Thus the four attributes assessed in this case study are: (1) house size, which may serve as a proxy for the economic well-being of the family, (2) number of rooms, a proxy for differentiation within the house, (3) number of decorated rooms, another proxy for wealth, and (4) distance to the forum, which may relate to the status of the owner (MORRIS 2005, WALLACE-HADRILL 1994). The case study then applies Gappy POD in order to identify the dominant trends of housing at Pompeii, where 93 houses were considered, and Herculaneum, where 29 houses were considered. These dominant trends allow us to (a) quantitatively represent characteristics of houses at the respective sites, and (b) reconstruct gaps in future data. Because of the unique level of preservation at these sites and the presence/absence nature of the decoration analysis, it can be assumed that the architectural and decorative remains at these sites represent a fairly full and accurate account of what was present in antiquity.

First, we computed the eigensamples of both the Herculaneum and Pompeii data sets. For Pompeii, the sample mean is μ=[409.6, 12,2, 5.0, 111.0]^T and the first eigensample is [-0.9982, -0.0184, -0.0139, -0.0053] ^{T}, where the entries of these vectors correspond to the housing attributes considered. Thus, the four attributes are related among the Pompeian houses primarily in the following way:

$$\begin{bmatrix} 409.6 \\ 12.2 \\ 5.0 \\ 111.0 \end{bmatrix} + c \begin{bmatrix} -0.9982 \\ -0.0184 \\ -0.0139 \\ 0.0553 \end{bmatrix}$$

That is, for most houses in Pompeii, there is a value of c that very closely describes it. For example, for house VI.15.1/27, the optimal value of c=-676.2 yields 1084.6 m^2 for the area, 24.6 rooms, 14.4 decorated rooms, and 73.7 m distance from the forum; these are very close to its true values of 1100 m^2, 24, 15, and 79 m , respectively. For Herculaneum, a similar, yet distinct trend emerges with the following dominant relationship:

$$\begin{bmatrix} 312.9 \\ 12.1 \\ 5.4 \\ 92.9 \end{bmatrix} + c \begin{bmatrix} -0.9984 \\ -0.0180 \\ -0.0146 \\ -0.0506 \end{bmatrix}$$

These rescaled eigensamples for Pompeii and Herculaneum are graphically depicted in Figures 2 and 3.

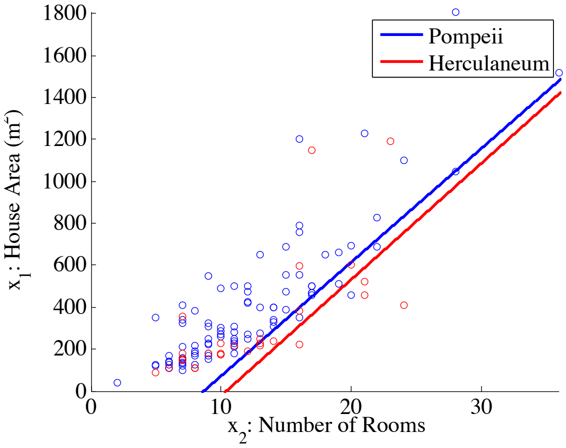

Figure 2: *Data points and rescaled eigensamples projected on the $x_2 - x_1$ plane*

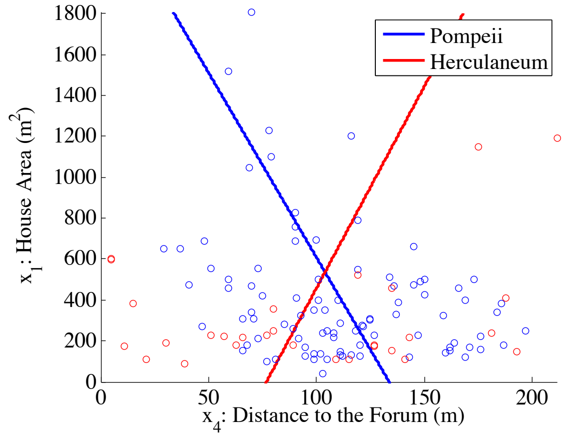

Figure 3: *Data points and rescaled eigensamples projected on the $x_4 - x_1$ plane*

A comparison between these rescaled eigensamples reveals the similarity between housing characteristics of the two towns. Pompeian houses tend to be slightly larger on average (by about 100 m^2), but the number of rooms and decorated rooms are almost identical (roughly 12 and 5, respectively). These three attributes trend very similarly as well for both Pompeii and Herculaneum, with both the total number of rooms and the number of decorated rooms increasing as house size increases (see Figure 2). The most prominent contrast lies in the relationship between house size and proximity to the forum. In Pompeii, as one moves away from the forum houses in the data set tend to get smaller, suggesting wealthy citizens chose to live near the town center. Herculaneum displays the opposite trend, where houses tend to be larger the further they are from the forum (see Figure 3). This observation constitutes an interesting topic for future investigation.

Next, we used these eigensamples to make predictions for gappy samples. In the first set of experiments, we used the first eigensample from the Herculaneum data set to make predictions for gappy samples of the Herculaneum data set. To generate each gappy sample, one house was removed from the Herculaneum data set and one of its attributes was treated as unknown. The Gappy POD procedure outlined in Section 3.2 was then applied to compute a 99% confidence interval for this “missing” value. It was then determined whether or not the true value of the removed attribute was contained in the Gappy POD confidence interval. As one illustrative example, for the *Casa del Papirio dipinto*, the number of rooms was removed as though it was a natural gap in the data. Gappy POD predicted with 99% confidence that there would be between 7.4 and 10 rooms; the true value of 8 indeed falls in this interval.

This experiment was run for all attributes and all houses in the Herculaneum set. Table 1 compares the success rates (i.e. percentage of correct confidence intervals) achieved by Gappy POD with those obtained by standard regression using both second- and third-order polynomials.

Method	x_1	x_2	x_3	x_4
Gappy POD	55.17%	62.07%	48.28%	20.69%
Regression (2nd order)	62.07%	55.17%	48.28%	48.27%
Regression (3rd order)	55.17%	48.27%	37.93%	37.93%

Table 1: *Success rates for Herculaneum as the complete set for predicting Herculaneum gappy samples.*

Table 1 indicates that Gappy POD performs similarly to standard regression in this case. Here, it generates a “correct” confidence interval more often when the second attribute is missing, but is less successful when the fourth attribute is missing.

The first Herculaneum eigensample was also used to predict missing attributes for houses in Pompeii. The gappy sample generation procedure is identical to that used above. The number of decorated rooms in house VI.14.25 provides an example of the Gappy POD predictions. In this case, Gappy POD predicted its value to be between 2.2 and 3.6 with 99% confidence. Its true value of 3 falls in this interval. Also, the distance to the forum for house VI.14.34 was correctly predicted to be between 71.4 m and 96.8 m; its actual value is 89 m.

Again, this experiment was run for all attributes and houses in Pompeii. The associated success rates are provided in Table 2. As before, Gappy POD exhibits similar performance to standard regression.

Method	x_1	x_2	x_3	x_4
Gappy POD	47.31%	52.69%	40.86%	19.35%
Regression (2nd order)	60.22%	44.09%	56.99%	40.86%
Regression (3rd order)	41.93%	40.86%	53.76%	43.01%

Table 2: *Success rates for Herculaneum as the complete set for predicting Pompeii gappy samples.*

These results highlight several advantages of Gappy POD. First, the method does not require as many arbitrary modeling choices as standard regression. For example, in the above experiments, 2nd-order polynomial basis functions generated more accurate confidence intervals than 3rd-order polynomial basis functions in some cases, but not others. It is not apparent *a priori* which basis functions are best to use.

Secondly, the eigensamples generated by Gappy POD are interesting in their own right. In the study above, they were a useful tool to quantitatively compare and contrast the housing characteristics of the two sites.

Finally, the study illustrates that eigensamples generated from one site (e.g. Herculaneum) can be employed to predict missing values from other sites (e.g. Pompeii). In the future, this implies that well-preserved sites could be used to accurately predict unknown information of unexcavated or poorly preserved sites.

5.2. Case study 2: Regio I, Pompeii

The goal of the second case study, in addition to providing a more quantitative analysis of housing characteristics at Pompeii, is to simulate a circumstance where there are more known attributes than complete samples. This would prohibit the (straightforward) use of standard regression due to the dimensionality limitations discussed in Section 4. These situations arise primarily at less well-preserved or more sparsely excavated sites, where there may only be a small number of complete samples. However, Pompeii is used again for this case study because it allows for a comparison between the values predicted by Gappy POD and the actual values. Here seven complete samples from Pompeii Regio I, Insula 11 were used to compute the eigensamples; we used the first of these eigensamples to predict the missing values for all houses from Regio I, Insulae 6–12.

A total of ten attributes are considered in this case study. The first four are the same as the previous study: (1) house size, (2) the number of rooms, (3) the number of decorated rooms, and (4) the distance to the forum. These are combined with six new attributes: (5) the primacy of the house within the insula, which measures how the size of the house compares to its neighbors, (6) the distance to the nearest fountain, which measures proximity to a water source, (7) the distance to the nearest brothel, a proxy for proximity to deviant locales, (8) courtyard size, measuring the open air space within the house, (9) the number of colonnades within the courtyard, a possible status marker, and (10) the number of attached shops, a measure for proximate economic activity. The majority of these data is again drawn from Wallace-Hadrill's *House and Society in Pompeii and Herculaneum* (WALLACE-HADRILL 1994). The data for fountain and brothel locations originates from Laurence's *Roman Pompeii: Space and Society* (LAURENCE 1996, 43, 66). The data set here is limited to Pompeii, Regio I, Insulae 6 through 12.

The experiment was run in a similar fashion to the previous study. That is, gappy samples were generated by removing one house from the Pompeian set and considering one of its attributes to be unknown.

The successful prediction rates for the ten attributes are: (1) 75.8%, (2) 51.5%, (3) 36.4%, (4) 45.5%, (5) 48.5%, (6) 21.2%, (7) 54.5%, (8) 60.6%, (9) 45.5%, (10) 39.4%. These rates appear to be relatively low in some cases, which could be attributed to a lack of correlation between attributes. However, we reiterate that standard regression cannot be applied to this problem in a straightforward manner; thus, the Gappy POD method enables results to be obtained for this problem without making *ad hoc* decisions regarding the omission of basis functions.

6. Conclusions and future work

This work has applied Gappy POD, a recently-developed methodology for reconstructing gappy data, to archeological problems. The three main advantages of the method over standard regression have been detailed. These include its applicability to previously difficult problems (i.e. more known attributes than complete samples), the utility of the eigensamples generated by the method, and the lack of *ad hoc* modeling decisions. Two case studies on Roman housing illustrated these benefits.

The potential archaeological applications of the Gappy POD method range far beyond what has been done here. For instance, it is possible to extend the geographical scope of housing research to include other areas of Italy and the Roman empire at large. This will allow for the comparison of statistical trends in regional housing. We may then create housing typologies using, at least in part, quantitative data, which will facilitate cross-cultural comparison. Furthermore, Gappy POD may be used to predict trends on a larger scale. Assessing the size, layout, average house size, building types, and geographical location of cities, would allow us to establish regional urban trends as well as predict urban composition for partially excavated towns. Finally, this method could be used in order to reconstruct the iconography of ceramic vessels (e.g., Greek vases), in much the same way that this method was originally employed in image reconstruction (EVERSON 1995). This has the potential to save both time and money by reducing the need for on-site ceramic analysis.

Additionally, the method could be applied to cases where no complete samples are available. This arises when individual sites tend to have a unique combination of excavated attributes. The procedure outlined in Section 3 of EVERSON, 1995 could be applied to such problems.

Overall, Gappy POD provides an innovative computer-based method through which archaeologists can generate new quantitative answers to traditional archaeological questions.

References

BARTON, I. M. (ED.)., 1996. Roman Domestic Buildings. University of Exeter Press, Exeter.

BOS, R., BOMBOIS, X., and VAN DEN HOF, P., 2004. Accelerating large-scale non-linear models for monitoring and control using spatial and temporal correlations. *Proceedings of the American Control Conference 4*, pp: 3705–3710.

BUI-THANH, T., DAMODARAN, M., and WILLCOX, K.., 2004. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *American Institute for Aeronautics and Astronautics Journal 42* (8), pp: 1505–1516.

CARLBERG, K., FARHAT, C., and BOU-MOSLEH, C., "Efficient Nonlinear Model Reduction via a Least-Squares Petrov-Galerkin Projection and Compressive Tensor Approximations," International Journal of Numerical Methods in Engineering, *submitted May 2010.*

CLARKE, J. R., 1991. The Houses of Roman Italy, 100 B.C. – A.D. 250: Ritual, Space, and Decoration. University of California Press, Berkeley.

ELLIS, S. P., 2000. Roman Housing. Duckworth, London.

EVERSON, R. and SIROVICH, L., 1995. Karhunen-Loeve procedure for gappy data. *Journal of the Optical Society of America A 12* (8), pp: 1657–1664.

GOLUB, G. and VAN LOAN, C., 1996, Matrix Computations, Johns Hopkins University Press, 3rd ed.

LAURENCE, R., 1996, Roman Pompeii: Space and Society, Routledge, London.

MCKAY, A. G., 1975. Houses, Villas, and Palaces in the Roman World. Cornell University Press, Ithaca.

MORRIS, I. 2005. Archaeology, standards of living, and Greek economic history. In J. Manning and I. Morris (eds.), The Ancient Economy: Evidence and Models, pp:91–126. Stanford University Press, Stanford.

ROBINSON, T., ELDRED, M., WILLCOX, K., and HAIMES, R., 2006. Strategies for Multifidelity Optimization with Variable Dimensional Hierarchical Models. AIAA Paper 2006-1819, 47th AIAA Structures, Structural Dynamics, and Materials Conference, May 1–4, 2006.

VENTURI, D. and KARNIADAKIS, G., 2004. Gappy data and reconstruction procedures for flow past a cylinder. *Journal of Fluid Mechanics 519*, pp: 315–336.

WALLACE-HADRILL, A., 1994. Houses and Society in Pompeii and Herculaneum. Princeton University Press, Princeton.

WILLCOX, K.., 2006. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers and Fluids 35*(2), pp: 208–226.