Principal Component and Independent Component Regression for Predicting the Responses of Nonlinear Base Isolated Structures

by

Sina Shirali

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Civil Engineering

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made available to the public.

Abstract

Peak base displacement is one of the most important quantities in the design of base-isolated buildings. During the preliminary stages of design, a nonlinear time-history analysis is often not possible or too expensive, and hence reliable measures for predicting peak base displacement must be obtained through other means. In this study, regression models are developed in order to predict the peak displacement using a series of intensity measures (*IMs*) as model inputs. This thesis utilizes two methods for this purpose, Principal Component Regression (PCR) and a newly proposed method known as Sorted-Input Independent Component Regression (SI-ICR). In the framework of PCR and SI-ICR, the problem that exists due to correlation of *IMs* is addressed, which allows the transformation of correlated components into uncorrelated ones. This step is followed by dimensionality reduction of the components that do not contribute significantly to the explained variance of the original data set. A regression model using only one *IM*, peak ground velocity (*PGV*), is also developed to compare the advantages of using multiple *IMs* as opposed to one. Prediction results are presented and compared to simulation results for building models with increasing degree of complexity, starting with a two degree of freedom uniaxial case to a twelve degree of freedom biaxial model. It is concluded that PCR and SI-ICR significantly outperform the *PGV* model with PCR slightly outperforming SI-ICR. PCR is regarded as a more suitable and practical regression method for predicting the responses of nonlinear base isolated structures.

Acknowledgements

First and foremost I would like to thank my supervisors Dr. Sriram Narasimhan and Dr. Mahesh Pandey for making this research possible. Second, I would also like to thank Mr. Budhayita Hazra for providing me with valuable advice on ICA, and Dr. Min Wang for providing me with the ground motion database and insights into PCA regression modelling. I would also like to thank my fellow colleagues in the Structural Dynamics and Control group, Mr. Paul Paquet and Mr. Aaron Roffel, for their support. Finally, I would like to thank Natural Sciences and Engineering Research Council (NSERC) of Canada, through their Discovery Grants Program and UNENE Industrial Research Chair programs, for providing the financial support to conduct this study.

Dedication

This is dedicated to my family and friends and to anyone who seeks to make the world a better place through the power of knowledge.

TABLE OF CONTENTS

1. INTRODUCTION – OBJECTIVES AND OVERVIEW

In recent years, base isolation has become a major design approach for structures in seismic-sensitive areas. The basic idea of base isolation is to lengthen the natural period of the structure using a system of either elastic or hysteretic isolation elements. The concept of base isolation is relatively simple but the design and engineering knowledge involved in constructing such systems is something that is yet to be perfected. Typically, base isolated structures use a specially designed isolation system, which provides flexibility and the necessary energy dissipation. For example, a popularly used isolation system is the Lead-Rubber bearing (LRB) system that consists of a lead core contained in a reinforced elastomer (Fig. 1.1). The flexibility is provided by the elastomer, whereas, the lead core provides the energy dissipation. There are other types of isolation systems such as linear elastomeric and friction pendulum systems, which utilize similar principles described here.

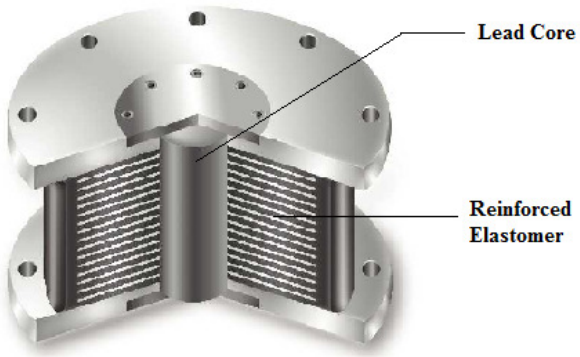

Figure 1.1. Lead Rubber bearing (LRB)

The end result is a reduction in interstory drifts and floor accelerations compared to their fixed base counterpart. Many hospitals, laboratories, and internet-based businesses in highly seismic areas have adopted this system. An example of this is the University of Southern California

hospital building in Northridge which performed well and reduced the response caused by the 1994 Northridge earthquake compared to a fixed-base structure (Nagarajaiah 2000).

With regards to the design of base isolated buildings, specifically the properties of the isolator, the calculation of peak displacement is critical. During preliminary design stages, a nonlinear time-history analysis required for this purpose is both time consuming and expensive. Hence, it is important to have a model that gives designers a reasonably accurate statistical estimate on the magnitude of this peak parameter. The problem with many building codes such as the International Building Code (ICC 2000) is that they use a linear single-degree of freedom system concept to calculate the base displacements of nonlinear base isolated buildings. Although relatively simple and fast, this method has been shown to underestimate the base displacements of nonlinear base isolated structures and alternative regression methods have been proposed (Ryan and Chopra 2004a, 2004b). In the current study, time-histories and spectral information of recorded earthquakes are used to calculate intensity measures (*IMs*) which are then used to model the response of nonlinear base isolated structures. An earthquake *IM* is a characteristic of a recorded ground motion that quantifies the intensity or the severity of a seismic event (Bianchini 2008). Some examples of *IMs* include peak ground acceleration (*PGA*), peak ground velocity (*PGV*), and peak ground displacement (*PGD*), which are all discussed in further detail during the course of this thesis.

PGV is regarded as a good measure of ground motion intensity in calculating the responses of base isolated buildings (Ryan and Chopra 2004a, 2004b). However, *PGV* alone does not give the best response prediction for nonlinear base isolated buildings as reported recently (Narasimhan *et al.* 2008). For this reason, this study utilizes other *IMs* in the model prediction. These *IMs* and others are studied in detail in this thesis followed by a correlation

analysis between the *IMs* to show their statistical relationships with one another. This is a key step that will enable the elimination of redundant *IMs* which may not contribute to useful information already present in other *IMs*.

The main objective of this thesis is to develop a multivariate (multidimensional) regression model to predict the response of nonlinear base isolated buildings. Since, as mentioned earlier, the *IMs* are correlated, a systematic procedure is required to address this problem of collinearity. Two approaches, namely Principal Component Regression (PCR), and Sorted-Input Component Regression (SI-ICR) are utilized for this purpose. This issue of collinearity has not been addressed in many studies on this subject (e.g. Riddel 2007). Dimensionality reduction is a major step in these methods which allows for discarding data that does not contribute significantly to the observed variance in the data.

PCR is a two step multivariate regression method. First, Principal Component Analysis (PCA) is applied to the data matrix followed by a multiple linear regression between the newly acquired transformed variables and the response variable. The response variable, namely the base displacement, is the quantity of interest that is to be modelled. PCA is a vector space transform that is used to reduce multidimensional data sets to lower dimensions for analysis. There are two main reasons why PCA is applied. First it transforms correlated *IMs* into uncorrelated variables. This allows for dimensionality reduction to be applied, which is the second reason why PCA is used. Models are developed and results compared according to the number of components retained in the regression model.

SI-ICR is motivated from a sorted PCR (SPCR) method which essentially has the same steps as PCR, but involves sorting the extracted scores with the response variable according to

the magnitude of correlation (Gustafsson 2005). SI-ICR has however another significant step, which includes applying Independent Component Analysis (ICA) (Gustafsson 2005). This step takes the newly uncorrelated variables, which are derived through a process called whitening or sphering (similar to PCA), and makes them statistically independent as well. After this step, the variables are sorted in terms of decreasing correlation to the response variable. Once this is done, linear regression is performed in the same fashion as PCR.

This study investigates the efficiency of using a combination of *IMs* as well as comparing the two regression methods, PCR and SI-ICR mentioned earlier. This analysis is done on a uniaxial base isolated (UBI) model as well as a biaxial base isolated (BBI) model. Uniaxial refers to excitation and response in a single (lateral) plane, whereas, biaxial refers to the case where the excitations and the responses of the structure occur in two orthogonal directions (planes). See Fig. 1.2.

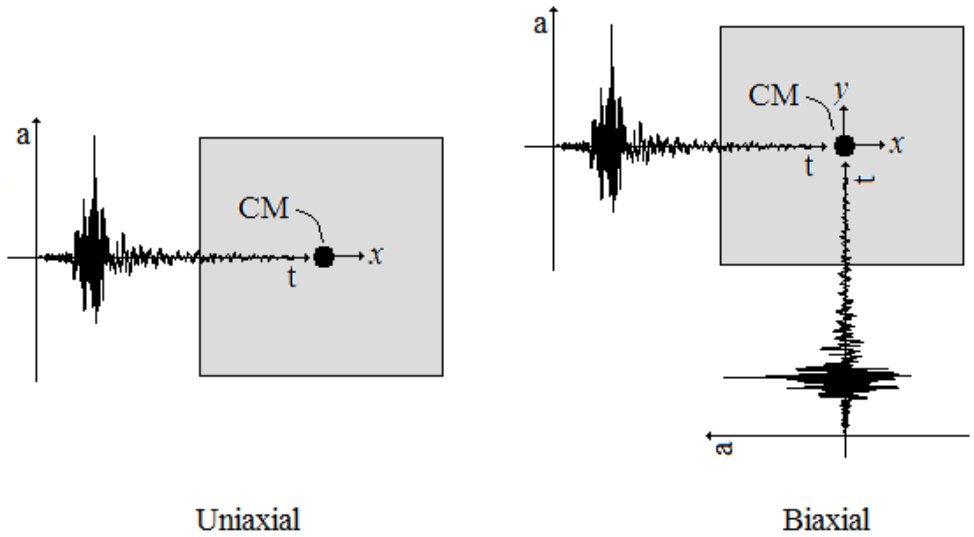

Figure 1.2. Uniaxial and biaxial models

Over 700 Japanese and North American earthquakes are used for the model development and validation of this study. Structural models with increasing degrees of complexity (degrees of freedom) are utilized for model development and validation. Both uniaxial and biaxial cases are utilized for this purpose. The superstructure is assumed to be linear, as is generally the practice for base isolated buildings, and the isolation system is assumed to consist of LRBs. The earthquake data base is divided into two large segments (different divisions for uniaxial and biaxial structures) and used for model development and validation purposes.

The thesis is organized as follows. Chapter 2 presents a detailed background on intensity measures, the proposed base isolated models, principal component regression, and sorted-input independent component regression. Chapter 3 outlines detailed steps on the problem statement by looking at state space representations, implementation of the earthquakes, correlation analysis, and detailed methodology of the proposed regression methods. This is followed by the results and discussion in Chapter 4 and conclusions in Chapter 5.

2. BACKGROUND, LITERATURE REVIEW AND RELATED WORKS

In this section, a detailed background is presented on the earthquake intensity measures (*IMs*), the proposed nonlinear base isolated models, principal component regression (PCR), and sorted-input independent component regression (SI-ICR). *IMs* are tools used for measuring the intensity of an earthquake and are mainly used as model inputs. There are two structural models that are utilized for model development and validation; one is a two degree of freedom (2DOF) *uniaxial* nonlinear base isolated structure and the other is a four degree of freedom (4DOF) *biaxial* model. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are data analyzing tools, which are combined with a regression method in order to develop the prediction models. These two regression methods are studied in detail and then compared for both uniaxial and biaxial nonlinear base isolated structures.

2.1 Intensity Measures

IMs characterize the strength of an earthquake needed for seismic analysis or design. It must be noted that no one *IM* can fully describe the strength of an earthquake and there is no "best" choice of *IM* for all structural systems. "It is inherently impossible to describe a complex phenomenon by a single number, a great deal of information is inevitably missed when it is attempted" (Housner and Jennings 1982). For this reason, many, not just one or two *IMs* have to be considered in any seismic analysis. Some *IMs* work better than others in predicting different types of responses based on their respective structure types. The work conducted by Riddel (2007) is motivated partly to finding the most suitable *IMs* for the structural systems of interest. Naturally, it is desirable to have the *IM* of choice to be correlated to the response; meaning, if the *IM* were to increase or decrease in strength, the response should follow the same pattern

respectively. Using several *IMs* for predicting the responses of nonlinear base isolated structures is not something that has been widely researched.

It is neither practical nor convenient to consider all *IMs* as a response predictor even though this method will probably result in the most accurate predicted response. The reason why this is inconvenient is largely due to the high level of correlation between various *IMs*. For example, the Arias intensity, to be later defined, for an un-damped case should be highly correlated to the earthquake power index proposed by Housner (1975). This is due to the fact that both *IMs* use an acceleration squared quantity in their calculation. Statistically speaking, if the same variables are used in various *IMs*, those *IMs* will be highly correlated to one another. Failure to account for correlation amongst *IMs* is likely to result in inflated estimates of the variance explained. Another reason why only a subset of *IMs* is generally used as opposed to the entire set has to do with feasibility. If one were to consider the computational effort in modelling the response of a base isolated structure, it would require much more computation to use 20 *IMs* as opposed to say, four or five. Again the four or five *IMs* may include most of the necessary information present in the 20 *IMs* of interest. For these reasons, one must determine which *IMs* are the most suitable for the given application.

Previous studies on this subject concluded that acceleration-related indices are the best for rigid systems, velocity-related indices are better for intermediate-frequency systems, and displacement-related indices are better for flexible systems, and some velocity related indices are also effective for the low-frequency region subject (Riddel 2007). Here, the terms indices and measures are interchangeable. For the sake of simplicity, peak ground motion parameters are strongly recommended as *IMs*. This is due to the fact that acquiring these parameters can be done with relative ease; and given their effectiveness, they make good choices for *IMs*. For example,

according to the aforementioned study (Riddel 2007), Housner's intensity was shown to be a poor predictor variable, but the Arias intensity, being an acceleration-related index, is a good indicator when the response of rigid systems is considered. Arias intensity is used as one of the *IMs* in this study.

It is unclear which *IMs* are best for predicting the response of non linear base isolated buildings. Many studies show that peak ground velocity is one of the more effective *IMs* for predicting the responses of such buildings (Ryan and Chopra 2004a 2005b). For linear systems, this tends to be easier by using different regions of frequency spectrum (Riddel 2007). However, for nonlinear structures, such as ones considered in this thesis, the selection process is unclear. The complete set of *IMs* used in this study consists of intensity indices presented in earlier studies (Riddel, 2007; Narasimhan *et al.,* 2008). This set is a summary of the intensity indices shown by Riddel (2007). The simplest index of intensity is its maximum or peak value. Peak ground acceleration (*PGA)*, peak ground velocity *(PGV),* and peak ground displacement *(PGD)* are all peak values of the time histories of ground motion and are shown below.

$$PGA, PGV, PGD = \max\left(|a(t), v(t), d(t)|\right) \tag{2.1.1}$$

Where, *a(t), v(t)*, and *d(t)* are the ground acceleration, velocity, and displacement respectively. It should be noted that all of the *IMs* presented subsequently, either directly or indirectly, use *a(t), v(t)*, and *d(t)*.

In order to describe the damage potential of earthquakes, energy based *IMs* are used. The intensity of the ground motion defined by Arias (1970) is the integral of the squared acceleration, which he interpreted as the sum of the energies dissipated, per unit mass, by a population of damped oscillators of all natural frequencies $(0 < \omega < \infty)$:

$$I_A = \frac{\pi}{2g}\int_0^{t_f}[a(t)]^2 dt \tag{2.1.2}$$

Where, t_f is the total duration of the ground motion, and g is the acceleration of gravity ($9.8\,m/s^2$). The Arias intensity is a good measure of energy input capabilities of the ground motion.

The following *IMs* are integral squared ground motion quantities. Nau and Hall (1984) presented these as normalizing factors for both elastic and inelastic spectra, which are closely related to the Arias intensity (see Eq. 2.1.2).

$$E_a = \int_0^{t_f}[a(t)]^2 dt \tag{2.1.3}$$

$$E_v = \int_0^{t_f}[v(t)]^2 dt \tag{2.1.4}$$

$$E_d = \int_0^{t_f}[d(t)]^2 dt \tag{2.1.5}$$

The roots square of Eqs. 2.1.3, 2.1.4, and 2.1.5 have been proposed by Housner (1970) and serve as alternate intensity measures.

$$a_{rs} = \sqrt{E_a} \tag{2.1.6}$$

$$v_{rs} = \sqrt{E_v} \tag{2.1.7}$$

$$d_{rs} = \sqrt{E_d} \tag{2.1.8}$$

Housner (1975) argued that a measure of seismic destructiveness could be given by the mean-square value of the acceleration history, which he termed “earthquake power index.” This index can be related to acceleration, velocity, and displacement and is given by,

$$P_a = \frac{1}{t_d}\int_{t_5}^{t_{95}}[a(t)]^2 dt \tag{2.1.9}$$

$$P_v = \frac{1}{t_d}\int_{t_5}^{t_{95}}[v(t)]^2 dt \tag{2.1.10}$$

$$P_d = \frac{1}{t_d}\int_{t_5}^{t_{95}}[d(t)]^2 dt \tag{2.1.11}$$

Where t_{95} and t_5 are the limits of the strong portion of motion and $t_d = t_{95} - t_5$ is selected as the significant duration of ground motion (Trifunac and Brady 1975). The interval between instants t_5 and t_{95} at which 5% and 95% of the total integral in Eqs. 2.1.9, 2.1.10, and 2.1.11 are attained, respectively.

The root-mean-square of the strong phase of the ground motion was introduced in earthquake engineering by Housner and Jennings (1964). The root-mean-square values directly correspond to the power indices.

$$a_{rms} = \sqrt{P_a} \tag{2.1.12}$$

$$v_{rms} = \sqrt{P_v} \tag{2.1.13}$$

$$d_{rms} = \sqrt{P_d} \tag{2.1.14}$$

The "potential destructiveness" of an earthquake was defined by Araya and Saragoni (1980) and is in the form

$$P_D = \frac{I_A}{v_0} \tag{2.1.15}$$

Where, I_A is the Arias intensity defined in Eq. 2.1.2 and v_0 is the number of zero crossings per unit of time of the accelerogram. The significance of this *IM* is the use of the frequency content of the acceleration record through the parameter v_0. As v_0 decreases, the index accounts for shifting of the record power to the intermediate-frequency range. A simple relationship was

proposed by Park *et al.* (1985) between the destructiveness of ground motions; it was expressed in terms of the "characteristic intensity,"

$$I_c = a_{rms}^{1.5} t_d^{0.5} \tag{2.1.16}$$

Fajfar *et al.* (1990) proposed the scaling parameter,

$$I_F = PGV \cdot t_d^{0.25} \tag{2.1.17}$$

I_F combines *PGV* and the duration of the ground motion as a measure of ground motion capacity to damage structures with fundamental periods in the velocity-sensitive region. Riddell and Garcia (2001) recommended a set of *IMs* which would minimize the dispersion of the hysteretic energy-dissipation spectra of inelastic single degree of freedom (SDOF) systems. These measures were given by,

$$I_a = PGA \cdot t_d^{1/3} \tag{2.1.18}$$

$$I_v = PGV^{2/3} \cdot t_d^{1/3} \tag{2.1.19}$$

$$I_d = PGD \cdot t_d^{1/3} \tag{2.1.20}$$

There are many *IMs* defined in various publications based on physics or statistical analysis. It should be noted again that the ones presented have not been chosen for means of finding the most effective one(s) but rather to determine the efficiency of selecting a combination of *IMs* based on statistical evaluation. All the measures presented are correlated to one another in some respect which naturally makes some of the *IMs* redundant. For this reason, part of this study involves a correlation analysis between the *IMs*, which will help in developing the proposed prediction model for non linear base isolated structures.

2.2 Models for Non Linear Base Isolated Structures

Base isolation has become a major feature of structural design in the past couple of decades and is now a widely accepted seismic protection system. The purpose of seismic base isolation is to mitigate the effects of the earthquake by *isolating* the structure and its contents from potentially dangerous ground motion, especially in the frequency range where the building is most affected (Skinner *et al.*, 1993; Naeim and Kelly, 1999). The concept of base isolation is one in which flexibility and energy dissipation capacity is provided by a specially designed isolation system between the superstructure and foundation (Nagarajaiah *et al.,* 1990). There are several types of base-isolated systems in use today. These include friction pendulum, lead rubber, and elastomeric bearings shown in Fig. 2.1. Figure 2.1 also shows the force displacement behaviour for the aforementioned systems. The most popular amongst them, also utilized in this study, is the lead rubber bearing (LRB). This isolation system usually comprises of a lead plug within a rubber unit, known as lead-rubber bearings. In the design of a structure with LRBs, the nonlinear behaviour is restricted to the base, and the superstructure is assumed to be elastic at all times. This allows the designer to design the superstructure elements such as beams and columns using the elements of linear elasticity (elementary mechanics approaches), and the complexities are limited to the isolation layer only.

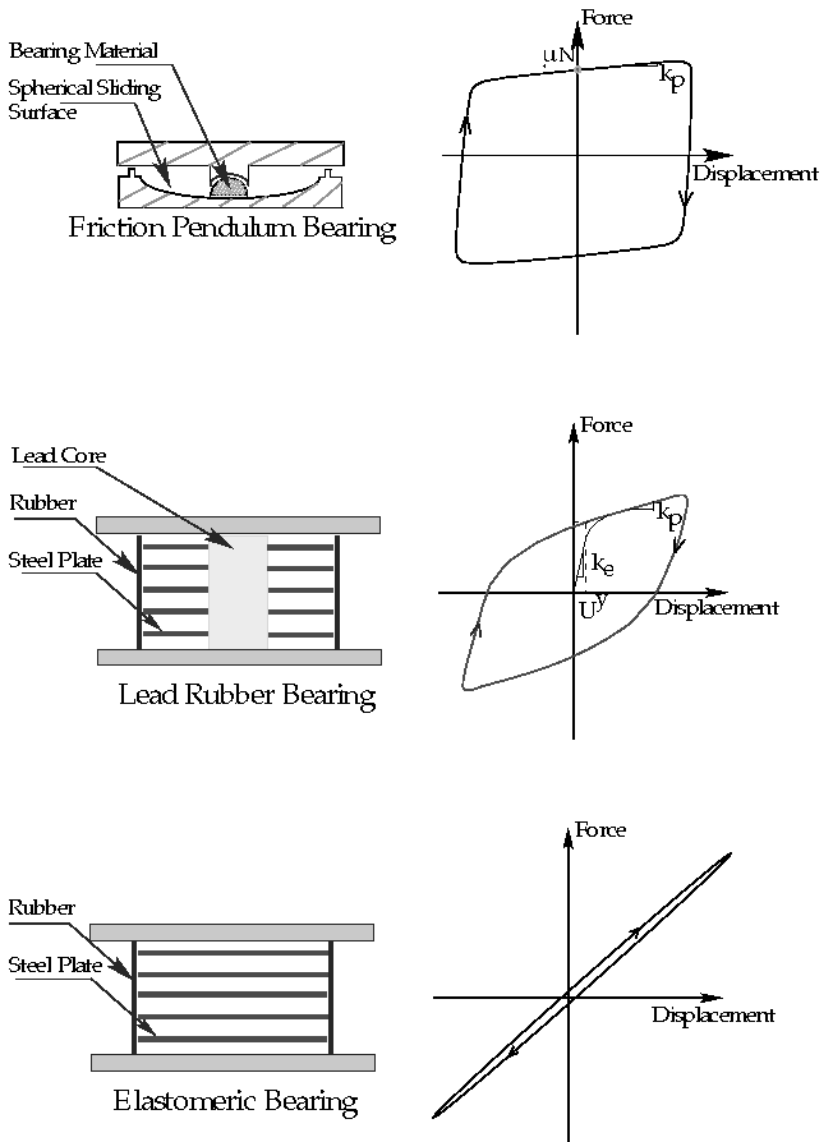

Figure 2.1. Force displacement characteristics of isolation bearings

This LRB system is capable of resisting the weight of the structure using the steel reinforced elastomer, while simultaneously allowing the foundation to move with the ground during an earthquake and isolating the superstructure from the acceleration, resonant frequencies, and the impulsive energy of the earthquake (Jacobs 2008). The lead core within the rubber unit provides energy dissipation and structural support while the rubber provides restoring force or recentering capacity to the system. The ultimate engineering goal of the entire system is to

reduce interstory drifts and floor accelerations. This will ensure that damage is limited or completely avoided for both the structure and its contents.

2.2.1 Uni-axial Base Isolated Model

The first structural model that is considered for this study is the Uni-Axial Base Isolated (UBI) model. The nonlinearities in the isolation system, namely the isolation force, are modelled using hysteresis. The most popular used hysteresis model for the isolation system is the Bouc-Wen model (Park *et al.,* 1986). This model has been used extensively by several researchers (Nagarajaiah, 1991, Romallo *et al.,* Narasimhan *et al.,* 2006). Since the lead plug is the nonlinear element in the isolation system, it is possible to separate the linear elastic part (provided by the rubber surrounding the lead plug) and the nonlinear part and treat them separately in the equations of motion. It must be noted that the nonlinearities are only present in the isolation layer, whether in the uniaxial or biaxial case.

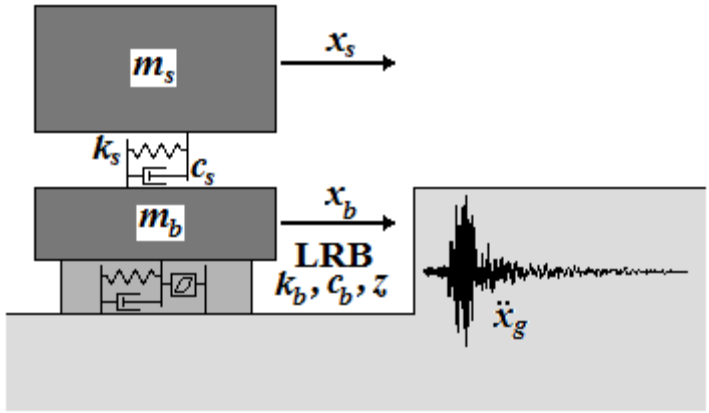

Figure 2.2. 2DOF uniaxial nonlinear base isolated model

The equation of motion for the nonlinear base isolated system shown in Fig. 2.2 can be summarized as,

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{\Lambda} f_H - \mathbf{M}\mathbf{R}\ddot{x}_g \tag{2.2.1}$$

Where f_H is the hysteretic force exerted by the LRB lead plug; $\mathbf{\Lambda} = [1\ 0]^{\mathrm{T}}$ gives the location of the LRB force. In $\mathbf{\Lambda}$**,** 1 corresponds to the isolation layer and the zero corresponds to the superstructure. **R** is the influence vector for the ground acceleration $\ddot{x}_g$ (ones at all the degrees of freedom). The mass, damping, and stiffness matrices are given by,

$$\mathbf{M} = \begin{bmatrix} m_{\boldsymbol{b}} & 0 \\ 0 & m_{\boldsymbol{s}} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} c_b + c_s & -c_s \\ -c_s & c_s \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} k_b + k_s & -k_s \\ -k_s & k_s \end{bmatrix} \tag{2.2.2}$$

The terms *'m', 'k',* and *'c'* are the mass, stiffness, and damping coefficients of the structure. The subscripts *'s'* and *'b'* represent the superstructure and base respectively. The hysteretic force of the LRBs is modelled using the Bouc-Wen model (Park *et al.,* 1986). This model includes an evolutionary variable z to account for the hysteretic component of the force f_H which is expressed as,

$$f_H = Q_H z \tag{2.2.3}$$

The evolutionary variable z is obtained by solving the differential equation,

$$D^Y \dot{z} + \gamma z \left|\dot{d}_b\right| |z|^{\eta-1} + \beta \dot{d}_b |z|^{\eta} - \alpha \dot{d}_b = 0 \tag{2.2.4}$$

where β, γ, and η are shape parameters of the hysteresis loop which are considered to be time invariant. D^Y is the yield displacement and d_b is the isolator deformation/displacement. Note that d_b is of primary interest in this analysis. α determines the yield strength of the isolator.

Q_H from Eq. 2.2.3 is defined as,

$$Q_H = (1 - K_r)Q_Y \tag{2.2.5}$$

Where, K_r is the ratio of the post-yield to pre-yield stiffness:

$$K_r = \frac{K_{yield}}{K_{initial}} \tag{2.2.6}$$

and Q_Y is the yield strength of the isolator (from both the lead plug and rubber):

$$Q_Y = \frac{K_{initial}}{\alpha} \tag{2.2.7}$$

The post-yield natural period of the 2DOF structure is defined as,

$$T = 2\pi\sqrt{\frac{m_b+m_s}{K_{yield}}} \tag{2.2.8}$$

Manipulating Eqs. 2.2.6, 2.2.7, and 2.2.8, the following function of *T* is derived,

$$Q_Y = \frac{m_b+m_s}{K_r\alpha}\left(\frac{2\pi}{T}\right)^2 \tag{2.2.9}$$

Equation 2.2.9 can be used to create a spectrum of time periods (presented in Section 3.1.3). It can be noted that the time period is inversely proportional to the yield strength of the isolator. So as time period increases, the strength of the isolator naturally decreases.

2.2.2 Bi-axial Base Isolated Model

The second structural model that is considered for this study is the biaxial base isolated (BBI, see Fig. 2.3) model. The BBI model is essentially the same as the UBI model with an added *y* direction, hence the name *biaxial*. As mentioned earlier, Nagarajaiah and Xiaohong

(2000) conducted a study to evaluate the seismic performance of the base isolated University of Southern California (USC) hospital building during the 1994 Northridge earthquake. They concluded that the base-isolated USC hospital building performed well and reduced the response when compared to a fixed-base structure. The model presented here uses the same nonlinear modelling procedure presented by Nagarajaiah and Xiaohong (2000) and Fuladgar & Shkib (2003).

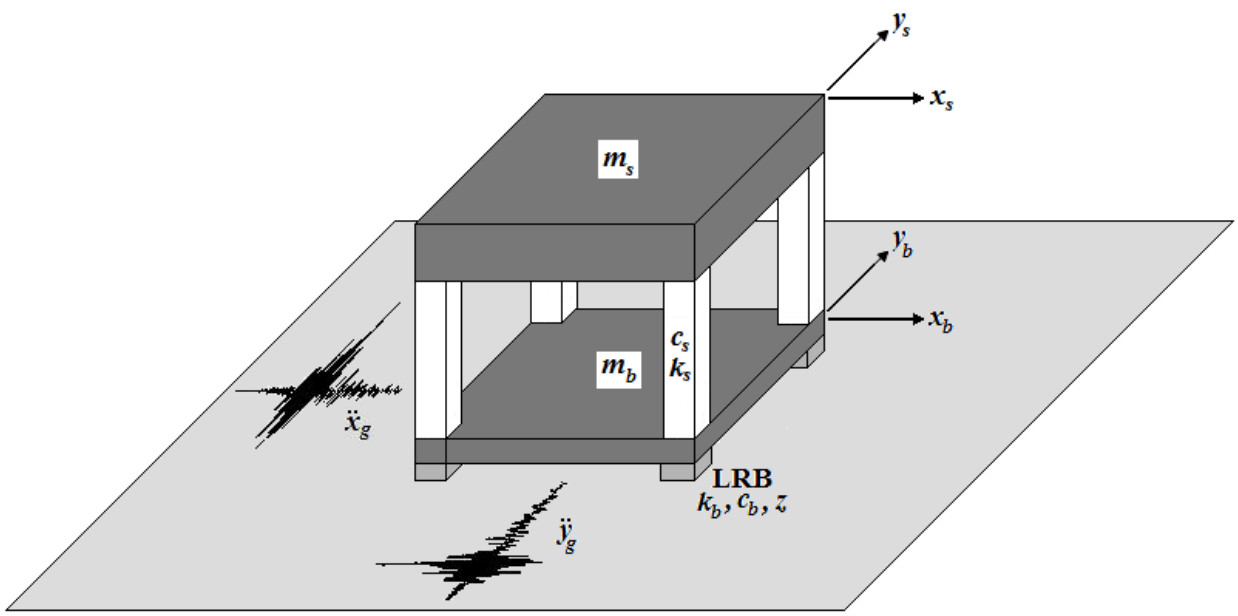

Figure 2.3. 4DOF biaxial nonlinear base isolated model

The equation of motion for this model can be summarized in the same manner as the UBI model (see Eq. 2.2.1). The sizes of the mass, damping, and stiffness matrices differ however, and can be summarized as,

$$\mathbf{M} = \begin{bmatrix} m_b & 0 & 0 & 0 \\ 0 & m_b & 0 & 0 \\ 0 & 0 & m_s & 0 \\ 0 & 0 & 0 & m_s \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} c_b + c_s & 0 & -c_s & 0 \\ 0 & c_b + c_s & 0 & -c_s \\ -c_s & 0 & c_s & 0 \\ 0 & -c_s & 0 & c_s \end{bmatrix},$$

$$\mathbf{K} = \begin{bmatrix} k_b + k_s & 0 & -k_s & 0 \\ 0 & k_b + k_s & 0 & -k_s \\ -k_s & 0 & k_s & 0 \\ 0 & -k_s & 0 & k_s \end{bmatrix} \tag{2.2.10}$$

The differential equations used for the dimensionless *z* variable was also proposed by Park *et al.* (1986). The evolutionary variable *z* is defined in a similar manner to the UBI model, except now there are *x* and *y* directions.

$$f_{H_x} = Q_H z_x \tag{2.2.11}$$

$$f_{H_y} = Q_H z_y \tag{2.2.12}$$

z for each corresponding direction is obtained by solving the differential equations (Narasimhan *et al.,* 2006),

$$D^Y \dot{z}_x + \gamma \left|\dot{d}_{b_x} z_x\right| z_x + \beta\, \dot{d}_{b_x} z_x^2 + \gamma \left|\dot{d}_{b_y} z_y\right| z_x + \beta \dot{d}_{b_y} z_x z_y - \alpha \dot{d}_{b_x} \tag{2.2.13}$$

$$D^Y \dot{z}_y + \gamma \left|\dot{d}_{b_y} z_y\right| z_y + \beta\, \dot{d}_{b_y} z_y^2 + \gamma \left|\dot{d}_{b_x} z_x\right| z_y + \beta \dot{d}_{b_x} z_y z_x - \alpha \dot{d}_{b_y} \tag{2.2.14}$$

The shape parameters η , β, and γ are the same as the UBI model and Eqs. 2.2.5 to 2.2.9 are kept the same and can be applied to the BBI model as well. The 4th and 5th terms in Eqs. 2.2.13 and 2.2.14 involving the products $z_x z_y$ are known as the interaction terms. These interaction terms simulate a more realistic response compared to the UBI model by making the structure stiffer. Without these terms, one will notice Eqs. 2.2.13 and 2.2.14 resemble differential Eq. 2.2.4 for the UBI model.

To summarize, the two models (uniaxial and biaxial) presented behave similarly. The challenge mainly lies in numerically implementing these systems and solving the differential equations in a consistent fashion. Several numerical methods have been utilized to solve the stiff differential equations (Eqs. 2.2.13 and 2.2.14). For example, Nagarajaiah (1990) used the Newmark's constant-average-acceleration method (pseudo-force method) with nonlinear forces being represented as pseudoforces and the solution of the differential equations using the Runge-Kutta method. Narasimhan *et al.* (2003) utilized MATLAB (MATLAB 2007b) and it's toolbox SIMULINK to conduct the nonlinear analysis and generate the structural responses, using ODE45 and ODE23 solvers (MATLAB 2007b). In this paper, the nonlinear models are implemented in SIMULINK as signal-flow diagrams using integrators, multipliers and adders. The linear portion of the system is modelled in state-space. Solving the equations in this manner eliminates the need to use pre-built solvers and iterative procedures (Narasimhan 2008).

2.3 Principal Component Analysis/Regression

Principal component analysis (PCA) is a multivariate technique in which a number of correlated variables are processed through a linear transformation into a set of uncorrelated variables. PCA is a useful statistical technique that has been used in face recognition and image compression software, and for finding patterns in data of high dimension (Jackson 1991). This method is primarily a data analyzing technique that obtains linear transformations of a group of correlated variables such that certain optimal conditions are met (Jackson 1991). The most important of these conditions is that the transformed variables are uncorrelated. Correlation of variables is basically an indication of the strength and direction of a linear relationship between

two variables (Weisberg 1980) and it must be considered if redundant data is to be acknowledged and eliminated.

The question remains, why is having uncorrelated variables more desirable than having correlated ones? The answer is simply that uncorrelated variables can be discarded easily if they are redundant. More specifically, this means that when the data is correlated, redundant information is hard to locate and discard but when data is transformed into an uncorrelated set, redundant information is a lot easier to identify and discard. Additionally, failure to account for correlation amongst variables may result in inflated estimates of the variance explained. Specifically, in modelling, it is important to uncorrelate data and identify redundant information because it allows one to find the optimal data retention point necessary to make a good model without sacrificing accuracy.

For purposes of this study, PCA is illustrated by means of a small hypothetical two-variable example. This will show the mechanics and the utility of the tool. Once PCA is explained, principal component regression (PCR), which initially uses PCA, is then studied in detail. PCR is a two step process, which first uses PCA then applies a multivariate linear regression (MLR) procedure. This second step regresses the newly acquired data with the response variable, which in this case is the nonlinear base isolated response. But first, correlation of data sets along with eigenvectors and eigenvalues and their relation to PCA must be explained.

2.3.1 Correlation of Variables

Correlation describes the degree of relationship between two variables in a linear sense. The correlation coefficient between two variables X and Y is defined as,

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \tag{2.3.1}$$

Where, *cov(X,Y)* represents the covariance matrix (later defined in Section 2.3.2), σ_X, σ_Y are the standard deviations of *X* and *Y* respectively and *E* represents the expected values or means of the variables. This correlation coefficient can range anywhere between -1 and 1; 0 describes the absolute lowest degree of relationship between the variables and ± 1 describes the absolute highest degree of relationship. For example, in Fig. 2.4, two mean centered random variables *X* and *Y* with correlation coefficient of 0.88 calculated using Eq. 2.3.1 are plotted. Quantitatively, $\rho = 0.88$ would be considered *strong* correlation between *X* and *Y*.

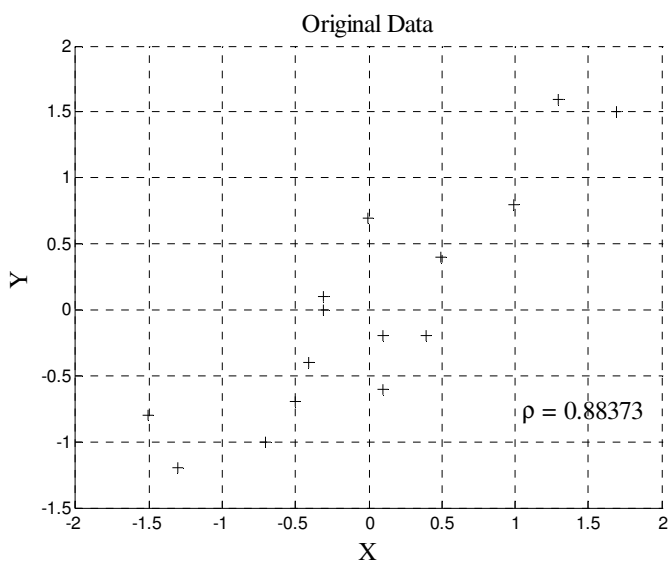

Figure 2.4. Scatter plot of highly correlated variables *X* and *Y* (Jackson 1991)

Completely uncorrelated variables have a correlation coefficient of 0, meaning that their covariance is 0, $E(XY) - E(X)E(Y) = 0$. This however does not mean that the variables are independent. The concept of independence is looked at in Section 2.4 and its significance explained in regards to independent component analysis.

2.3.2 Eigenvectors and Eigenvalues and their Relation to Principal Component Analysis

In this section a brief but necessary overview is presented on eigen-decomposition which includes eigenvectors and eigenvalues and their significance in relation to PCA. Eigenvectors and eigenvalues are vectors and numbers associated with square matrices, and together they provide the eigen-decomposition of a matrix which analyzes the structure of this matrix (Abdi 2007). In relation to this, PCA is obtained from the eigen-decomposition of a covariance matrix (defined in Section 2.3.3) and gives the least square estimate of the original data matrix. This means that the sum of the squares of the residuals has the least possible value. A residual is the difference between an observed value and the value given by a model.

The eigenvector of a matrix $\mathbf{A}$ is given by $\mathbf{u}$ and satisfies the following equation:

$$\mathbf{Au} = \lambda\mathbf{u} \tag{2.3.2}$$

This can be rewritten as follows:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \tag{2.3.3}$$

where λ is a scalar called the eigenvalue associate to the eigenvector $\mathbf{u}$.

For example, the matrix:

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}$$

has eigenvectors $\mathbf{u_1} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ with eigenvalue $\lambda_1 = 4$ and $\mathbf{u_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ with eigenvalue $\lambda_2 = -1$. This is calculated by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$.

In PCA, $m \times n$ matrix $\mathbf{X}$ is linearly transformed into *scores* such that these scores *explain* as much of the variance of $\mathbf{X}$ as possible, and such that the scores are uncorrelated to one another. This linear transformation is defined as follows:

$$\mathbf{Z} = \mathbf{XP} \tag{2.3.4}$$

Where $\mathbf{Z}$ is known as the scores and $\mathbf{P}$ are the eigenvectors. This transformation must satisfy the following relationships:

$$\mathbf{Z}^{\mathbf{T}}\mathbf{Z} = \mathbf{P}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{XP} \tag{2.3.5}$$

$$\mathbf{P}^{\mathbf{T}}\mathbf{P} = \mathbf{I} \tag{2.3.6}$$

$\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$ is a diagonal matrix (*i.e.*, $\mathbf{Z}$ and $\mathbf{P}$ are orthogonal: matrix whose transpose is its inverse). There are several ways of obtaining $\mathbf{P}$. One possible approach is to use the technique of the Lagrangian multipliers where the constraint from Eq. 2.3.6 is expressed as the multiplication with a diagonal matrix of Lagrangian multipliers denoted $\mathbf{\Lambda}$ in order to give the following expression (Abdi 2007)

$$\mathbf{\Lambda}(\mathbf{P}^{\mathbf{T}}\mathbf{P} - \mathbf{I}) \tag{2.3.7}$$

This transforms Eq. 2.3.5 to the following:

$$\mathcal{L} = \mathbf{Z}^{\mathbf{T}}\mathbf{Z} - \mathbf{\Lambda}(\mathbf{P}^{\mathbf{T}}\mathbf{P} - \mathbf{I}) = \mathbf{P}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{XP} - \mathbf{\Lambda}(\mathbf{P}^{\mathbf{T}}\mathbf{P} - \mathbf{I}) \tag{2.3.8}$$

Taking the derivative of Eq. 2.3.8 yields the following:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = 2\mathbf{X}^{\mathbf{T}}\mathbf{XP} - 2\mathbf{\Lambda}\mathbf{P} \tag{2.3.9}$$

Setting Eq. 2.3.9 to zero will yield the optimum value:

$$\mathbf{X^T X P} - \mathbf{\Lambda P} = 0 \Leftrightarrow \mathbf{X^T X P} = \mathbf{\Lambda P} \tag{2.3.10}$$

Since $\mathbf{\Lambda}$ is diagonal, it is clear that Eq. 2.3.10 resembles Eq. 2.3.2 which is an eigen-decomposition problem. This means that $\mathbf{\Lambda}$ is the matrix of eigenvalues of the covariance matrix of $\mathbf{X}$, ($\mathbf{X^T X}$) ordered from the largest to the smallest and $\mathbf{P}$ is the matrix of eigenvectors of $\mathbf{X^T X}$. The covariance of the scores, $\mathbf{Z}$, is equal to the eigenvalues:

$$\mathbf{Z^T Z} = \mathbf{\Lambda^{1/2} P^T P \Lambda^{1/2}} = \mathbf{\Lambda} \tag{2.3.11}$$

Since the sum of the eigenvalues is equal to the trace of $\mathbf{X^T X}$, the first score, Z_1, extracts as much of the variance of the original data as possible, the second score, Z_2, extracts as much of the variance left unexplained by the first score, and so on for the remaining scores (Abdi 2007). This theory can be used to further explain PCA in the following section.

2.3.3 Principal Component Analysis and Hypothetical Example

The basic idea of PCA is to transform a set of correlated random variables causing the problem of collinearity into a new set of uncorrelated ones. The transformation is achieved using the eigenvector of the covariance matrix of random variables outlined in Section 2.3.2. The problem of collinearity is approached by eliminating those dimensions of the independent variable space that are causing the collinearity problem. This is similar, in concept, to dropping an independent variable to contribute meaningful information on the dependent response variable. The basic steps involved in the PCA methodology are best explained through a numerical example.

Step 1: Acquire Data Matrix

This original data are presented in Table 2.1 under the *X* and *Y* columns.

Table 2.1. Data for example presented by Jackson (1991)

Obs. No.	*X*	*Y*	$X - \mu_X$	$Y - \mu_Y$	*Z1*	*Z2*
1	10	10.7	-0.01	0.7	0.4802	0.5093
2	10.4	9.8	0.39	-0.2	0.145	-0.4167
3	9.7	10	-0.31	0	-0.2211	0.2125
4	9.7	10.1	-0.31	0.1	-0.1518	0.2846
5	11.7	11.5	1.69	1.5	2.2603	-0.0918
6	11	10.8	0.99	0.8	1.2705	-0.1115
7	8.7	8.8	-1.31	-1.2	-1.7736	0.0401
8	9.5	9.3	-0.51	-0.7	-0.8504	-0.1537
9	10.1	9.4	0.09	-0.6	-0.3484	-0.4973
10	9.6	9.6	-0.41	-0.4	-0.5704	-0.0066
11	10.5	10.4	0.49	0.4	0.6329	-0.0534
12	9.3	9	-0.71	-1	-1.2024	-0.2314
13	11.3	11.6	1.29	1.6	2.0412	0.2575
14	10.1	9.8	0.09	-0.2	-0.0713	-0.2089
15	8.5	9.2	-1.51	-0.8	-1.6407	0.4671
μ	10.01	10.00				

Step 2: Subtracting the mean

For PCA to work properly, the mean must be subtracted from each of the data dimensions. This is also presented in Table 2.1 under the $X - \mu_X$ and $Y - \mu_Y$ columns. The plot of the mean centered data is shown in Fig. 2.4. The correlation coefficient is also shown and it can be noted that the original mean centered data are highly correlated to one another, leading to the problem of collinearity.

Step 3: Calculate the covariance matrix

The normalized covariance of a data set can simply be calculated by the following formula,

$$\mathbf{S} = \frac{[X-\mu_X \quad Y-\mu_Y]'[X-\mu_X \quad Y-\mu_Y]}{(n-1)} \tag{2.3.12}$$

Since the data in this case is two dimensional, the covariance matrix is 2×2. In this example, the covariance matrix is

$$\mathbf{S} = \begin{bmatrix} 0.7878 & 0.6721 \\ 0.6721 & 0.7343 \end{bmatrix}$$

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix

Since the covariance matrix is square, the eigenvalues and eigenvectors can easily be obtained. Solving Eq. 2.3.10 for the eigenvectors **P** and eigenvalues **Λ** yields:

$$\mathbf{P} = \begin{bmatrix} 0.6929 & -0.7210 \\ -0.7210 & -0.6929 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 1.4337 \\ 0.0884 \end{bmatrix}$$

These values provide good information about the original data. The eigenvectors are perpendicular to one another (see Fig. 2.5). One goes through the middle of the points, similar to drawing a line of best fit. This eigenvector shows the relation between the two data sets along that line. The second eigenvector is perpendicular to the first and shows that all the points follow the main line, but are off to the side of the main line. Geometrically, the eigenvectors are the principal axis rotation of the original coordinate axes *X* and *Y* about their means. In this case the means are both zero because the data were mean centered.

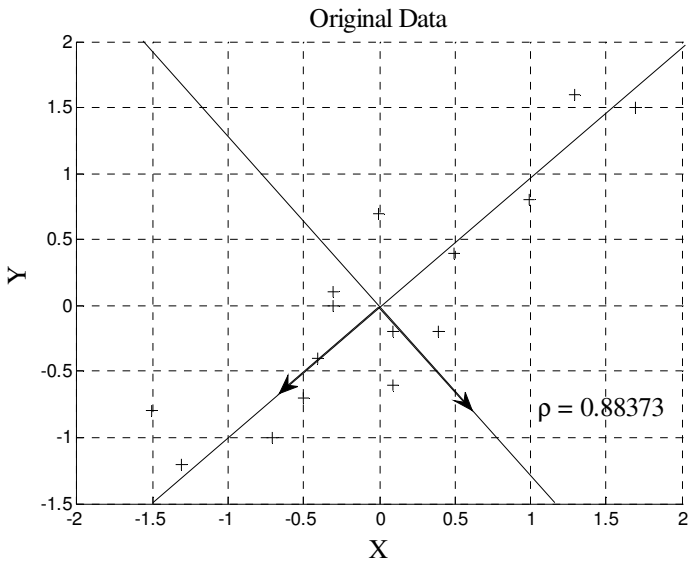

Figure 2.5. Scatter plot of correlated variable *X* and *Y* showing eigenvectors

Step 5: Deriving the new data set Z

Once the eigenvectors **P** are determined, applying Eq. 2.3.4 will yield the new uncorrelated scores **Z**. The new uncorrelated scores *Z1* and *Z2* are presented in Table 2.1 and plotted in Fig. 2.6. Note that the coefficient of correlation is now zero.

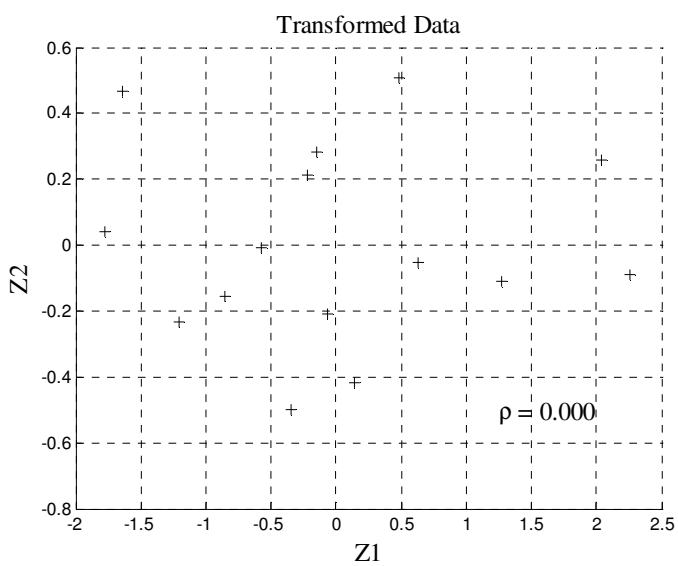

Figure 2.6. New uncorrelated scores, *Z1* and *Z2*

2.3.4 Dimensionality Reduction and Multiple Linear Regression

The next step in PCR is dimensionality reduction, followed by multiple linear regression (MLR). Dimensionality reduction basically consists of eliminating the scores which do not contribute to the variance explained of the original data matrix **X**. Since the scores, **Z**, are now uncorrelated to one another, the elimination of unwanted variables corresponding to relatively low eigenvalues, will not result in any substantial loss of information. Once the $m \times n$ matrix, **Z**, is reduced to a $m \times p$ matrix ($p < n$), MLR can be applied. In MLR, the newly reduced scores, **Z** ($m \times p$), are used to model a single response variable, Y (Weisberg 1980). The model is specified by a linear equation,

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_p Z_p + \varepsilon \tag{2.3.13}$$

where β's are unknown parameters and the ε's are statistical standard errors or residuals. In the case of the example presented in Section 2.3.3, Eq. 2.3.13 becomes the equation of a two dimensional plane shown below.

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon$$

At this point, MLR can be applied to the above model. Regression of this equation will yield the β coefficients which do not have a physical meaning in their present form due to the fact that regression was done between transformed data, **Z**, and real data, Y. β's have to be converted back to the original space. This is explained in detail in Section 3.3.1. Readers are referred to Weisberg (1980) and Jackson (1991) for details on multiple linear regression which is not presented here for the sake of brevity. This thesis utilizes MATLAB's *regress* command for this regression step (MATLAB 2007b).

2.4 Independent Component Analysis/Regression

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals (Hyvarinen *et al.*, 2001). In other words, ICA is a method for finding components from multivariate statistical data, something that this thesis sets out to do. It has been tested that ICA typically performs best with large sets of data (Gustafsson 2005). Gustafsson (2005) concluded that when the number of training examples decreases, the performance of the ICA algorithm decrease significantly both in terms of prediction of the model and recovery of the original data. Data in ICA are assumed to be linear or nonlinear mixtures of some unknown latent variables. These latent variables are similar to the scores, **Z**, determined through the PCA transformation (see Eq. 2.3.4). In ICA, the latent variables are assumed to be nongaussian and mutually independent and are therefore called the independent components of the observed data.

ICA can be seen as an extension to PCA and other factor analysis techniques, but ICA has been argued to perform better in some cases (Gustafsson 2005). Gustafsson (2005) reported that the goodness of fit of several regression methods including two PCA motivated methods yielded inferior results to two ICA motivated methods. For this reason, it is important to test ICA motivated algorithms in comparison to PCA techniques for predicting nonlinear base isolated responses.

2.4.1 Motivation for Independent Component Analysis

The motivation behind ICA stems from a problem called the *cocktail-party problem* (Hyvarinen *et al.*, 2001). The scenario is as follows. There are two people in a room speaking simultaneously. Two microphones, which are held at different locations, are picking up the

speech signals. The question is, could the original speech signals be estimated using only the recorded signals in the microphones? This is what ICA sets out to solve. Mathematically, this problem can be presented in the following form:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \tag{2.4.1}$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \tag{2.4.2}$$

where $x_1(t)$ and $x_2(t)$ are the recorded time signals by the microphones; $s_1(t)$ and $s_2(t)$ are the weighted sums of the speech signals projected by the two speakers; and a_{11}, a_{12}, a_{21}, and a_{22} are some unknown parameters that depend on the distances of the microphones from the speakers. So restating the earlier question, could the speech signals $s_1(t)$ and $s_2(t)$ be estimated using only $x_1(t)$ and $x_2(t)$? Before presenting details as to how ICA attempts to solve this problem, independence of variables and key principles of ICA must be studied first.

2.4.2 Independence of Variables

Two variables, *X* and *Y* are said to be independent if information on the value of *X* does not provide any information on the value of *Y*, and vice versa. This can be defined by their respective probability densities. If one were to denote the joint probability density function (PDF) of *X* and *Y* as *p(X,Y)* and the marginal PDF's of *X* and *Y* as *p(X)* and *p(Y)*, then independence of two variables can be defined as,

$$p(X,Y) = p(X)p(Y) \tag{2.4.3}$$

which is a product of their marginal densities. This means, given two functions, h_1 and h_2, the following relationship also holds.

$$E\{h_1(X)h_2(Y)\} = E\{h_1(X)\}E\{h_2(Y)\} \tag{2.4.4}$$

where E is the expected or mean value.

2.4.3 Definition and Principles of Independent Component Analysis

To best define ICA, a statistical *latent variables* model must be used (Jutten and Herault, 1991; Comon, 1994). In this model, it is assumed that n linear mixtures $x_1, \dots, x_n$ of n independent components is observed.

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n, \text{ for all } j. \tag{2.4.5}$$

It should be noted that the time index t is now dropped because in the ICA model it is assumed that each mixture x_j as well as each independent component s_k is a random variable, instead of a sample in time (Hyvarinen *et al.*, 2001). Re-writing Eq. 2.4.5 in vector form, the model can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.4.6}$$

Equation 2.4.6 is known as the ICA model. **s** must be determined in such a way that its components s_i are statistically independent to one another. The fundamental limitation of ICA is that the variables must be nongaussian (non-normal) for ICA to be possible (Hyvarinen *et al.*, 2001). This means if one or more of the original variables have normal distributions, an ICA algorithm will fail to extract the independent components. After the matrix **A** is estimated, the independent components can be calculated simply by the following equation:

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \tag{2.4.7}$$

In order to illustrate ICA with regards to Eqs. 2.4.6 and 2.4.7, consider two independent components, s_1 and s_2, that have uniform distributions with a continuous uniform distribution interval [0, 1]. After mean centering, s_1 and s_2 take the shape shown in Fig. 2.7. It can be seen that knowing the value of s_1 does not in any way help in determining the value of s_2. This is the definition of independence.

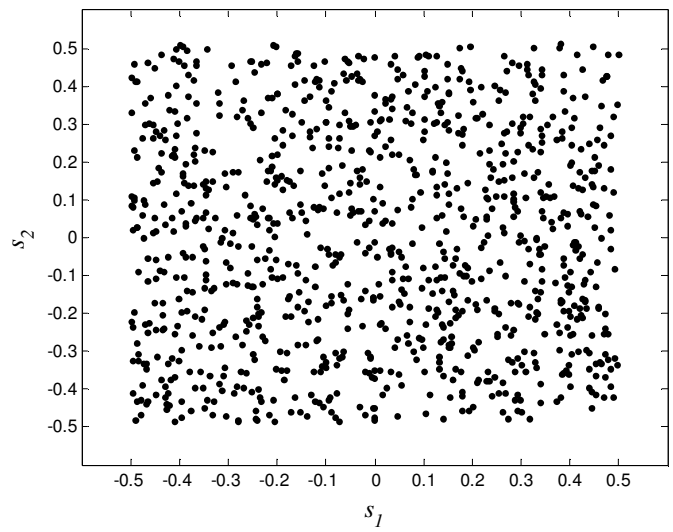

Figure 2.7. Two independent components with uniform distributions

Now assuming that the mixing matrix **A** is given by the following hypothetical example,

$$\mathbf{A} = \begin{bmatrix} \mathbf{2} & \mathbf{3} \\ \mathbf{2} & \mathbf{1} \end{bmatrix}$$

yields the mixed variables x_1 and x_2 by applying Eq. 2.4.6. The mixed data has a distribution presented in Fig. 2.8. The random variables are now not independent anymore. This means that information about the value of x_1 *will* help in determining the value of x_2. For example, consider the maximum value of x_1; this clearly shows what the corresponding value of x_2 is.

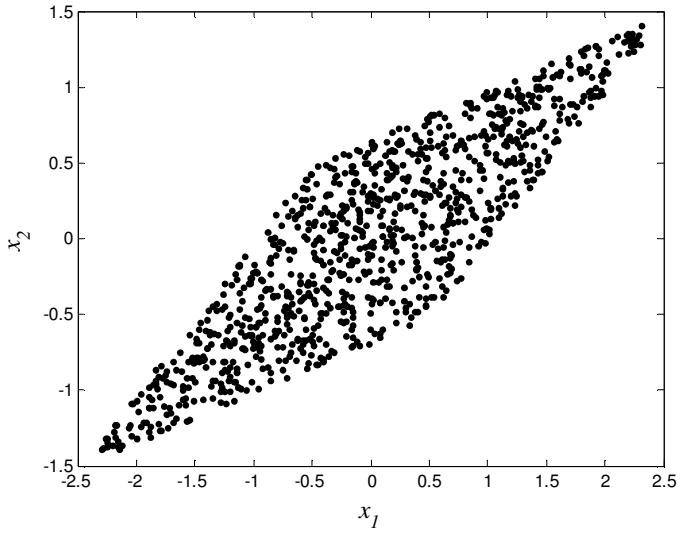

Figure 2.8. Joint distribution of observed mixtures x_1 and x_2

As mentioned earlier, the fundamental restriction in ICA is that the independent components, **s**, must be as nongaussian as possible. The estimation of the independent components is not possible without non-gaussianity. If non-gaussianity can be defined mathematically, finding the mixing matrix, **A**, would become an optimization problem, similar to PCA. The classical method of non-gaussianity is calculated through a measure known as kurtosis. Kurtosis of a variable *y* is defined by

$$\text{kurt(y)} = E\{y^4\} - 3(E\{y^2\})^2 \tag{2.4.8}$$

One of the main assumptions of ICA is that the standard deviation of the independent variables is equal to one ($E\{y^2\} = 1$). This simplifies the right side of Eq. 2.4.8 to $E\{y^4\} - 3$. Typically, non-gaussianity is measured by the absolute value of kurtosis. If, s_1 and s_2 are two independent random variables, the following relationships hold true.

$$\text{kurt}(s_1 + s_2) = \text{kurt}(s_1) + \text{kurt}(s_2) \tag{2.4.9}$$

and

$$\mathrm{kurt}(\alpha s_1) = \alpha^4 \mathrm{kurt}(s_1) \tag{2.4.10}$$

Considering a two dimensional model, $\mathbf{x} = \mathbf{As}$, s_1 and s_2 would have kurtosis values kurt(s_1) and kurt(s_2) respectively both of which are nonzero (kurtosis is zero for a gaussian random variable). One of the independent components would be given by, $y = \mathbf{w^T x}$, where $\mathbf{w^T} = \mathbf{A^{-1}}$. If the transformation, $\mathbf{z} = \mathbf{A^T w}$, is made, then $y = \mathbf{w^T x} = \mathbf{w^T A s} = \mathbf{z^T s} = z_1 s_1 + z_2 s_2$. Based on Eqs. 2.4.9 and 2.4.10, $\mathrm{kurt}(y) = \mathrm{kurt}(z_1 s_1) + \mathrm{kurt}(z_2 s_2) = z_1^4 \mathrm{kurt}(s_1) + z_2^4 \mathrm{kurt}(s_2)$. Based on the assumption that the variance of *y* is equal to 1, this would put a constraint on **z**: $E\{y^2\} = z_1^2 + z_2^2 = 1$. Geometrically, this means that **z** is constrained to the unit circle on the two dimensional plane. Therefore the function, $|\mathrm{kurt}(y)| = |z_1^4 \mathrm{kurt}(s_1) + z_2^4 \mathrm{kurt}(s_2)|$ must be optimized on the unit circle. Once *y* is determined, the mixing matrix can then be determined.

Kurtosis is one of the ways to compute the independent components of ICA. Other methods outlined by Hyvarinen *et al.* (2001) include negontropy, mutual information, maximum likelihood, and the infomax principle which all use different criteria for optimization, conceptually similar to using kurtosis. Some are more suitable for certain data than others, and details of these methods are beyond the scope of this thesis. For purposes of this thesis, a MATLAB defined function *aci*, which uses the Comon's method of optimizing cumulants (Comon 1994), is used to determine the mixing matrix. An ICA motivated regression method known as sorted-input independent component regression is studied next.

2.4.4 Sorted-Input Independent Component Analysis

Whitening of the observed variables is recommended as a pre-processing strategy in ICA (Gustafsson, 2005; Hyvarinen *et al,*. 2001). Whitening is similar to PCA; it transforms the correlated components to uncorrelated ones with their variances equal to unity. For this reason SI-ICR method involves a whitening step. The most popular method for whitening is using the eigen-value decomposition, similar to the one presented for PCA (see Section 2.3.2). Essentially, SI-ICR is a PCR-motivated regression tool that uses both PCA whitening techniques as well as the ICA algorithm. The major steps involved in SI-ICR (Gustafsson 2005) are outlined below with minor modifications made for purposes of this thesis.

Step 1: Acquire Data Matrix

Similar to the first step in PCA (Section 2.3.3), the first step is to get a data matrix **X** here as well. The notation is kept similar to the parameters presented in Section 2.3.3 for simplicity.

Step 2: Subtracting the mean

Perform conventional mean centering on all variables of **X**.

Step 3: Estimate the covariance matrix

Based on the *N* number of samples, estimate the covariance matrix using the standard estimate.

$$\hat{\mathbf{C}}_{\mathrm{xx}} = \frac{1}{N-1}\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathbf{X}_{\mathrm{n}}\mathbf{X}_{\mathrm{n}}^{\mathrm{T}} \tag{2.4.11}$$

Step 3: Perform PCA

Compute eigenvectors $\boldsymbol{u}_k$, and eigenvalues λ_k (see Section 2.3.3, step 4), where $k = 1,2,\ldots,K_e$ and K_e is the number of nonzero eigenvalues.

Step 4: Dimensionality reduction

Reduce dimensionality from K to H by transforming the input $\mathbf{X}$ into the reduced vector

$$\mathbf{X}_{red} = \mathbf{\Lambda}^{-1/2}\mathbf{U}_H'\mathbf{X} \tag{2.4.12}$$

Where, $\mathbf{U}_H$ is a $K \times H$ matrix, $\mathbf{U}_H = [u_1, u_2, \dots, u_H]$, and $\mathbf{\Lambda}^{-1/2}$ is a diagonal $H \times H$ matrix with diagonal elements $\mathbf{\Lambda}^{-1/2} = 1/\sqrt{\lambda_h}$, $h = 1, 2, \dots, H$.

Step 5: Perform ICA algorithm

This step extracts the most independent components from the $\mathbf{X}_{red}$ matrix.

$$\mathbf{X}_{red} = \mathbf{A}_{red}\mathbf{s} \tag{2.4.13}$$

$\mathbf{A}_{red}$ is the $H \times H$ mixing matrix and $\mathbf{s}$ contains the independent components of $\mathbf{X}_{red}$.

Step 6: Estimate the linear transformation matrix

In this step, $\widehat{\mathbf{A}}$, which is the $K \times H$ linear transformation matrix that relates $\hat{\mathbf{s}}$ to $\mathbf{X}$, is estimated through the following relationship:

$$\widehat{\mathbf{A}} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{A}_{red} \tag{2.4.14}$$

Where,

$$\hat{\mathbf{s}} = \widehat{\mathbf{A}}^{-1}\mathbf{X} \tag{2.4.15}$$

Step 7: Compute the correlation and sort

In this step, the correlation coefficients between $\hat{\mathbf{s}}$ and the response variable $\boldsymbol{y}$ is computed and $\hat{\mathbf{s}}$ is sorted according to decreasing order of correlation. The same is done with the mixing matrix $\widehat{\mathbf{A}}$. Therefore the new sorted scores become

$$\mathbf{Z}_{\boldsymbol{sort}} = \mathbf{A}_{\mathbf{sort}}^{-\mathbf{1}}\mathbf{X} \tag{2.4.16}$$

Where, $\mathbf{A}_{\boldsymbol{sort}}$ is the new mixing matrix according to the new orientation based on the correlation between $\hat{\mathbf{s}}$ and the response variable $\boldsymbol{y}$.

Step 8: Linear regression

Linear regression of the new scores $\mathbf{Z}$ and the response $\boldsymbol{y}$ is performed next. This step is similar to the one presented in the PCR methodology (see Section 2.3.4).

3. METHODOLOGY AND NUMERICAL SIMULATIONS

In this chapter, the problem statement is presented, followed by details of the numerical simulations carried out on the nonlinear base isolated models. In order to simulate the proposed UBI and BBI models, State Space representation of the models is created and then implemented in MATLAB and it's toolbox, SIMULINK. This is followed by a section that presents the details of the ground motion database and the implementation of the earthquake simulations. This section also presents a correlation analysis between responses and proposed *IMs*, as well as the regression models for both PCR and SI-ICR methods. Model predictions for an MDOF structure is proposed using PCR, SI-ICR and peak ground velocity (*PGV)* methods. Finally, PCR and SI-ICR methodology for the UBI and BBI systems, and their respective model predictions are studied.

3.1 State Space Representation of Structural Models and Model Parameters

The differential equations of motion for the dynamical structure can be cast into a set of n first order differential equations in a matrix form, where n is the order of the system. This representation of the differential equations is particularly convenient to handle for systems with multiple inputs and multiple outputs. The relationship between the inputs and the outputs is through a linear operation of the state variables using system matrices. The state variables are the smallest subset of variables that are necessary to represent the solution of the system at any time. State space representation is a very practical and useful tool for what is being proposed and can be conveniently implemented in MATLAB's SIMULINK toolbox to simulate the responses according to the input earthquakes. System matrices are defined in accordance to either the

uniaxial or biaxial model. The details of the numerical implementation of UBI and BBI models are presented next.

General state space representation relating a system of inputs and outputs is of the form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\boldsymbol{u} \tag{3.1.1}$$

$$\boldsymbol{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\boldsymbol{u} \tag{3.1.2}$$

The matrix **A** is a known *feedback* matrix, as its elements determine the *feedback* connections of the realization (Narasimhan 2008). The matrix **B** is known as the input matrix and **C** is known as the output matrix. **D** is known as the transmission matrix. Matrices **A** and **B** are defined by the structure whereas matrices **C** and **D** are user defined. Meaning, matrices **C** and **D** can be changed depending on the desired output. $\dot{\mathbf{x}}$ are the states, $\boldsymbol{y}$ are the desired responses and $\boldsymbol{u}$ is the input vector containing the earthquake as well as the nonlinear forces in the isolation system.

For this system, the equation of motion 2.2.1, is rearranged in a fashion which effectively results in a set of first order differential equations as follows:

$$\ddot{\mathbf{x}} = -\frac{\mathbf{K}}{\mathbf{M}}\mathbf{x} - \frac{\mathbf{C}}{\mathbf{M}}\dot{\mathbf{x}} - \mathbf{R}\ddot{x}_g + \frac{\mathbf{\Lambda}}{\mathbf{M}}f_H \tag{3.1.3}$$

Equation 3.1.3 can be represented in the matrix form

$$\frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \dot{\mathbf{x}}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{K}/\mathbf{M} & -\mathbf{C}/\mathbf{M}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \dot{\mathbf{x}}\end{bmatrix} - \begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}\ddot{x}_g + \begin{bmatrix}\mathbf{0}\\ \mathbf{\Lambda}/\mathbf{M}\end{bmatrix}f_H \tag{3.1.4}$$

Here, $\mathbf{x}$ and $\dot{\mathbf{x}}$ represent the displacement and velocity vector of the system.

3.1.1 State Space Representation for an Uni-axial Model

For a one story UBI model (refer to Fig. 2.2), there are 2 degrees of freedom, the coordinates corresponding to the center of mass of the base and the first floors. The states corresponding to the formulation to be presented here, are the displacements and velocities at each degree of freedom. This means that **A** is a 4×4 matrix, **B** is a 4×2 matrix, and **C** and **D** can be varied depending on the desired output. For example, if the accelerations are to be measured in addition to the already desired states, **C** will be a 6×4 and **D** will be a 6×2. For the purposes of this thesis, **C** and **D** matrices are adapted to measure only the base and the superstructure displacement and velocity. For the UBI model, the **A, B, C,** and **D** matrices are presented below.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{4\times4} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ [-\frac{\mathbf{K}}{\mathbf{M}}] & & [-\frac{\mathbf{C}}{\mathbf{M}}] & \end{bmatrix}_{4\times4}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ -1 & -1/\mathrm{m_b} \end{bmatrix}_{4\times2}$$

$$\mathbf{C} = [\mathbf{I}]_{4\times4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4\times4}$$

$$\mathbf{D} = [\mathbf{Zeros}]_{4\times2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{4\times2}$$

It can be seen that the user defined matrices **C** and **D** measure the superstructure and base displacement and velocity. If the above matrices are put back into Eqs. 3.1.1 and 3.1.2, the following state space representation is obtained:

$$\underbrace{\frac{d}{dt}\begin{bmatrix} x_s \\ x_b \\ \dot{x}_s \\ \dot{x}_b \end{bmatrix}_{4\times1}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ [-\frac{\mathbf{K}}{\mathbf{M}}] & & [-\frac{\mathbf{C}}{\mathbf{M}}] & \end{bmatrix}_{4\times4}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_s \\ x_b \\ \dot{x}_s \\ \dot{x}_b \end{bmatrix}_{4\times1}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ -1 & -1/\mathrm{m_b} \end{bmatrix}_{4\times2}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \ddot{x}_g \\ f_H \end{bmatrix}_{2\times1}}_{\boldsymbol{u}}$$

$$\underbrace{\begin{bmatrix} x_s \\ x_b \\ \dot{x}_s \\ \dot{x}_b \end{bmatrix}_{4\times1}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{4\times4}}_{\mathbf{C}} \underbrace{\begin{bmatrix} x_s \\ x_b \\ \dot{x}_s \\ \dot{x}_b \end{bmatrix}_{4\times1}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}_{4\times2}}_{\mathbf{D}} \underbrace{\begin{bmatrix} \ddot{x}_g \\ f_H \end{bmatrix}_{2\times1}}_{\boldsymbol{u}}$$

The LRB (force f_H) is modelled using the Bouc-Wen model mentioned in Section 2.2.1. Equations 2.2.3 and 2.2.4 are used to find the evolutionary variable z. This procedure implemented in SIMULINK is shown in Fig. A.1 (appendix) and a sample force vs. displacement relationship taken from the 1966 Parkfield Earthquake is shown in Fig. 3.1. Note that the units 'gal' are in cm/s^2.

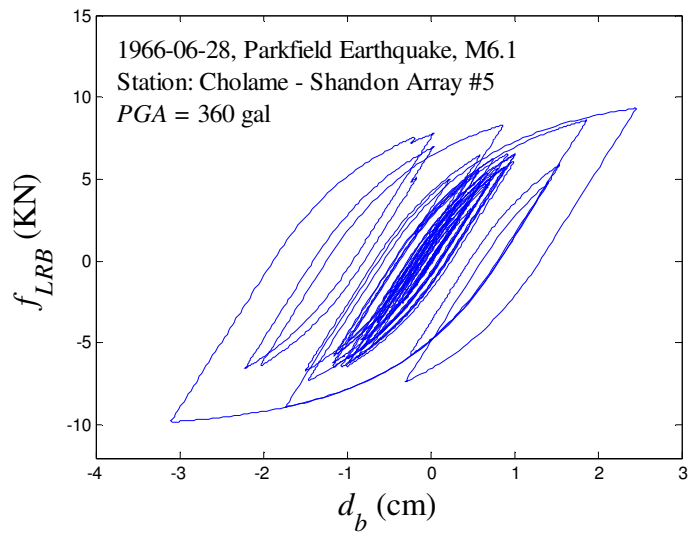

Figure 3.1. Force-deformation behaviour of the LRB (T = 3.0 sec.) for Parkfield 1966

3.1.2 State Space Representation for Bi-axial Model

For the BBI model, there are four degrees of freedom, the superstructure displacement and the base displacement, each in the x and y direction (refer to Fig. 2.3). This means that if one were to calculate displacement and velocity states for each degree of freedom, **A** will be an 8×8 matrix, **B** will be an 8×2 matrix and **C** and **D** are an 8×8 identity matrix and an 8×4 zero matrix respectively. For the BBI model, the **A, B, C,** and **D** matrices are presented below.

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{8\times 8} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \left[-\frac{\mathbf{K}}{\mathbf{M}}\right]_{4\times 4} & & & & \left[-\frac{\mathbf{C}}{\mathbf{M}}\right]_{4\times 4} & & & \end{bmatrix}_{8\times 8}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & -1/\mathrm{m_b} & 0 \\ 0 & -1 & 0 & -1/\mathrm{m_b} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}_{8\times 4}$$

$$\mathbf{C} = [\mathbf{I}]_{8\times 8} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{8\times 8}$$

$$\mathbf{D} = [\mathbf{Zeros}]_{8\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{8\times 4}$$

If the above matrices are substituted into Eqs. 3.1.1 and 3.1.2, it results in the following state space representation:

$$\underbrace{\frac{d}{dt}\begin{bmatrix} x_{bx} \\ x_{by} \\ x_{sx} \\ x_{sy} \\ \dot{x}_{bx} \\ \dot{x}_{by} \\ \dot{x}_{sx} \\ \dot{x}_{sy} \end{bmatrix}_{8\times 1}}_{\dot{\mathbf{x}}} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \left[-\frac{\mathbf{K}}{\mathbf{M}}\right]_{4\times 4} & & & & \left[-\frac{\mathbf{C}}{\mathbf{M}}\right]_{4\times 4} & & & \end{bmatrix}_{8\times 8}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_{bx} \\ x_{by} \\ x_{sx} \\ x_{sy} \\ \dot{x}_{bx} \\ \dot{x}_{by} \\ \dot{x}_{sx} \\ \dot{x}_{sy} \end{bmatrix}_{8\times 1}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & -1/\mathrm{m_b} & 0 \\ 0 & -1 & 0 & -1/\mathrm{m_b} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}_{8\times 4}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \ddot{x}_{gx} \\ f_{H_x} \\ \ddot{x}_{gy} \\ f_{H_y} \end{bmatrix}_{4\times 1}}_{\boldsymbol{u}}$$

$$\underbrace{\begin{bmatrix} x_{bx} \\ x_{by} \\ x_{sx} \\ x_{sy} \\ \dot{x}_{bx} \\ \dot{x}_{by} \\ \dot{x}_{sx} \\ \dot{x}_{sy} \end{bmatrix}_{8\times1}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{8\times8}}_{\mathbf{C}} \underbrace{\begin{bmatrix} x_{bx} \\ x_{by} \\ x_{sx} \\ x_{sy} \\ \dot{x}_{bx} \\ \dot{x}_{by} \\ \dot{x}_{sx} \\ \dot{x}_{sy} \end{bmatrix}_{8\times1}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{8\times4}}_{\mathbf{D}} \underbrace{\begin{bmatrix} \ddot{x}_{gx} \\ f_{H_x} \\ \ddot{x}_{gy} \\ f_{H_y} \end{bmatrix}_{4\times1}}_{u}$$

Similarly, the LRB for the biaxial system is modelled after the Bouc-Wen model mentioned in Section 2.2.1. Equations 2.2.11 to 2.2.14 are used to find the evolutionary variable z for both x and y directions. This procedure implemented in SIMULINK is shown in Fig. A.2.

3.1.3 Structural Parameters for Uni-axial and Bi-axial Models

To be consistent, the structural parameters for both models (UBI and BBI) are kept the same. This way, conclusions about the performance pertaining to PCR and SI-ICR methods can be made in a consistent fashion. The structural parameters for both UBI and BBI systems are presented in Table 3.1. For a detailed description of these quantities, refer to Section 2.2.

Table 3.1. Structural Parameters for both UBI and BBI systems

Parameter	Value	Units
m_b	6,800	kg
m_s	29,485	kg
k_b	232,000	N/m
k_s	11,912,000	N/m
c_b	3,740	$N \cdot s/m$
c_s	23,710	$N \cdot s/m$
D^Y	0.013	m
η	1(UBI), 2(BBI)	-
α	1	-
β	0.5	-
γ	0.5	-
K_r	1/6	-

In order to achieve a spectrum of isolation periods, the ratio, K_r is kept constant and the initial stiffness is varied. For this study, the time period T varies from 1.3 to 5.5 seconds if the ratios of the yield force Q_y to the total weight $W = (m_b + m_s)g$ ranges from 1% to 20%. This means if T = 1.3 to 5.5 is applied to Eq. 2.2.9, Q_y would range from 66,134 N to 3,695 N (Fig. 3.2). The time period in this case is varied logarithmically from 1.3 to 5.5 and divided into 20 equal points. For most cases, as the yield force of the isolator decreases, the base displacement is expected to increase.

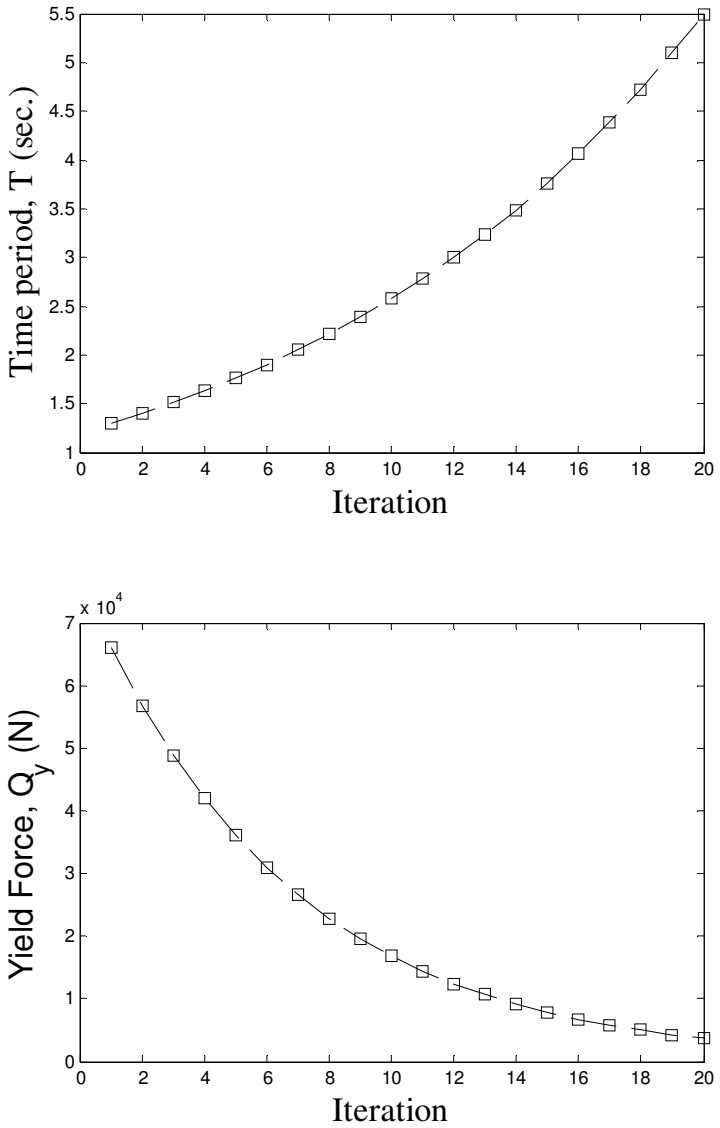

Figure 3.2. Post-yield time period and yield force for 20 iterations

3.2 Correlation Analysis and Regression Models

In this section, the proposed ground motion databases are presented first. This is followed by a correlation analysis between the *IMs.* Subsequently, the probabilistic seismic demand regression model is presented. The process of elimination of several redundant *IMs* is studied next. Finally, a multi-degree of freedom (MDOF) structure is proposed to validate the developed

UBI and BBI regression models. A popular model based on the peak ground velocity (*PGV*) is also studied and compared with the proposed regression models.

3.2.1 Ground Motion Database and Implementation of Earthquakes

A total of 774 ground motion records (387 pairs of two horizontal components) from 32 historical earthquakes (Table A.1, Appendix) are used for this study. These components are designated as north-south and east-west components of a particular site and event. For example, Parkfield earthquake has two sites and each contains a north-south and east-west component. The ground motions consist of world-wide earthquakes mostly taken from North America and Japan. These ground motions cover a wide range of hazard parameters such as intensity, tectonic environment, and site geology. Earthquakes 1 to 17 in Table A.1 (112 pairs of two horizontal components) were obtained from the PEER Strong Motion Database (http://peer.berkeley.edu/smcat/index), and earthquakes 18 to 32 were obtained from K-Net and KiK-NET in Japan. The source parameters of the Japan earthquakes are adopted from the data published by the Japan Meteorological Agency (JMA).

Out of these 774 ground motions, two equal data sets of 387 were generated for the UBI model. These data sets are denoted DS1 and DS2, which are used for analysis and validation of the UBI model respectively. More specifically, DS1 is used to construct the model while DS2 is used to validate and predict the results for the UBI system. DS1 is a compilation of all the north-south components of the 387 sites while DS2 contains the east-west components. For the BBI model, since both horizontal (north-south and east-west) components must be used in each application, two equal data sets of 193 sites are generated. The last Site (387) is discarded to make the number even for equal division. So in total there are two sets of 193 (total 386) and

these data sets are denoted DS3 and DS4 which are used for the analysis and validation of the BBI model respectively. Both UBI and BBI systems are modelled using both PCR and SI-ICR methods and the results compared.

The selected ground motions satisfy the following conditions: 1) The moment magnitude **M** is greater than 5.5 (Richter scale); 2) *PGA* values of two horizontal components are both larger than 100 gal (cm/sec^2); 3) the selected ground motions from one earthquake are no more than 25; and, 4) all the ground motions are observed in free field conditions. All these requirements allow for a well-rounded and unbiased prediction model.

3.2.2 Correlation Analysis

The correlation coefficients presented in Table A.2 (calculated using Eq. 2.3.1) are between the demand parameter d_b and the *IMs* calculated from DS1. In this case d_b is the maximum structural response of the UBI system in accordance with a 3.0 sec. post-yield time period. These correlation coefficients are presented on a logarithmic scale (natural log) and are arranged in the order of the acceleration sensitive, velocity sensitive, and displacement sensitive as shown by Chopra (1995) and Riddel (2007). These results are consistent with those presented by Riddel (2007). Riddel (2007) pointed out that "acceleration-related indices" in general present high correlation among them, while their correlation with "velocity-related indices" is poor, and even poorer correlation is observed with respect to "displacement-related indices". "Velocity-related indices" in general present high correlation among them, while some of them also present good correlation with "displacement-related indices." Finally, "displacement-related indices" are strongly correlated among them. Similar observations can be made with regards to DS2 and the demand parameter d_b presented in Table A.3. The correlation coefficients presented in Tables

A.4 and A.5 correspond to the BBI model response d_b and data sets DS3 and DS4 respectively. Details regarding DS1, DS2, DS3, and DS4 were presented in the previous section.

The correlation between duration, t_d, and other *IMs* as well as response d_b is very low. This is not surprising given that the length of the ground motion is independent of its intensity and the peak response it may produce in the structure. d_b has a high degree of correlation between velocity sensitive and displacement sensitive measures, while it shows low correlation with acceleration sensitive measures, and almost no correlation with the duration t_d. The correlation between d_b and *PGA, PGV,* t_d, and *PGD* for DS1 and post-yield period of 3.0 sec. are presented in Figs. 3.3, 3.4, A.3, and A.4 respectively. The correlation between d_b and t_d, *PGA, PGV,* and *PGD* for other data sets is similar.

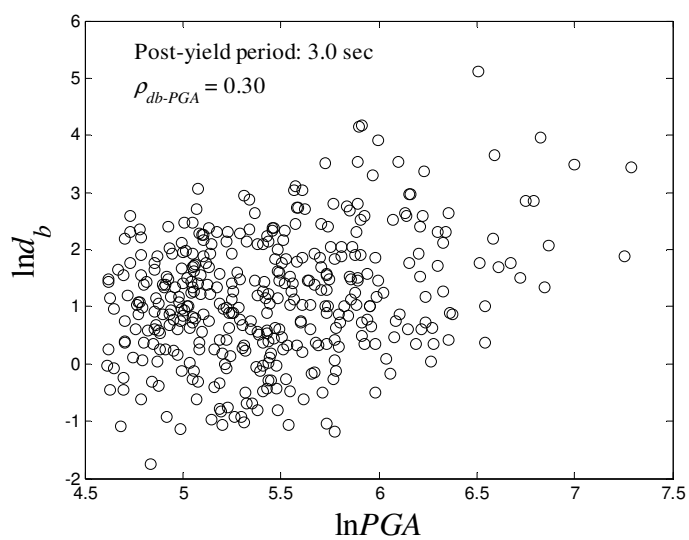

Figure 3.3. Correlation between d_b and *PGA* (T = 3.0 sec., DS1)

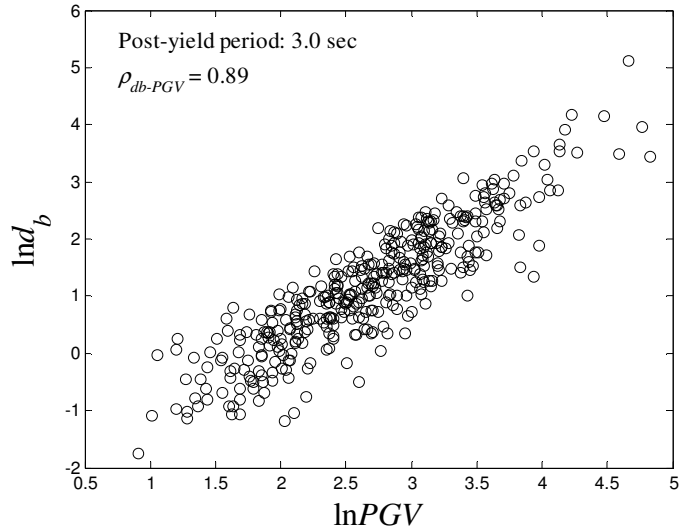

Figure 3.4. Correlation between d_b and *PGV* (T = 3.0 sec., DS1)

From the plots presented in Figs. 3.4 and A.4, it would be reasonable to assume that *PGV* and *PGD* would serve to be good predictors of d_b due to their high correlation with the response variable. But which ones would be most suitable depends on their correlations and dependency with one another as well as their practical limitations on estimating them with confidence. This is further explained in the next section. It should be pointed out that this thesis does not set out to define the most suitable *IMs* as predictors but rather to investigate the suitability of certain selected *IMs* based on their dependencies to other *IMs*. This is presented in the next section.

3.2.3 Probabilistic Seismic Demand Model and Analysis

As mentioned previously, in this study, practical implementation of several *IMs* is investigated as opposed to providing the best *IM(s) a priori*. Utilizing all the *IMs* in the model is impractical. Hence, it is important to sift through the entire set, and discard those *IMs* that contain redundant information. There are a number of ways in which *IMs* are discarded. First, the

IMs, which can be obtained through a linear combination of the other *IMs* in a log-scale, are excluded. This includes *IMs* such as I_c, I_F, I_a, I_v, and I_d. Second, *IMs* that only differ only by a constant from other *IMs* are also discarded; for example E_a as it is a multiple of I_A, root-mean-square values a_{rms}, v_{rms}, and d_{rms}, as well as root-square values a_{rs}, v_{rs}, and d_{rs}. Since the earthquake power values P_a, P_v, and P_d are combinations of the square of the ground motion and duration, they are also excluded from the regression model. As is apparent from the correlation tables, t_d almost has no correlation with the response variable d_b and thus not used in the regression model. Because it may be difficult to estimate *PGD* due to its sensitivity to the filter parameters used to process the acceleration record, it is eliminated from the regression model. Finally, E_d is also eliminated since it has no physical meaning in terms of energy dissipating characteristics. Thus, out of the 23 *IMs* originally developed, only five are retained for the regression model. They are *PGA,* I_A, *PGV,* E_v, and P_D.

In this study, the multiple regression model is set up as a linear combination of the *IMs* presented above in a logarithmic scale as estimators for the base displacement d_b for both the UBI and BBI models. The regression model has the following form and it is of the same form as Eq. 2.3.13.

$$\ln d_b(T) = \alpha_0 + \sum_n \alpha_i(T) \ln IM_i + \varepsilon(T) \tag{3.2.1}$$

IM_i is the ith intensity measure selected out of the five retained. α_0 and α_i are regression coefficients, ε is a normal variable with mean zero and standard deviation σ_ε. Every parameter except the *IMs* is a function of the post-yield period, *T*. Applying the five retained *IMs*, Eq. 3.2.1 can be expanded to the following. Note that *T* is omitted in the following expression for convenience.

$$\ln d_b = \alpha_0 + \alpha_1 \ln PGA + \alpha_2 \ln I_A + \alpha_3 \ln PGV + \alpha_4 \ln E_v + \alpha_5 \ln P_D + \varepsilon \quad (3.2.2)$$

The general statistical properties of the ground motions in databases DS1, DS2, DS3, and DS4 are presented in Table 3.2 for the five intensity measures selected.

Table 3.2. Statistics for the Ground Motions in DS1, DS2, DS3, and DS4

		PGA	I_A	PGV	E_v	P_D
Maximum	DS1	1467	3540	125	28105	91
	DS2	1211	2596	124	38748	231
	DS3	1467	1176	125	14688	66
	DS4	1415	3540	124	38748	231
Minimum	DS1	101	7	2	5	0.02
	DS2	100	6	3	6	0.01
	DS3	107	8	3	6	0.02
	DS4	101	11	3	7	0.05
Mean	DS1	271	150	19	1010	3
	DS2	271	152	20	1126	3
	DS3	296	143	23	1029	3
	DS4	313	214	22	1609	5
Standard Deviation	DS1	180	273	16	2591	8
	DS2	174	264	18	2987	15
	DS3	184	180	19	1988	5
	DS4	208	414	20	4361	21

Units are in centimetres and seconds

3.2.4 Model Predictions for Multi-Degree of Freedom (MDOF) Structure

Based on the models developed using PCR and SI-ICR for the BBI system, predicted responses are generated and compared with simulated responses of a MDOF base isolated structure. The superstructure is five floors plus the base with *x* and *y* directions at each mass making it a 12DOF structure. The plan and 3D representation of the 12DOF structure is shown in Fig. 3.5. The LRB yield force, $Q_Y = 11{,}484\ N$, and the post to pre-yield stiffness ratio, $K_r = 1/6$, is kept the same as the BBI model. This corresponds to a post-yield time period of approximately $T = 3.1$ sec. Other structural parameters for the 12DOF structures are presented in Table 3.3.

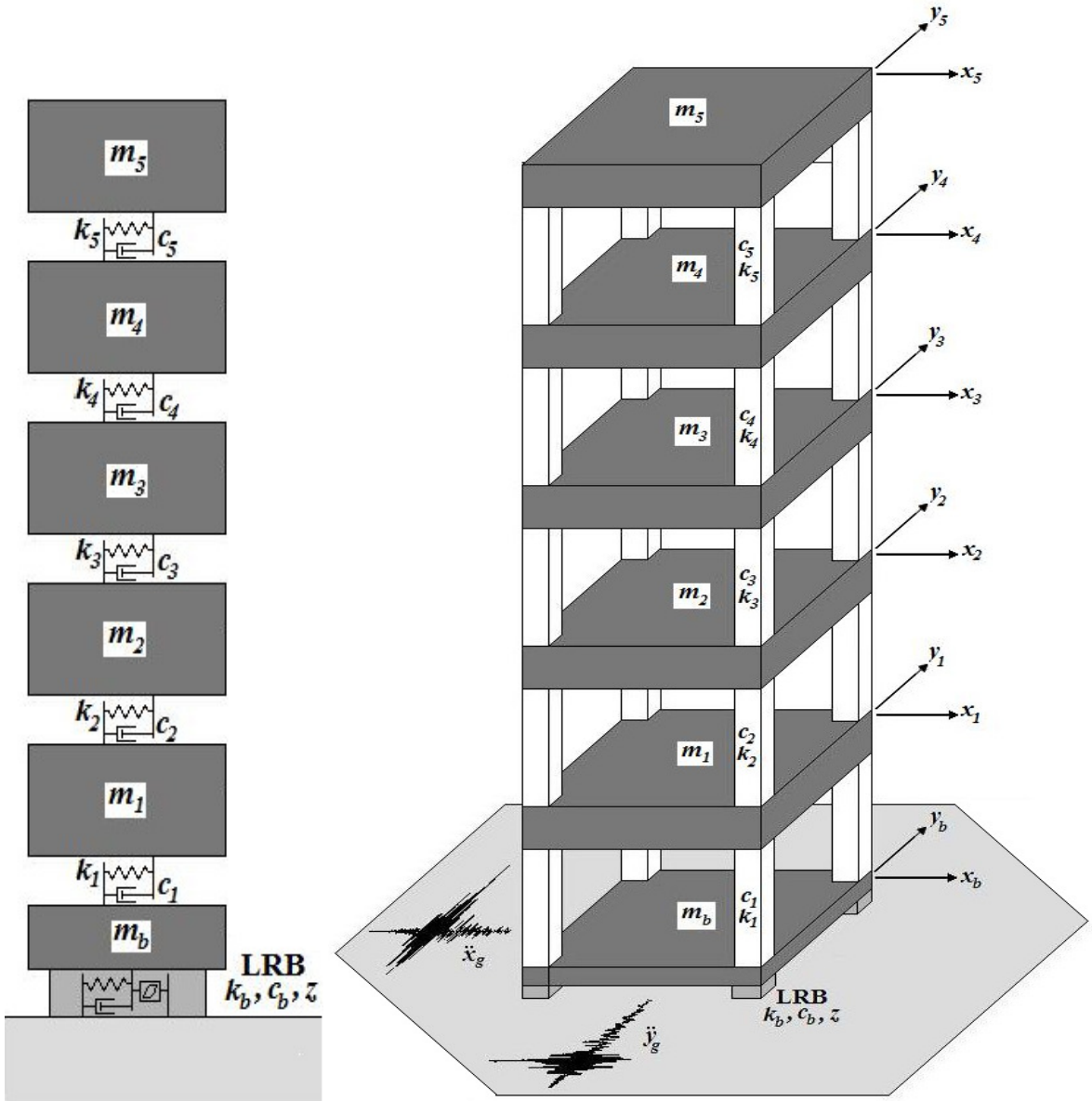

Figure 3.5. 12DOF structure plan and 3D representation

Table 3.3. Structural Parameters for 12DOF model

Parameter	Value	Units
m_1, m_2, m_3, m_4, m_5	5,897	kg
m_b	6,800	kg
k_1	33,732,000	N/m
k_2	29,093,000	N/m
k_3	28,621,000	N/m
k_4	24,954,000	N/m
k_5	19,059,000	N/m
k_b	232,000	N/m
c_1	67,000	$N \cdot s/m$
c_2	58,000	$N \cdot s/m$
c_3	57,000	$N \cdot s/m$
c_4	50,000	$N \cdot s/m$
c_5	38,000	$N \cdot s/m$
c_b	3,740	$N \cdot s/m$
Q_Y	11,484	N
D^Y	0.013	m
η	2	-
α	1	-
β	0.5	-
γ	0.5	-
K_r	1/6	-

In addition to the PCR and SI-ICR methods, a *PGV* model is developed and used to predict the responses of the MDOF system. As can be seen from Tables A.2 to A.5, the intensity measure *PGV* is highly correlated to d_b with correlation coefficients as high as 0.92 for DS2. As stated earlier, this suggests that *PGV* should be a good predictor of d_b. This is something that has been observed by a number of earlier studies regarding base isolation (Ryan and Chopra 2004a, 2005b). Hence, a regression model is developed between d_b and *PGV* using the least squares approach. This is a good tool to measure the effectiveness of using only one highly correlated intensity measure as opposed to multiple *IMs* in the PCR and SI-ICR methods. The least squares regression model between d_b and *PGV* is straight forward and details pertaining to it is not presented.

3.3 Regression Methodology for the Proposed Uni-axial and Bi-axial Systems

The application of PCR and SI-ICR is studied in relation to both UBI and BBI systems in detail in this section. This is followed by the model prediction for both UBI and BBI systems using either PCR or SI-ICR.

3.3.1 Principal Component Regression for Uni-axial and Bi-axial systems

Although least squares regression is the best linear unbiased estimator, the problem of collinearity still exists amongst the *IMs* due to the correlation that exists among them. It is difficult to separate the effects of the correlated *IMs* and estimate the regression coefficients with good accuracy using the traditional least squares approach. For example, looking at Table A.2, which corresponds to DS1, correlation coefficients between *PGA* and I_A, *PGV* and E_v, *PGV* and P_D, are 0.88, 0.92, and 0.84 respectively. Similar observations can be made from the other data sets as well. These *IMs* are highly correlated to one another and thus contain redundant information which can be eliminated by applying PCA followed by dimensionality reduction.

PCR handles the collinearity problem by eliminating those dimensions of the independent variable space, ln$\boldsymbol{IM}$, that are causing the collinearity problem. As explained in Section 2.3, PCR is a two step multivariate calibration method: In the first step, PCA is applied to the data matrix ln$\boldsymbol{IM}$**.** The newly acquired variables are denoted as the scores, $\mathbf{Z}$, which are now uncorrelated to one another. After the transformation to the scores, a multiple linear regression is conducted between the scores and the demand parameter $\ln d_b$. The d_b values for the UBI system are simply the maximum displacement recorded pertaining to each applied ground motion and the intensity measures are the ones recorded from those same ground motions. For the BBI system, a slightly different approach is followed. The d_b values for the BBI system are the maximum of

either *x* or *y* directions and the intensity measures are the maximum of either the *x* or *y* applied ground motion. *x* and *y* directions correspond to north-south and east-west ground motion components respectively. For example, for the Parkfield Earthquake – Site 1, the maximum displacement for post-yield time period 3.0 sec. is 4.77 *cm* which happens to be in the *y* direction and the maximum *PGA* value is 432.77 cm/s^2 which is also in the *y* direction. This makes sense since the higher *PGA* value, which in this case is in the *y* direction, also produced the higher displacement value in the same direction. This however is not always the case but it is important to take the maximum *IM* value because they would be the best predictors for the maximum displacement values.

Following the steps outlined in Section 2.3, the original data matrix **X** must be mean centered. Meaning $\mathbf{X} = \ln\boldsymbol{IM} - \boldsymbol{\mu}_{\ln\boldsymbol{IM}}$ where each column represents one of the five intensity measures and $\boldsymbol{\mu}_{\ln\boldsymbol{IM}}$ is a vector of mean $\ln\boldsymbol{IM}$. Therefore for the UBI system,

$$\mathbf{X} = \{[\ln\boldsymbol{PGA} - \boldsymbol{\mu}_{\ln\boldsymbol{PGA}}]\left[\ln\boldsymbol{I_A} - \boldsymbol{\mu}_{\ln\boldsymbol{I_A}}\right][\ln\boldsymbol{PGV} - \boldsymbol{\mu}_{\ln\boldsymbol{PGV}}]\left[\ln\boldsymbol{E_v} - \boldsymbol{\mu}_{\ln\boldsymbol{E_v}}\right]\left[\ln\boldsymbol{P_D} - \boldsymbol{\mu}_{\ln\boldsymbol{P_D}}\right]\}_{387\times5}$$

and for the BBI system,

$$\mathbf{X} = \{[\ln\boldsymbol{PGA} - \boldsymbol{\mu}_{\ln\boldsymbol{PGA}}]\left[\ln\boldsymbol{I_A} - \boldsymbol{\mu}_{\ln\boldsymbol{I_A}}\right][\ln\boldsymbol{PGV} - \boldsymbol{\mu}_{\ln\boldsymbol{PGV}}]\left[\ln\boldsymbol{E_v} - \boldsymbol{\mu}_{\ln\boldsymbol{E_v}}\right]\left[\ln\boldsymbol{P_D} - \boldsymbol{\mu}_{\ln\boldsymbol{P_D}}\right]\}_{193\times5}$$

Z, as stated before, is the scores of matrix **X** and is related to a linear transformation **P** through the following transformation.

$$\mathbf{Z} = \mathbf{XP} \tag{3.3.1}$$

Where the columns of **P** are the eigenvectors or loadings of the covariance matrix of **X**. The **Λ** matrix, not shown here, contains the eigenvalues of the covariance matrix of **X** and the square

root of those values are the λ_i values. This means that $\lambda_1, \lambda_2, \ldots \lambda_5$ correspond to the scores on principal components $\mathrm{PC}_1, \mathrm{PC}_2, \ldots \mathrm{PC}_5$ and λ_i is sorted in decreasing order of importance. The higher the eigenvalues for the corresponding original variables, the more important those new scores are in regards to how well they explain the variance of the original data.

Once the transformation is applied on Eq. 3.2.2, the new scores **Z** are uncorrelated to one another. Dimensionality reduction can now be applied to the principal components (PCs). For example, if the first four PCs explain 99% of the original variance, retaining the fifth PC wouldn't add much significance in predicting the demand parameter. More generally, *m* PCs are retained from an original dimension of *n*, where $m < n$. Dimensionality reduction does not have to be applied, but that would be the same as applying multiple linear regression to the original *IMs* and the response without the transformation in Eq. 3.3.1. But in this study, dimensionality reduction is investigated and this means the transformation in Eq. 3.3.1 must be applied.

The next step in PCR is to perform linear regression of the scores **Z** and the demand parameter, $\ln d_b$. The linear model between $\ln d_b$ and **Z** for each post-yield time period, *T*, is shown below. Note the response variable is mean centered as were the original variables.

$$\ln d_b - \mu_{\ln d_b} = \mathbf{Z\beta} + \varepsilon \tag{3.3.2}$$

where,

$$\boldsymbol{\beta} = (\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}^{\mathbf{T}}(\ln \boldsymbol{d_b} - \mu_{\ln \boldsymbol{d_b}}) \tag{3.3.3}$$

β's are the regression coefficients in the principal component space but have no physical meaning in regards to the original *IM*-space in Eq. 3.2.2. They have to be converted to

coefficients that relate the original variables to the demand parameter. They are denoted α_0 and $\boldsymbol{\alpha}^{\mathbf{T}} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]$ and can be expressed through the eigenvectors $\mathbf{P}$,

$$\alpha_0 = \mu_{\ln d_b} - \boldsymbol{\mu}_{\ln \boldsymbol{IM}} \mathbf{P}\boldsymbol{\beta} \tag{3.3.4}$$

$$\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta} \tag{3.3.5}$$

3.3.2 Sorted-Input Independent Component Regression for Uni-axial and Bi-bxial systems

As mentioned in Section 2.4.4, whitening is recommended for the original variables as a pre-processing strategy in ICA. Whitening, similar to PCA, transforms the correlated components to uncorrelated ones. SI-ICR as previously outlined is essentially an application of PCA on the original variables followed by an application of ICA. This is followed by a sorting of the variables, in decreasing order of correlation to the response variable. Careful attention must be made when sorting the variables because the loading matrix must be sorted in a similar fashion for the method to work properly.

A brief summary of the SI-ICR algorithm is presented as it is the same as the one presented in Section 2.4.4. The first step is the mean centering of the original data matrix $\mathbf{X}$ as shown in Section 3.3.1. Based on the *N* number of samples (387 for UBI model and 193 for BBI model), the covariance matrix using Eq. 2.4.11 is estimated. Eigenvectors $\boldsymbol{u}_k$, and eigenvalues λ_k are computed using PCA and dimensionality reduction is done according to Eq. 2.4.12. After obtaining $\mathbf{X}_{red}$, ICA is performed on the newly reduced vectors using Eq. 2.4.13 to compute the reduced mixing matrix, $\mathbf{A}_{red}$. Equations 2.4.14 and 2.4.15 are then used to compute the scores, $\hat{\mathbf{s}}$. The newly acquired independent variables and their respective loadings are sorted in decreasing order of correlation to the response variable. Once the new scores and loadings are

obtained (Eq. 2.4.16), multiple linear regression is then applied in a similar fashion to PCR. Equations 3.3.2 to 3.3.5 are applied in the same fashion for the SI-ICR method.

3.3.3 Model Prediction for Uni-axial and Bi-axial systems

After obtaining the β and α coefficients from Eqs. 3.3.3, 3.3.4, and 3.3.5, the prediction model, obtained using DS1 for the UBI system and DS3 for the BBI system, can be used to determine new response values for a separate set of *IMs*, namely, DS2 for UBI and DS4 for BBI. For both UBI and BBI systems, the prediction model is attained the same way. The new score, obtained from either DS2 or DS4, is denoted $\mathbf{z}_*$ and is used to predict the values of the response d_b for their respective systems (UBI and BBI). d_b follows a t-distribution with v degrees of freedom (Weisberg 2005).

$$\frac{\ln d_b - \widehat{\ln d_b}}{\sigma_\varepsilon \sqrt{1+\mathbf{z}_*^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})\mathbf{z}_*}} \sim t_v \tag{3.3.6}$$

Where, $\widehat{\ln d_b} = \mu_{\ln d_b} + \mathbf{Z\beta}$ is a point estimate of $\ln d_b$, σ_ε is the standard error of the regression model, $v = n - m$ is the degrees of freedom, n is the size of the sample used in the regression model, and m is the number of scores retained in the regression model. For example, in the UBI model v = 387 – 4 = 383, and in the BBI model v = 193 – 4 = 189, if four PCs or four independent components (ICs) are retained. A 1-δ percent predictive interval can be calculated from a t-distribution as shown below.

$$\widehat{\ln d_b} - \sigma_\varepsilon \sqrt{1 + \mathbf{z}_*^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})\mathbf{z}_*} \cdot t_v(\delta/2) \leq \ln d_b \leq \widehat{\ln d_b} + \sigma_\varepsilon \sqrt{1 + \mathbf{z}_*^{\mathrm{T}}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})\mathbf{z}_*} \cdot t_v(\delta/2) \tag{3.3.7}$$

In this study, 90% prediction intervals are used to judge the efficiency of the model predictions.

4. RESULTS AND DISCUSSION

In this section, results for both PCR and SI-ICR methods are presented for the UBI and BBI systems. These results include eigenvectors, eigenvalues, regression coefficients, explained variances, component scatter plots, probability plots, standard errors, goodness of fit values, and prediction models for all cases. For the sake of brevity, a post-yield time period of 3.0 sec., considered representative of base isolated systems, is presented here as a general case. The 12DOF model presented in Section 3.2.4 is also studied in the context of PCR, SI-ICR and *PGV* models.

4.1 Principal Component Regression Results

PCR is performed on both UBI and BBI models using the procedure outlined in Section 3.3.1. As stated before, DS1 was used for the UBI system for model development. The eigenvector matrix, $\mathbf{P}_{\mathrm{DS1}}$, and the corresponding eigenvalues, $\mathbf{\Lambda}_{\mathrm{DS1}}$, of the covariance matrix of $\mathbf{X}_{\mathrm{DS1}} = \ln \boldsymbol{IM}_{\mathrm{DS1}} - \boldsymbol{\mu}_{\ln \boldsymbol{IM}_{\mathrm{DS1}}}$ is shown below.

$$\mathbf{P}_{\mathrm{DS1}} = \begin{bmatrix} -0.14 & 0.39 & -0.20 & 0.53 & 0.71 \\ -0.35 & 0.79 & -0.21 & -0.34 & -0.31 \\ -0.28 & -0.12 & -0.25 & 0.71 & -0.59 \\ -0.65 & -0.46 & -0.47 & -0.31 & 0.22 \\ -0.60 & 0.005 & 0.80 & 0.08 & 0.05 \end{bmatrix}, \quad \mathbf{\Lambda}_{\mathrm{DS1}} = \begin{bmatrix} 5.55 \\ 0.80 \\ 0.26 \\ 0.09 \\ 0.02 \end{bmatrix}$$

Similarly, DS3 was used for the BBI system for model development. The eigenvector matrix, $\mathbf{P}_{\mathrm{DS3}}$, and the corresponding eigenvalues, $\mathbf{\Lambda}_{\mathrm{DS3}}$, of the covariance matrix of $\mathbf{X}_{\mathrm{DS3}} = \ln \boldsymbol{IM}_{\mathrm{DS3}} - \boldsymbol{\mu}_{\ln \boldsymbol{IM}_{\mathrm{DS3}}}$ is shown below.

$$\mathbf{P}_{\mathrm{DS3}} = \begin{bmatrix} -0.11 & 0.41 & -0.24 & -0.56 & -0.67 \\ -0.28 & 0.83 & -0.12 & 0.35 & 0.30 \\ -0.30 & -0.11 & -0.31 & -0.64 & 0.63 \\ -0.68 & -0.37 & -0.46 & 0.36 & -0.25 \\ -0.59 & 0.002 & 0.79 & -0.16 & -0.05 \end{bmatrix}, \quad \mathbf{\Lambda}_{\mathrm{DS3}} = \begin{bmatrix} 4.84 \\ 0.97 \\ 0.34 \\ 0.08 \\ 0.01 \end{bmatrix}$$

Considering $\mathbf{\Lambda}_{\mathrm{DS1}}$, the higher eigenvalues corresponds to the more important component PCs. The same can be concluded from the percentage of variance explained. The percentage explanation of the PCs to the variance for DS1 and DS3 are shown in Figs. 4.1 and 4.2 respectively.

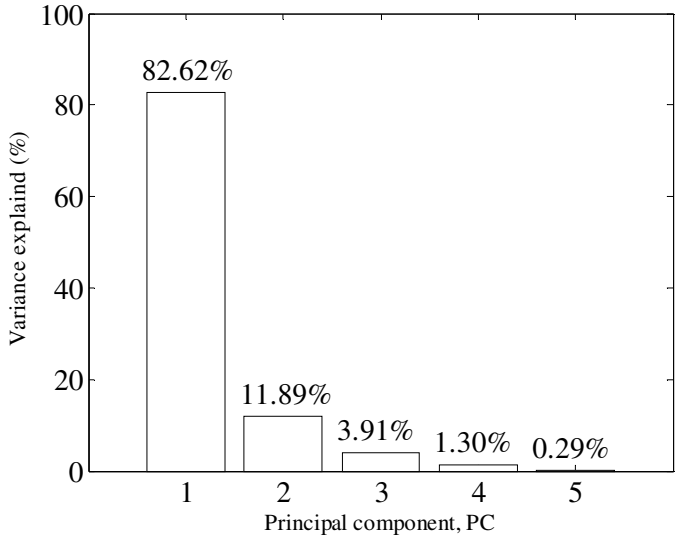

Figure 4.1. Percentage of variance explained for DS1

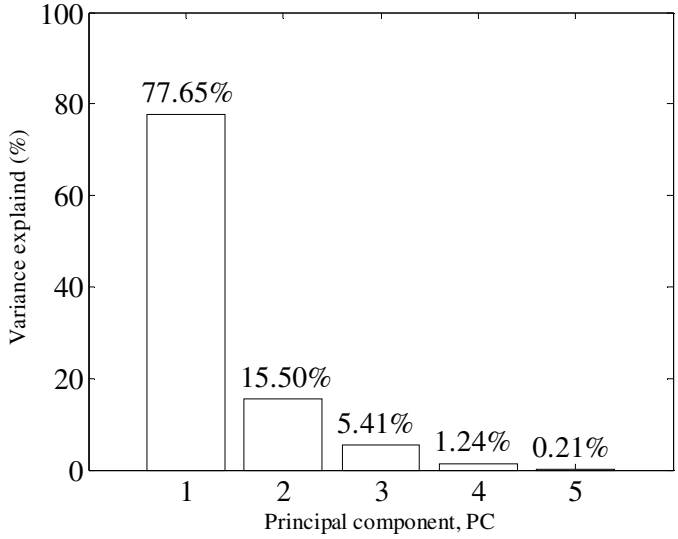

Figure 4.2. Percentage of variance explained for DS3

It can be seen from Figs. 4.1 and 4.2 that the first two PCs alone explain about 94.5% and 93.0% of the variance in the *IMs* calculated from DS1 and DS3 respectively. This shows that if the last one or two PCs are discarded, the effect on the overall model in terms of predicting the base displacement will be relatively small.

Recalling that the main reason why PCA is applied is to uncorrelate variables, Figs. 4.3 and 4.4 show the scatter plots between PC_1 and PC_2 for both DS1 and DS3 respectively.

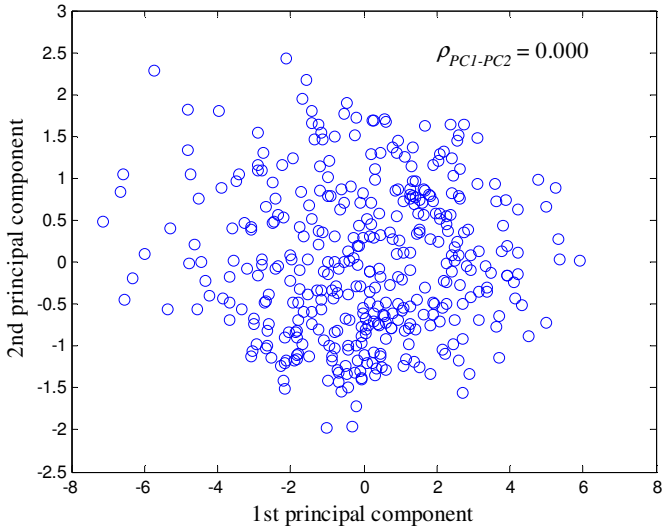

Figure 4.3. Scatter plot between PC_1 and PC_2 for DS1

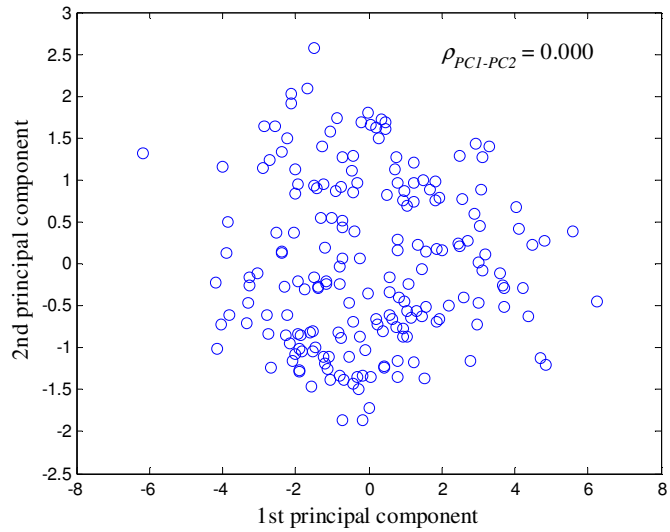

Figure 4.4. Scatter plot between PC_1 and $PC_2$for DS3

All the components after the transformation are now uncorrelated to one another. The same is true for the other components. The scatters of d_b with PC_1 and PC_5 for DS1 are shown in Figs. 4.5 and 4.6 respectively. The scatters of d_b with PC_2, PC_3, and PC_4 for DS1 are shown in Figs. A.5, A.6, and A.7 respectively.

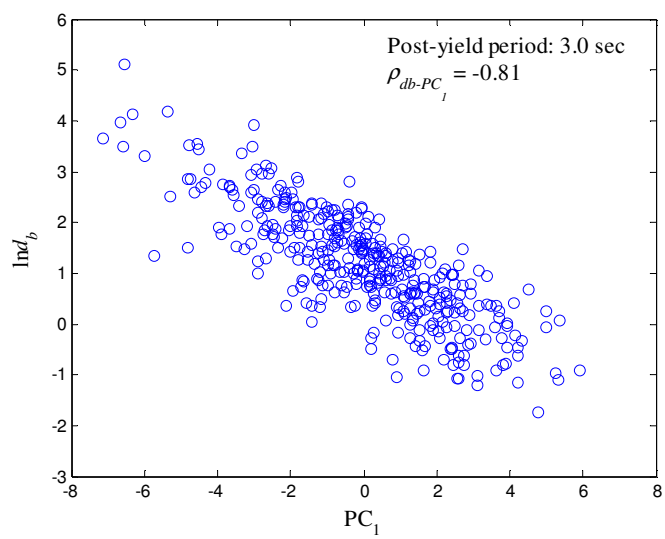

Figure 4.5. Scatter plot of d_b to PC_1 for DS1

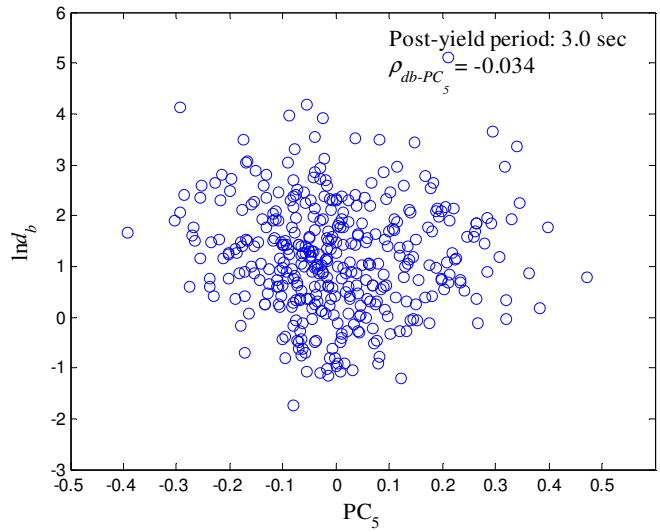

Figure 4.6. Scatter plot of d_b to PC_5 for DS1

It can be observed that the correlation between the PCs and $\ln d_b$ reduces from 1 to 5 components. Correlation coefficients between $\ln d_b$ and the five PCs for DS1 are -0.81, -0.43, 0.22, 0.13, and -0.034, respectively. This reaffirms the fact that the first couple of PCs are most important in predicting the response variable and that if the last one or two are discarded, not much accuracy would be lost. Similar correlation coefficients can be seen between $\ln d_b$ and PCs at other periods (1.3 to 5.5 seconds) as well. The same is true for the BBI system as can be seen from Figs. A.8 to A.12.

The advantage of using PCR can be gained from observing that the correlations with the physical components decreases as the eigenvalues decrease. This enables using only a subset of the PCs to develop the regression model. Initially, all five components are retained.

Tables A.6 and A.7 show the estimation of the β coefficients, according to Eq. 3.3.3, their standard deviations, t-values, and p-values for DS1 and DS3, respectively. Since the PCs are now uncorrelated to one another, the t-test of the coefficients can be applied. The t-value is the ratio

of the estimate to its standard error; the p-value is the conditional probability of observing a value of a t-statistic that is large or larger in magnitude given the null hypothesis that the true coefficient value is zero. The standard error of estimate σ_ε is a measure of precision in the prediction, which provides a basis for judging the reliability of the estimate. The two-sided p-value is calculated as Pr (> | t |) from the t distribution in Tables A.6 and A.7. It can be seen that the p-values of the coefficients of the first four PCs for both UBI and BBI cases are very close to zero, while the p-values for β_5 are relatively large compared to the rest. The t-test in this case shows that the regression coefficients for the first four PCs are significant, while the regression coefficient for the fifth component is not significant. Thus the fifth PC can be eliminated and as a result, no important information regarding $\ln d_b$ would be lost. This statement agrees with Figs. 4.1 and 4.2 which show the fifth PC explains only 0.29% and 0.21% of the variance in DS1 and DS3 respectively. Figs. 4.7 and A.13 show the regression coefficients β(T) using four significant PCs for DS1 and DS3 respectively.

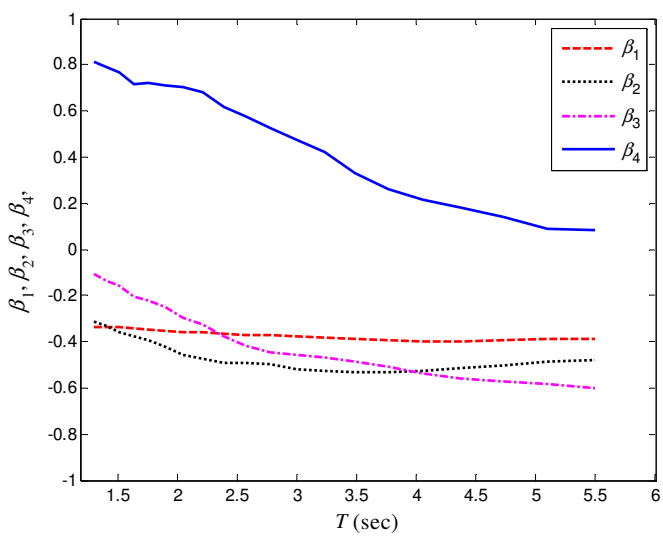

Figure 4.7. β using four PCs for DS1

$\boldsymbol{\alpha}$'s can be obtained using the transformations shown in Eqs. 3.3.4 and 3.3.5. Tables A.8 and A.9 show the regression coefficients, $\boldsymbol{\alpha}$, using four PCs for dataset DS1 and DS3 respectively. The standard errors for DS1 range from 0.32-0.37 and for DS3 range from 0.33-0.39 so it is obvious that the uniaxial model can be predicted with better precision than the biaxial model. The standard errors remain relatively constant for both UBI and BBI systems throughout the post-yield time period spectrum. This is an important result, for it shows that the combination of the *IMs* is capable of predicting the structural responses for both systems throughout a wide range of periods consistent with the design of base isolated structures.

The best way to check the effectiveness and accuracy of the model, as well as the standard error values presented earlier, is using the general R^2 statistics test. R^2 is a measure of goodness of fit of a prediction. There are several ways in which R^2 can be defined but for purposes of this study, the R^2 value used is shown below (MATLAB 2007b),

$$R^2 = 1 - \frac{\mathrm{norm}(Y-\hat{Y})^2}{\mathrm{norm}(Y-\mu_Y)^2} \tag{4.1.1}$$

Where, $\hat{Y}$ is the predicted response, Y is the actual response, and μ_Y is the response mean. The norm of a matrix is defined below.

$$\mathrm{norm}(A) = \|A\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2} \tag{4.1.2}$$

Figure 4.8 shows the R^2 values with respect to various *T* values for all PCs for DS1. It can be seen that using the first four PCs is sufficient, since the fifth PC does not improve the model's accuracy by much. The same conclusion can be made from Fig. 4.9 which shows the σ_ε values

with respect to the various time periods used for DS1. R^2 and σ_ε for DS3 are shown in Figs. A.14 and A.15 respectively.

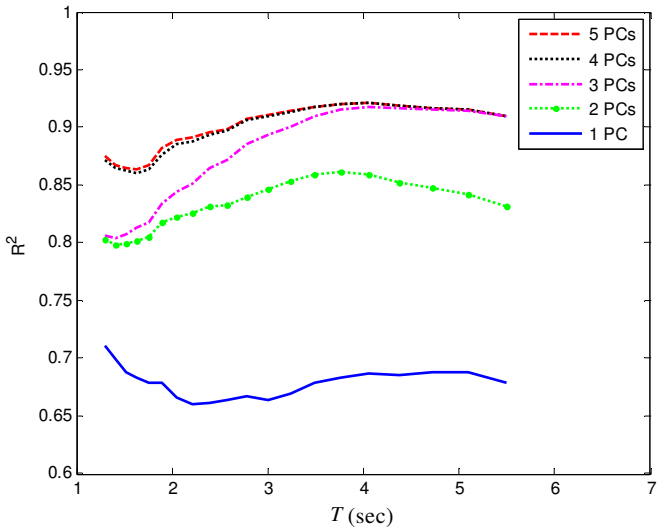

Figure 4.8. R^2 values using various PCs for DS1

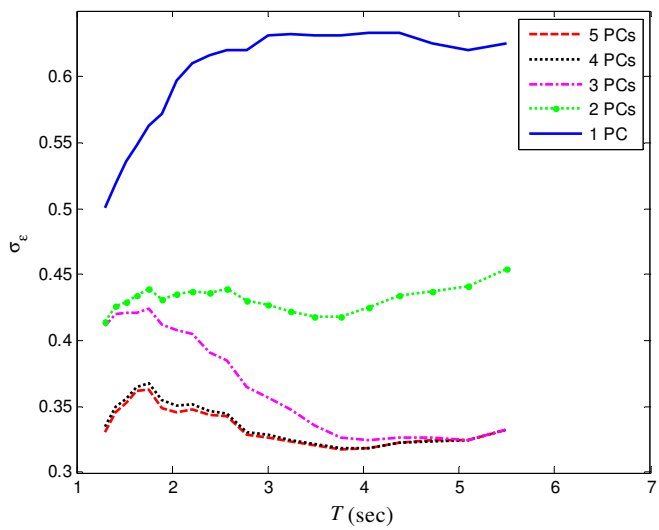

Figure 4.9. σ_ε values using various PCs for DS1

By analyzing the residuals, one can better assess the effectiveness of the regression model. Figs. 4.10 and A.16 show the normal probability paper plot of the residuals for DS1 and

DS3 respectively. From these plots, it can be inferred that the regression model shown in Eq. 3.2.2 is approximately normal.

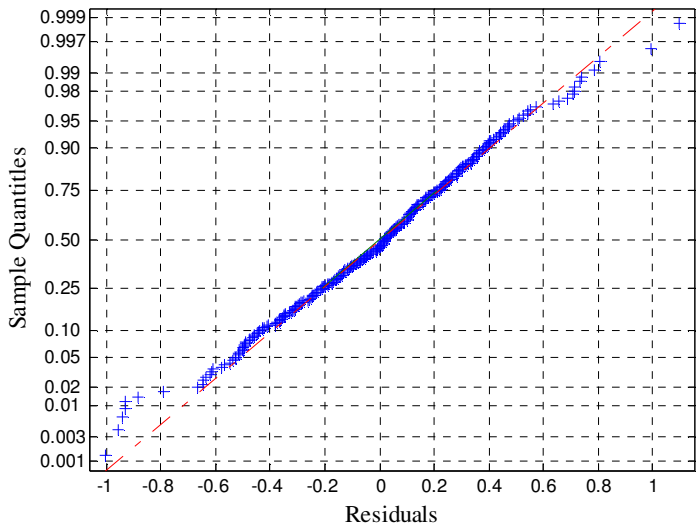

Figure 4.10. Normal probability plot of residuals for DS1 using PCR (T = 3.0 sec.)

Figs. 4.11 and 4.13 show the simulated values against the predicted values as well as the 90% prediction intervals for DS1 and DS3 respectively. It can be seen that most of the predicted vs. simulated values lie well within the 1:1 lines, meaning that the 90% prediction intervals shown in Eq. 3.3.7 includes most of the data. Figs. 4.10 and A.16 show that the normality assumption in Eq. 3.2.2 is approximately satisfied. The results in Figs. 4.12 and 4.14 show the predicted values obtained from Eq. 3.3.6 and the simulated values for DS2 and DS4 respectively. The mean value of the residuals are 0.008 and -0.010 for DS2 and DS4 respectively which means that both the proposed UBI and BBI models are unbiased (mean value of the residuals is below 5%). The standard error of the residuals are 0.30 and 0.36 for the DS2 and DS4 cases, which implies that the proposed UBI model can predict the base displacement responses with a better degree of confidence than the BBI model. Both models remain relatively unbiased for other

period ranges (T = 1.3 to 5.5 sec.). The mean value of the residuals and the standard error of the residuals are defined below.

$$\mu_{\ln\,(residual)} = \mu_{\ln y} - \mu_{\ln y_{predict}} \tag{4.1.3}$$

$$\sigma_{\ln\,(residual)} = \sigma_{\ln y} - \sigma_{\ln y_{predict}} \tag{4.1.4}$$

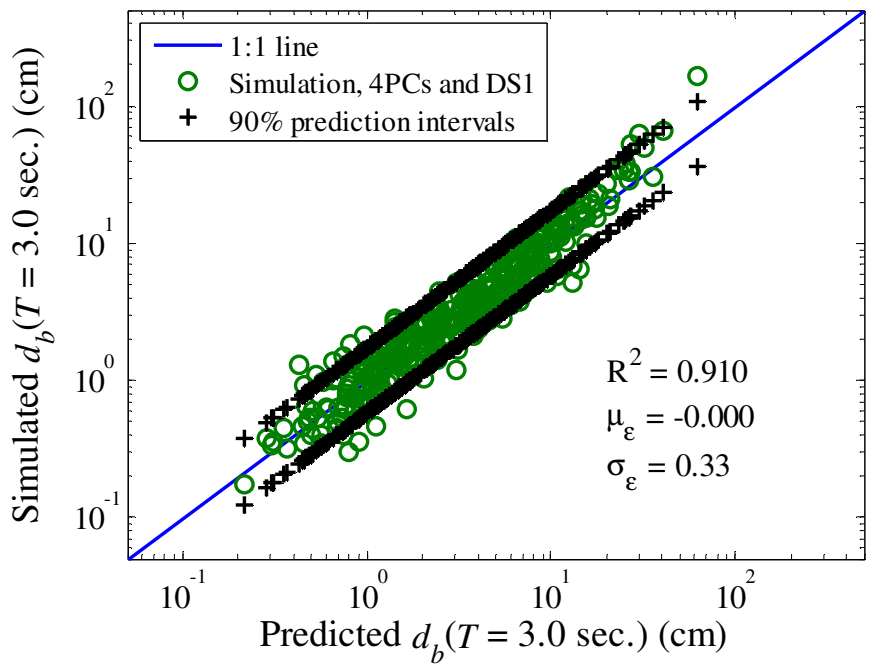

Figure 4.11. PCR: Prediction intervals for DS1 (2DOF UBI model)

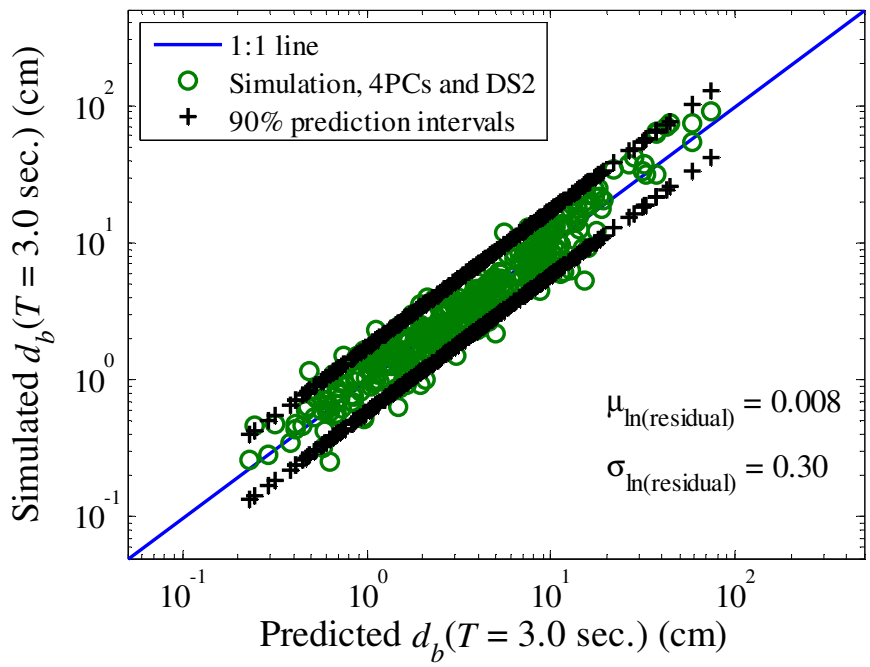

Figure 4.12. PCR: Prediction intervals for DS2 (2DOF UBI model)

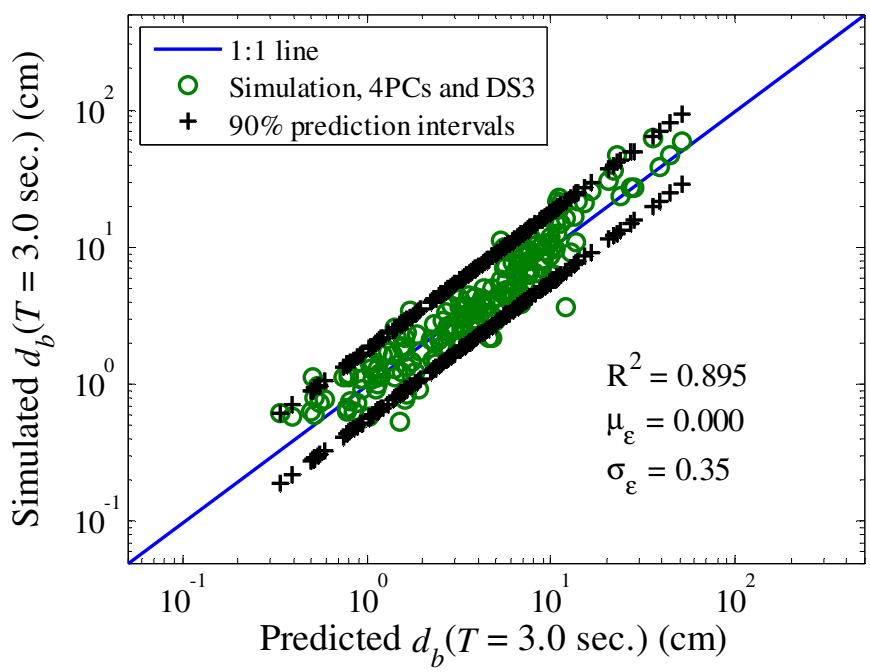

Figure 4.13. PCR: Prediction intervals for DS3 (4DOF BBI model)

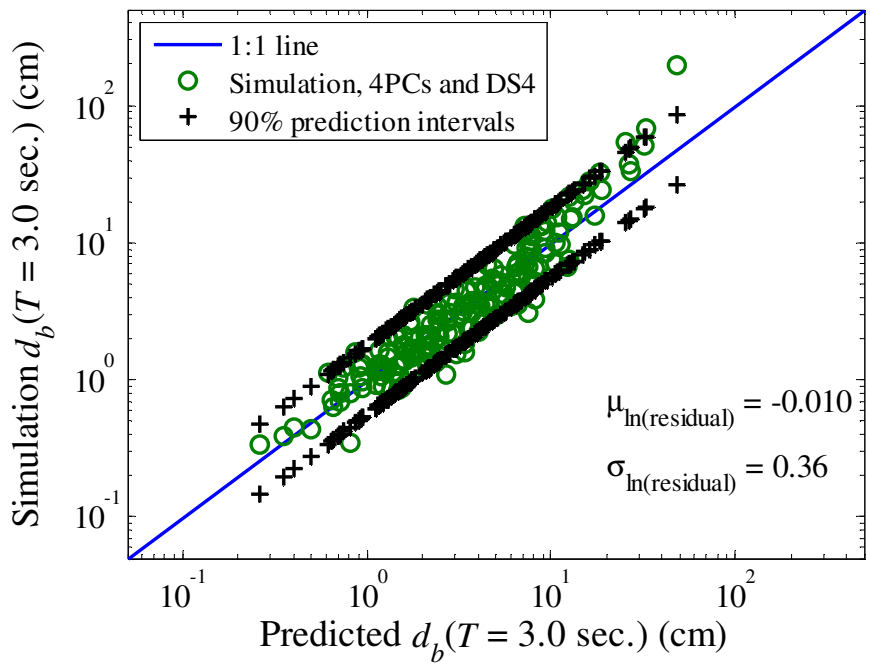

Figure 4.14. PCR: Prediction intervals for DS4 (4DOF BBI model)

4.2 Sorted-Input Independent Component Regression Results

SI-ICR is performed on both UBI and BBI models using the procedure outlined in Section 3.3.2. Similar to PCR, DS1 was used for the UBI system for model development. For comparison, four independent components (ICs) were retained and regressed with the response. It is expected that retaining all five ICs would yield the same results as retaining five PCs. The eigenvector matrix, $\mathbf{P}_{\mathrm{DS1}}$, is presented below.

$$\mathbf{P}_{\mathrm{DS1}} = \begin{bmatrix} -0.63 & 3.02 & 1.31 & -0.59 \\ -0.85 & -1.27 & -1.64 & -1.05 \\ -0.02 & 3.75 & 2.07 & -0.49 \\ 0.95 & -1.53 & -0.68 & -1.81 \\ 0.06 & 0.66 & -0.35 & 2.86 \end{bmatrix}$$

Similarly, DS3 was used for the BBI system for developing the model. The eigenvector matrix, $\mathbf{P}_{\mathrm{DS3}}$, is presented below.

$$\mathbf{P}_{\mathrm{DS3}} = \begin{bmatrix} 4.73 & 0.65 & -0.30 & 2.22 \\ -1.72 & 0.72 & -0.61 & -2.14 \\ -3.92 & 0.02 & -0.54 & 2.17 \\ -1.93 & -0.62 & -1.70 & -1.15 \\ 1.41 & -.026 & 2.45 & 0.14 \end{bmatrix}$$

It goes without mention that after SI-ICR, the ICs will be uncorrelated and independent to one another. Figs. 4.15 and 4.16 show the scatter plots between IC_1 and IC_2 for both DS1 and DS3 respectively.

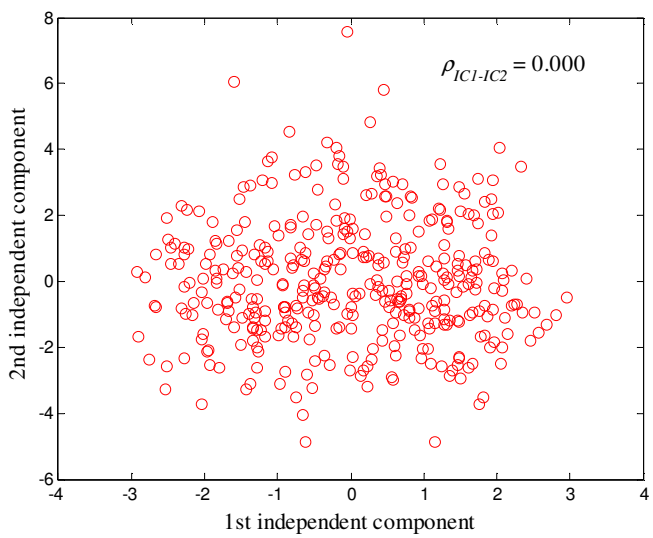

Figure 4.15. Scatter plot between IC_1 and IC_2 for DS1

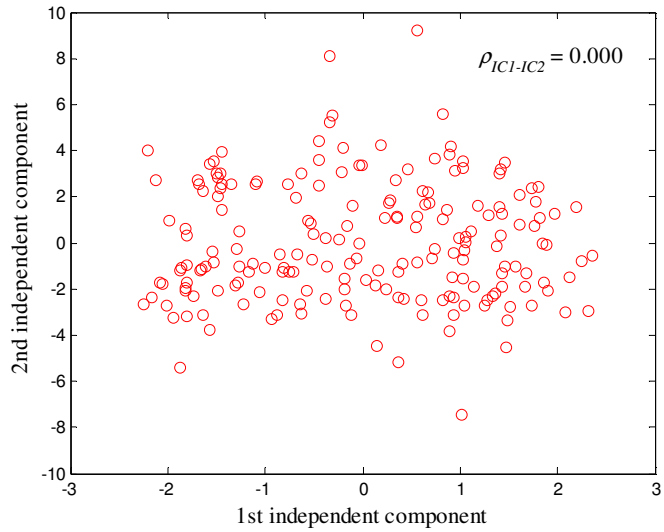

Figure 4.16. Scatter plot between IC_1 and IC_2 for DS3

Components 1 and 2 are now uncorrelated and independent to one another for both cases. The scatters of $\ln d_b$ with IC_1 and IC_4 for DS1 are shown in Figs. 4.17 and 4.8 respectively. The scatters of $\ln d_b$ with IC_2 and IC_3 for DS1 are shown in Figs. A.17 and A.18 respectively.

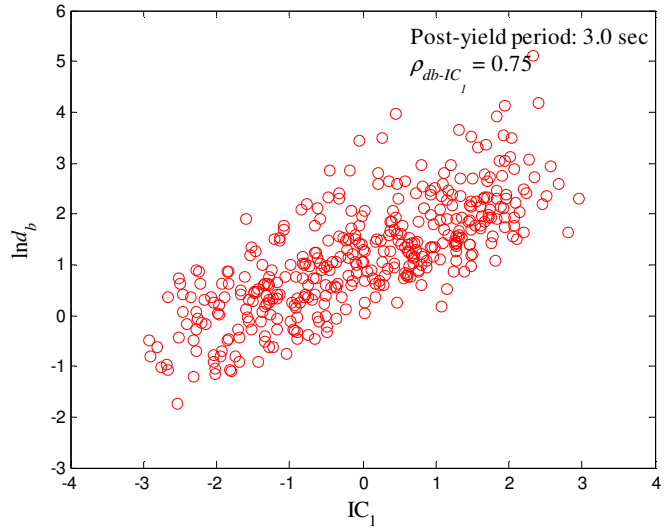

Figure 4.17. Scatter plot of d_b to IC_1 for DS1

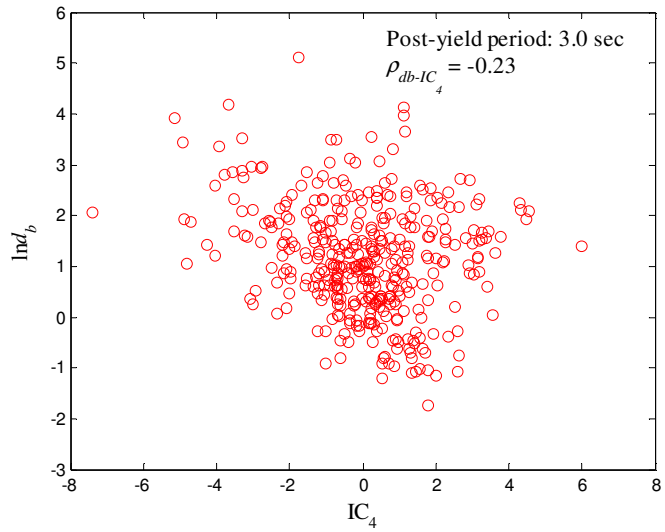

Figure 4.18. Scatter plot of d_b to IC_4 for DS1

Similar to PCA, correlation between the ICs and $\ln d_b$ gets lower from the first to the fourth component. Correlation coefficients between $\ln d_b$ and the four ICs are 0.75, 0.45, -0.30, and -0.23 respectively. Similarly, the first two ICs are the most important in predicting the response variable. Similar correlation coefficients can be seen between $\ln d_b$ and ICs at other periods (1.3 to 5.5 sec.). The same is true for the BBI system as can be seen from Figs. A.19 to A.22.

Tables A.10 and A.11 show the regression coefficients, $\boldsymbol{\alpha}$, using four ICs for datasets DS1 and DS3 respectively. The standard errors for DS1 range from 0.32-0.37 and for DS3 range from 0.33-0.39 which are similar to what PCR produces. Figure 4.19 shows the R^2 values with respect to various T values for all ICs. It can be seen that using the first four ICs is sufficient, since the fifth IC does not improve the model's accuracy, similar to what was observed with PCR. The same conclusion can be made from Fig. 4.20 which shows the σ_ε values with respect to the various time periods used. R^2 and σ_ε for DS3 are shown in Figs. A.23 and A.24 respectively.

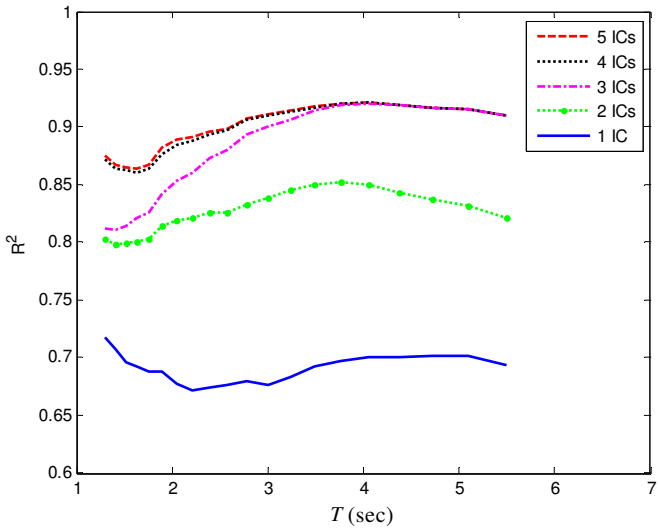

Figure 4.19. R^2 values using various ICs for DS1

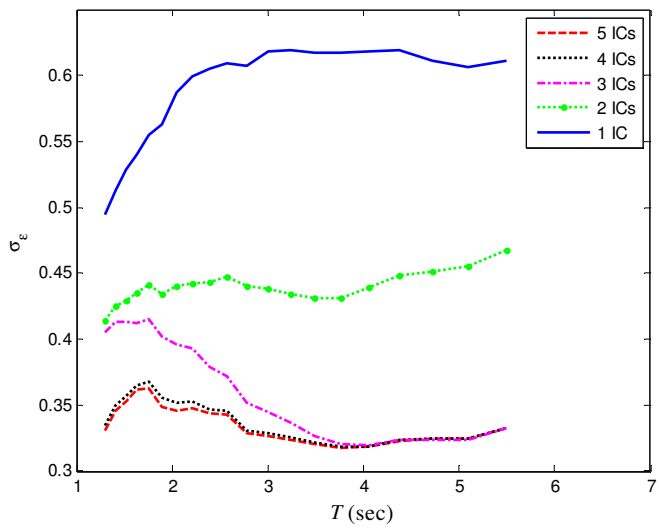

Figure 4.20. σ_ε values using various ICs for DS1

Figs. 4.21 and A.25 show the normal probability plot of the residuals for DS1 and DS3 respectively. From these plots, it can be inferred that the regression model shown in Eq. 3.2.2 is also approximately normal when using SI-ICR.

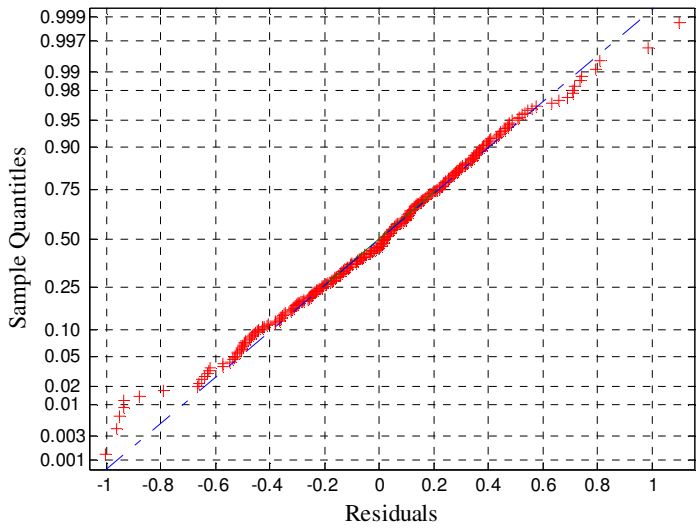

Figure 4.21. Normal probability plot of residuals for DS1 using SI-ICR (T = 3.0 sec.)

Figs. 4.22 and 4.24 show the simulated values against the predicted values as well as the 90% prediction intervals for DS1 and DS3 respectively. Similar to PCR results, most of the predicted vs. simulated values lie well within the 1:1 lines, meaning that the 90% prediction intervals shown in Eq. 3.3.7 includes most of the data. The results in Figs. 4.23 and 4.25 show the predicted values obtained from Eq. 3.3.6 and the simulated values for DS2 and DS4 respectively. The mean values of the residuals are 0.009 and -0.038 for DS2 and DS4 respectively which means that both the proposed UBI and BBI models are relatively unbiased. The standard error of the residuals are 0.30 and 0.37 for the DS2 and DS4 respectively, which implies that the proposed UBI model can predict the base displacement responses with a better degree of confidence than the BBI model, similar to PCR. Both models remain relatively unbiased for other period ranges (T = 1.3 to 5.5 sec.). Results pertaining to other time periods are not presented here due to length considerations.

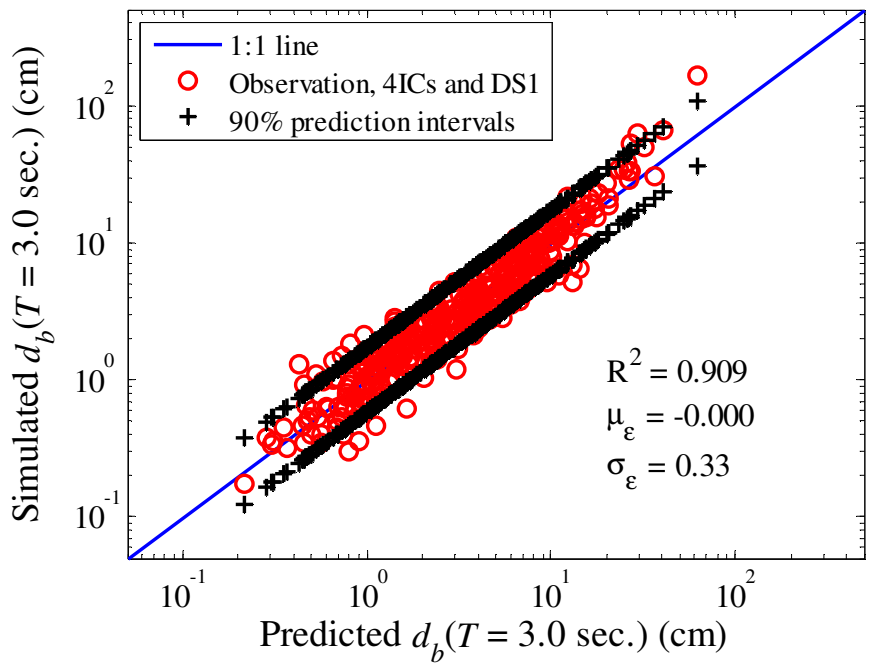

Figure 4.22. SI-ICR: Prediction intervals for DS1 (2DOF UBI model)

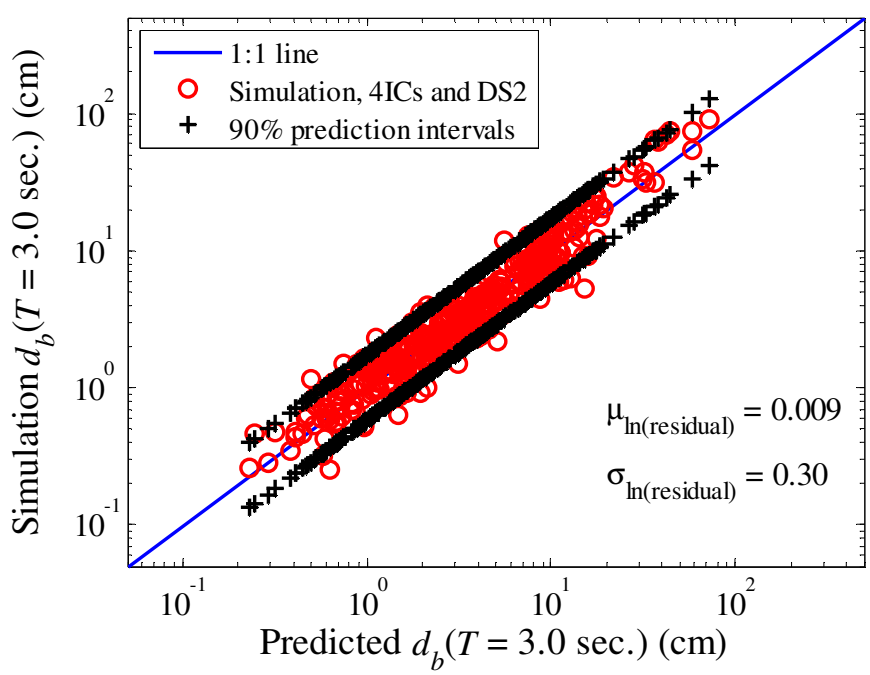

Figure 4.23. SI-ICR: Prediction intervals for DS2 (2DOF UBI model)

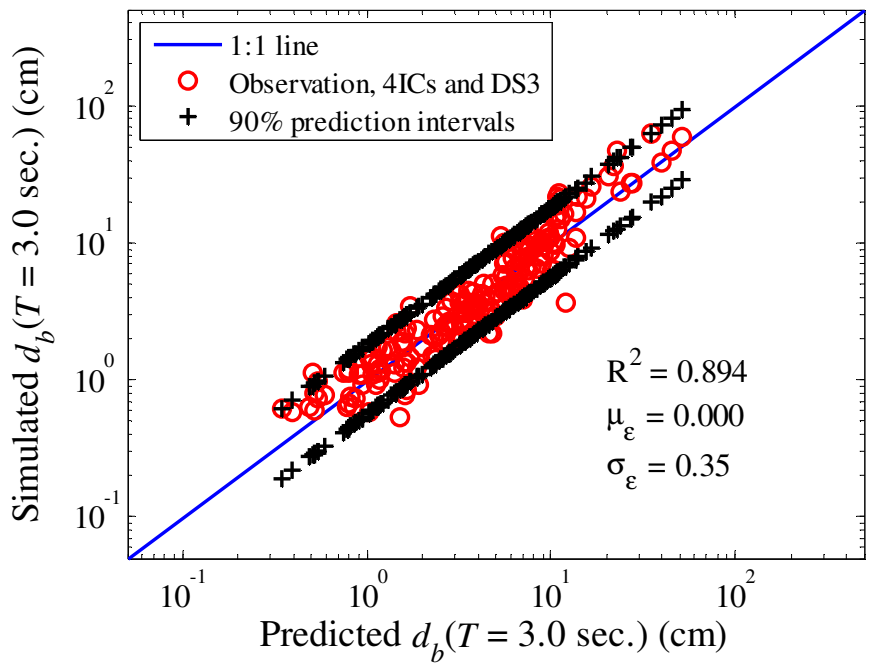

Figure 4.24. SI-ICR: Prediction intervals for DS3 (4DOF BBI model)

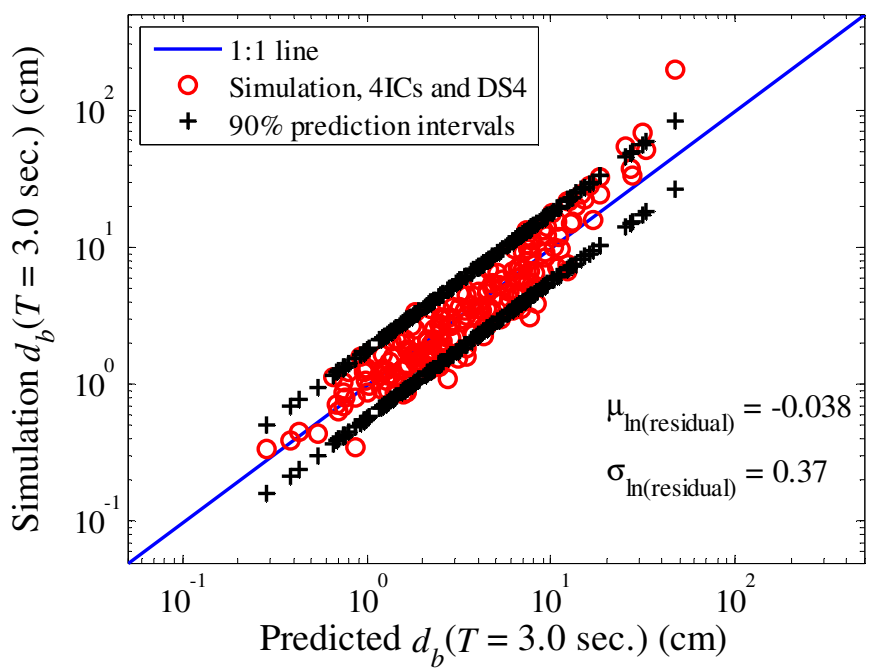

Figure 4.25. SI-ICR: Prediction intervals for DS4 (4DOF BBI model)

4.3 Peak Ground Velocity (PGV) Model Results

As stated in Section 3.2.4, a *PGV* model is developed in order to see the effectiveness of using one highly correlated *IM* as opposed to several proposed *IMs* in the PCR and SI-ICR methods. Similar to the other two models, the *PGV* model is first developed for the UBI and BBI cases using DS1 and DS3 respectively and checked using DS2 and DS4 respectively. Tables A.12 and A.13 show the regression coefficients, $\boldsymbol{\alpha}$, for the *PGV* model for datasets DS1 and DS3 respectively. The coefficient of determination, R^2, and the standard error, σ_{PGV}, of regression for both UBI (DS1) and BBI (DS3) models are presented in Figs. 4.26 and 4.27 respectively. R^2 values range from 0.76 to 0.83 and 0.77 to 0.83 for UBI and BBI models respectively. The σ_{PGV} values range from 0.38 to 0.54 and 0.38 to 0.53 for UBI and BBI models respectively. R^2 values tend to decrease with increasing post-yield periods and σ_{PGV} values tend to increase with increasing post-yield periods. This shows that using *PGV* alone as a model predictor is more suitable for certain post-yield time periods as opposed to the entire spectrum. Note that the standard error is relatively constant for all *T* for the case of PCR and SI-ICR models.

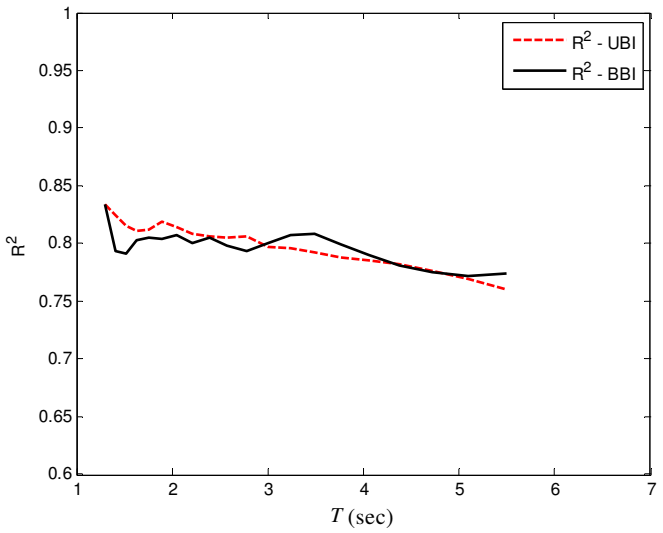

Figure 4.26. R^2 values for UBI and BBI systems

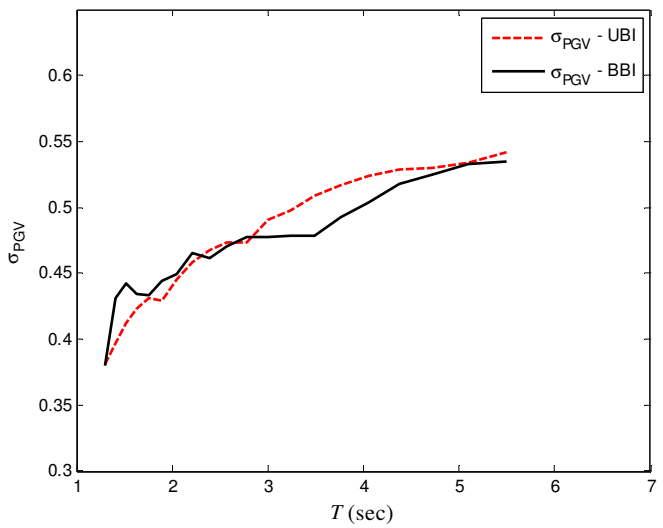

Figure 4.27. σ_ε values for UBI and BBI systems

Figs. 4.28 and 4.30 show the simulated values against the predicted values as well as the 90% prediction intervals for DS1 and DS3 respectively. Similar to the PCR and SI-ICR results, most of the predicted vs. simulated values lie well within the 1:1 lines. The results in Figs. 4.29 and 4.31 show the predicted values obtained from Eq. 3.3.5 and the simulated values for DS2

and DS4 respectively. The mean values of the residuals are 0.000 and 0.035 for DS2 and DS4 respectively which means that both the proposed UBI and BBI models are relatively unbiased (less than 5%). However, the standard error of the residuals are 0.46 and 0.48 for the DS2 and DS4 cases, which implies that the *PGV* model is relatively inaccurate and unreliable compared to PCR and SI-ICR methods especially at higher post-yield time periods.

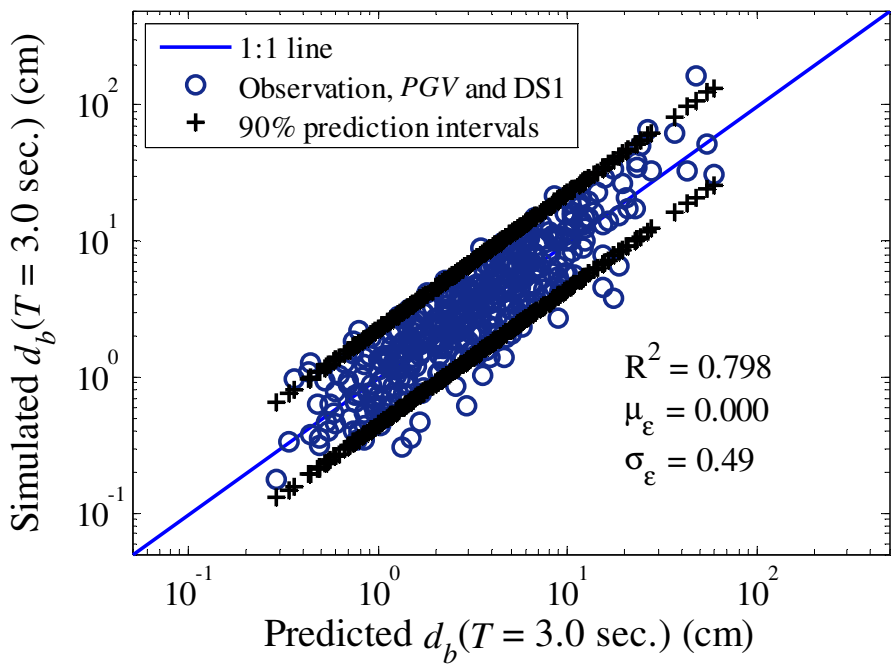

Figure 4.28. *PGV:* Prediction intervals for DS1 (2DOF UBI model)

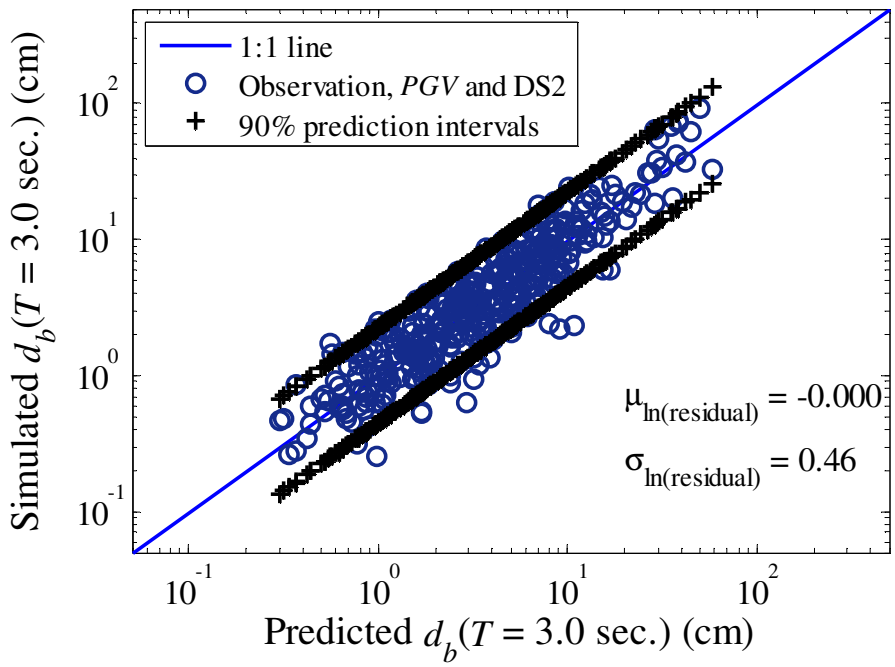

Figure 4.29. *PGV:* Prediction intervals for DS2 (2DOF UBI model)

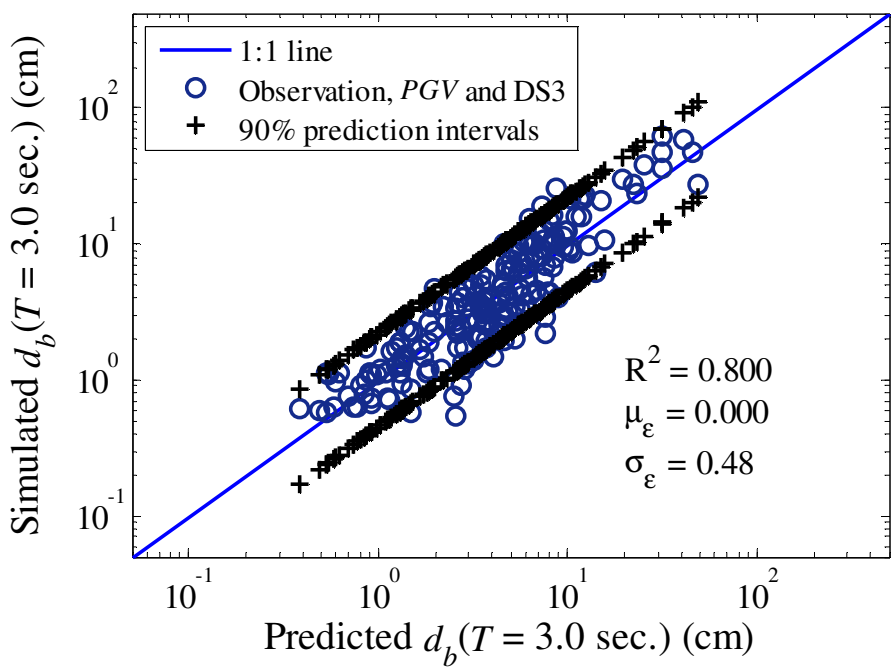

Figure 4.30. *PGV*: Prediction intervals for DS3 (4DOF BBI model)

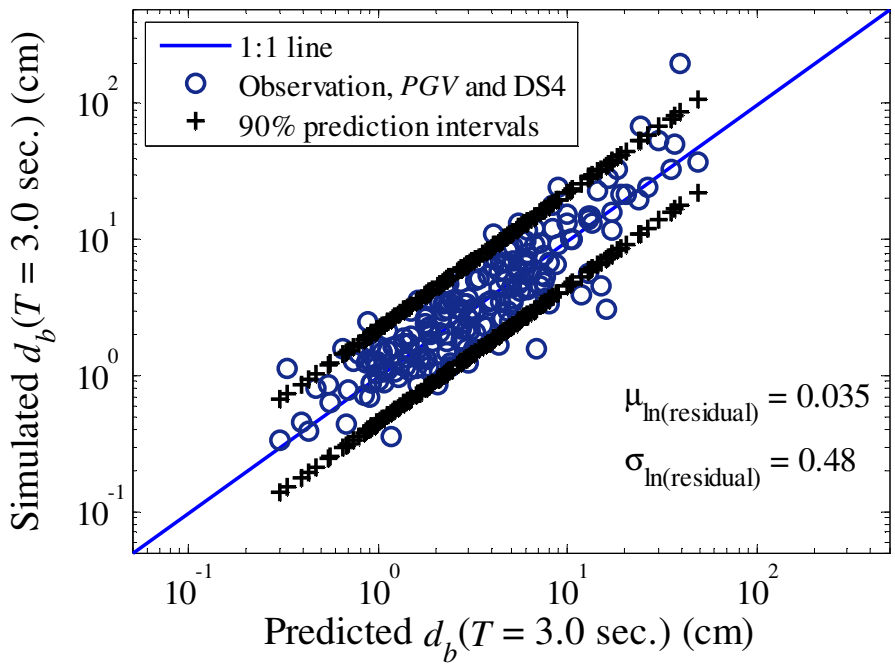

Figure 4.31. *PGV*: Prediction intervals for DS4 (4DOF BBI model)

4.4 Multi-Degree of Freedom (MDOF) Results

Figs. 4.32, 4.33, and 4.34 show the predicted versus simulated responses for the 12DOF structure presented in Section 3.2.4 using PCR, SI-ICR, and *PGV* models respectively. For all the three methods, most of the observations fall within the 90% prediction intervals with relatively low standard error of residuals for PCR and SI-ICR but the error is high for the *PGV* model. The bias is fairly low for all models with -0.006, -0.009, and 0.042 for PCR, SI-ICR, and *PGV* models respectively. All three models produce a bias of less than 5% which means they are all capable of predicting responses with a good degree of confidence.

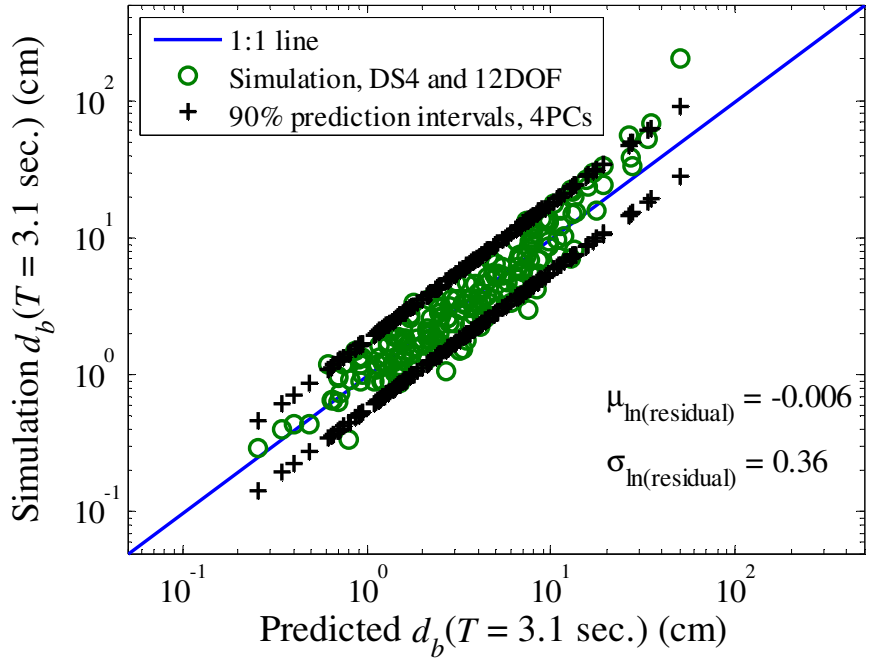

Figure 4.32. PCR: Predicted versus simulated responses for the MDOF structure

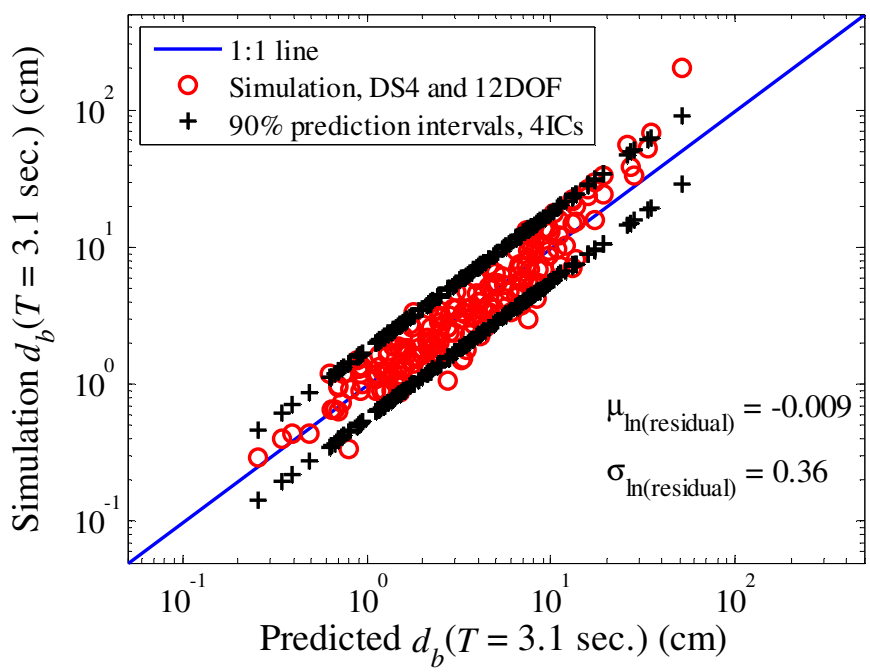

Figure 4.33. SI-ICR: Predicted versus simulated responses for the MDOF structure

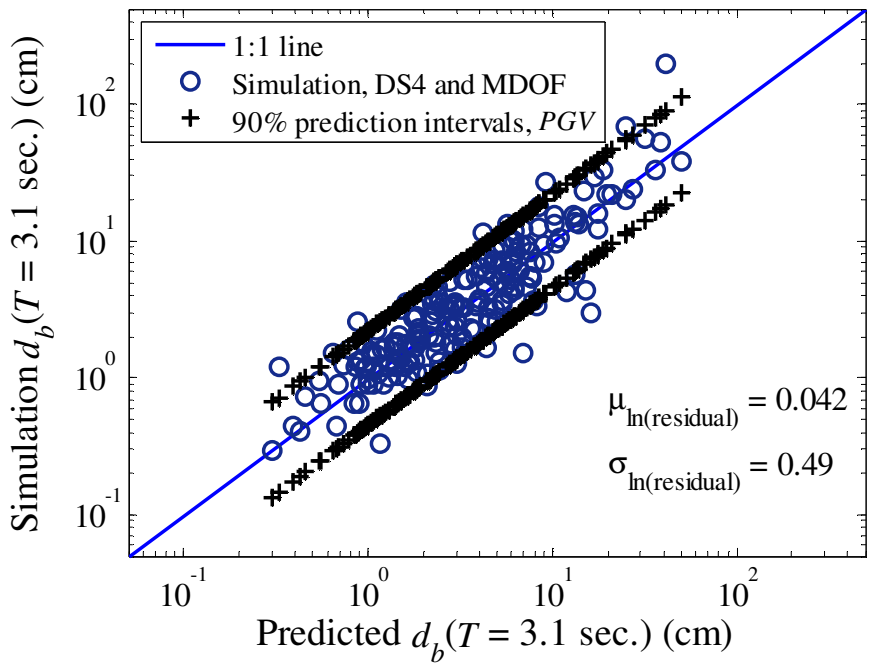

Figure 4.34. *PGV*: Predicted versus simulated responses for the MDOF structure

A good way to compare these methods is to look at their coefficients of determination (R^2) and standard error of the residuals (σ_ε) and compare these values between specific post-yield time periods as well as look at their averages across the time period spectrum. Table 4.1 shows the R^2 and σ_ε for selected post-yield time periods as well as average R^2 and σ_ε values across the T = 1.3-5.5 sec. spectrum. PCR and SI-ICR perform the best out of the three but on average, PCR is just slightly better than SI-ICR. If the numbers are observed closely, one will notice that the R^2 and σ_ε values, on average, for PCR and SI-ICR are at most within 0.03% and 0.14% difference of one another respectively. This is a very insignificant difference and it is fair to conclude that both PCR and SI-ICR methods in terms of goodness of fit and precision in the prediction are relatively the same. However, as components are dropped, SI-ICR performs slightly better than PCR (results not shown here). This however would be impractical because as more components are dropped, more data is lost. Thus, PCR is regarded as a better method when

pragmatically retaining four components. As expected, the *PGV* model does not perform as well as the other two methods for reasons previously outlined. On average, PCR and SI-ICR perform 5-6% better in terms of goodness of fit and precision compared to the *PGV* model. For the column labelled "Better" in Table 4.1, the '~' symbol indicates that the result is the same as above. Also, when written 'PCR/SI-ICR', it means that PCR is slightly better than SI-ICR and the same logic can be applied for 'SI-ICR/PCR'.

Table 4.1. R^2 and σ_ε for selected post-yield time periods and averages across spectrum

			PCR (4 PCs)	SI-ICR (4 ICs)	*PGV*	Better
DS1 (UBI)	T = 1.3 (sec)	R^2	0.8719	0.8716	0.8334	PCR/SI-ICR
		σ_ε	0.3348	0.3352	0.3808	~
	T = 2.6 (sec)	R^2	0.8970	0.8967	0.8051	~
		σ_ε	0.3449	0.3454	0.4732	~
	T = 5.5 (sec)	R^2	0.9101	0.9101	0.7599	SI-ICR/PCR
		σ_ε	0.3325	0.3324	0.5418	~
	Average	R^2	0.8957	0.8955	0.8001	PCR/SI-ICR
		σ_ε	0.3384	0.3388	0.4732	~
DS3 (BBI)	T = 1.3 (sec)	R^2	0.8626	0.8620	0.8344	~
		σ_ε	0.3482	0.3490	0.3803	~
	T = 2.6 (sec)	R^2	0.8854	0.8853	0.7982	~
		σ_ε	0.3561	0.3563	0.4702	~
	T = 5.5 (sec)	R^2	0.8913	0.8912	0.7744	~
		σ_ε	0.3734	0.3735	0.7744	~
	Average	R^2	0.8793	0.8790	0.7972	~
		σ_ε	0.3628	0.3633	0.4715	~

Table 4.2 shows the residual means and standard deviations for PCR, SI-ICR, and *PGV* methods. For the UBI system, the *PGV* model has a bias of zero which is lower than the PCR and SI-ICR values but it has a much higher standard error of 0.46. So even though the mean of the residuals is zero, the *PGV* model cannot predict the DS2 responses with as much precision as PCR and SI-ICR. The same can be observed for the standard error of residuals from DS4 and the MDOF responses. Overall, it can also be seen from Table 4.2 that PCR and SI-ICR perform relatively the same but significantly better than the *PGV* model.

Table 4.2. Residual means and standard deviations for PCR, SI-ICR, and *PGV* methods

		PCR (4PCs)	SI-ICR (4 ICs)	*PGV*	Better
DS2 (UBI)	$\mu_{\ln\,(residual)}$	0.008	-0.009	0.000	*PGV*
	$\sigma_{\ln\,(residual)}$	0.30	0.30	0.46	PCR, SI-ICR
DS4 (BBI)	$\mu_{\ln\,(residual)}$	-0.010	-0.038	0.0035	PCR
	$\sigma_{\ln\,(residual)}$	0.36	0.37	0.48	PCR
12DOF	$\mu_{\ln\,(residual)}$	-0.006	-0.009	0.042	PCR
	$\sigma_{\ln\,(residual)}$	0.36	0.36	0.49	PCR, SI-ICR

As mentioned in Section 2.4.3, the fundamental limitation of applying ICA is that the variables must be nongaussian (non-normal) for ICA to be possible (Hyvarinen *et al.*, 2001). Gustafsson (2005) reaffirms this point by stating that if one or more variables are gaussian, an ICR motivated algorithm will fail to recover the independent variables. This can definitely be used to explain what is happening in terms of how PCR and SI-ICR methods perform relative to one another in this application. A fact that was confirmed in this study is that the original variables (ln*IM*) have in fact gaussian distributions. Recall from Section 2.4.2 that kurtosis is a measure of non-gaussianity. The kurtosis of a perfectly normal distribution is 3. Distributions

that are more outlier-prone than a normal distribution have a kurtosis value of greater than 3; distributions that are less outlier-prone have a kurtosis value of less than 3 (MATLAB 2007b). Note that some definitions of kurtosis subtract 3 from the computed value, so that the normal distribution has a kurtosis value of 0 (see Eq. 2.4.8). However, the kurtosis function in MATLAB does not use this convention. So if one were to measure the kurtosis value of the original variables, it would be easy to see if an ICR algorithm (only applying ICA as a linear transformation) is suitable for this application.

The kurtosis values for lnPGA, lnI_A, lnPGV, lnE_v, and lnP_D for the UBI model (DS1) are 3.26, 2.97, 2.83, 2.76, and 3.05 respectively. For the BBI model (DS3), they are 2.48, 2.47, 2.97, 2.72, and 3.03 respectively. All these values indicate that the original variables are quite gaussian because they are very close to the value 3. For instance, Fig. 4.35 shows the histogram and normal probability plot of lnPGV for DS1. This indicates that the variable is quite normal; this agrees with its kurtosis value which is 2.83. The same is true for the other variables in both DS1 and DS2 datasets.

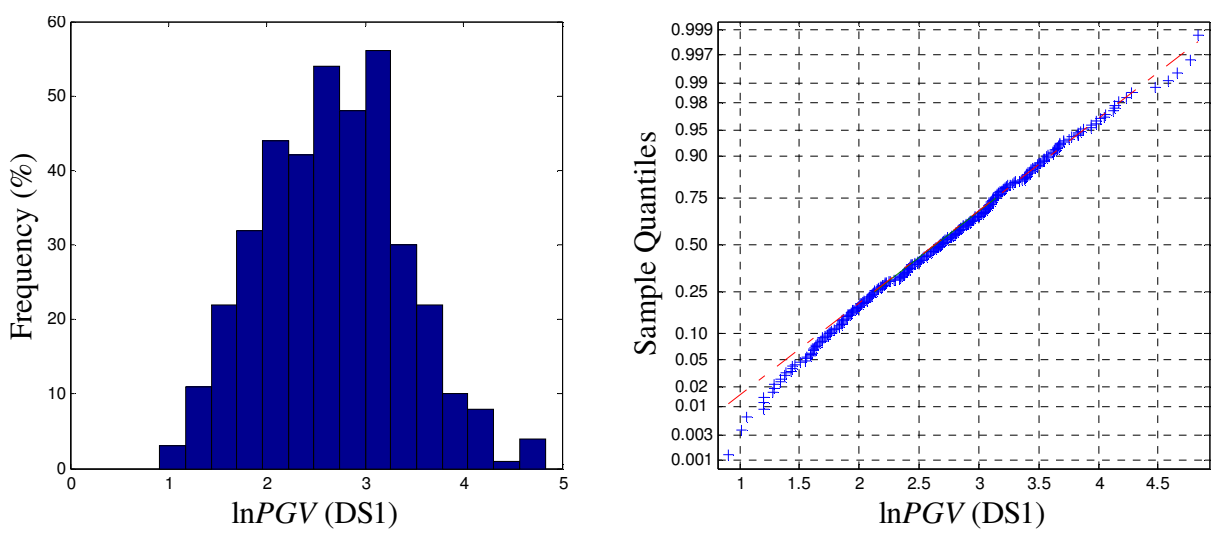

Figure 4.35. Histogram and normal probability plot of lnPGV for DS1

In order to explore this aspect further, ICR is performed on the UBI model without any pre-whitening or sorting. The following regression results are obtained for the UBI model (DS1, T = 3.0 sec.). Retaining four components, an R^2 and σ_ε value of 0.882 and 0.38 are obtained respectively. This compared with PCR's R^2 and σ_ε value of 0.910 and 0.33 for the same data set, is a noticeable decrease in performance. Even when all five components are retained in ICR, the R^2 and σ_ε values are still lower than PCR when retaining only four components. This reaffirms earlier observations that having Gaussian variables results in poor performance with ICR.

As mentioned in Section 2.4, SI-ICR works best with large sets of data. It was concluded that SI-ICR's effectiveness decreases as the size of data decreases. Based on this, it can be assumed that SI-ICR was not tested to its full effectiveness in this study. If the data sets were larger (greater than 387 and 193 samples for UBI and BBI models), then the performance of SI-ICR is expected to improve. In addition, the PCA algorithm organizes the scores in an order of most to least important based on the eigenvalues of the covariance matrix. However, SI-ICR's sorted input step arranges the scores differently in relation to decreasing correlation to the response variable. This means that if noise is added to the response measurements, SI-ICR is expected to perform better. Again, this would work better in conjunction with larger data sets.

The question remains, why do PCR and SI-ICR perform relatively the same for this application? The reason is not obvious, but perhaps it is due to the fact that SI-ICR contains a whitening step (same as PCA) which uncorrelates variables before applying ICA. Any processing done thereafter, with the exception of sorting, is just another linear transformation of the already uncorrelated data. This means that if ICA is excluded from SI-ICR (making it sorted PCR method), the performance will probably not be affected.

5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The main purpose of this study is to investigate the effectiveness of different regression methods with regards to estimating the base displacement responses of nonlinear base isolated structures, using a combination of carefully selected intensity measures as model inputs. Two methods, PCR and SI-ICR were studied in detail, and results from these models were compared with those from a single parameter *PGV* regression model for predicting responses of a uniaxial, biaxial, and multi-degree of freedom base isolated structure. Ground motions for this study were selected from both North American and Japanese earthquakes which were divided up into different data sets for the uniaxial and biaxial models. The probabilistic seismic demand model between the demand parameter and the intensity measures was modelled using multiple linear regression. Collinearity between the intensity measures was tackled by applying whitening to the variables. Once variables were uncorrelated, dimensionality reduction was possible. The effect of retaining only a subset of components was also studied. The *PGV* model was of course regressed using only the peak ground velocity as a model input and did not require any processing.

With respect to goodness of fit and standard error of residuals, PCR outperformed SI-ICR and *PGV* models. However, PCR's edge over SI-ICR was relatively small, while both the methods significantly outperformed the *PGV* model in many respects. On average across the post-yield time period spectrum, the R^2 values for PCR, SI-ICR, and *PGV* models in DS1 were 0.8957, 0.8955, and 0.8001 respectively. This results in PCR having a 0.02% and 11.27% performance edge over SI-ICR and *PGV* models respectively. Similarly, the σ_ε values for PCR, SI-ICR, and *PGV* models in DS1 were 0.3384, 0.3388, and 0.4732 respectively. This results in PCR having a 0.12% and 33.22% performance edge over SI-ICR and *PGV* models respectively. These results are similar to those in DS3 which dealt with the model development of the biaxial

base isolated structure. On average, PCR outperformed SI-ICR and *PGV* models by 0.03% and 9.79% in terms of goodness of fit respectively. In terms of σ_ε, PCR outperformed SI-ICR and *PGV* models by 0.14% and 26.06% respectively.

PCR and SI-ICR both predict the uniaxial and biaxial simulation results with a good degree of accuracy with standard deviation of errors staying within a constant range throughout the isolation period spectrum. For DS1, the standard deviation of errors ranged from 0.32 to 0.37 for both PCR and SI-ICR. However, this is not the case for the *PGV* model which had a range of 0.33 to 0.54. Goodness of fit values ranged from 0.86 to 0.92 for both PCR and SI-ICR while the *PGV* model had a range of 0.76 to 0.83. These results are similar to those computed for DS3. This shows that using a single parameter model can only perform well in certain ranges of the post-yield time period spectrum. The predicted results pertaining to the 12DOF model produced by PCR, SI-ICR, and *PGV* had relatively low biases with -0.006, -0.009, and 0.042 respectively. This means all models are capable of predicting the responses of a multi-degree freedom base isolated structure with a good level of confidence. However, as expected, the *PGV* model produced a standard deviation of errors of 0.49 which is approximately 31% higher than those produced by the PCR and SI-ICR models.

It is clear that when developing the regression model, PCR gives an overall better performance than the other regression methods. These results also confirm that using multiple *IMs* and properly dealing with their correlation effects with respect to one another, is much more effective than using only one highly correlated *IM* for developing the model, as has been undertaken extensively in many studies in the literature. Due to the small percent improvement that PCR has over SI-ICR, it is difficult to conclude which method is better quantitatively. But in terms of practicality with respect to the computational effort involved, it is fair to say that PCR is

a better method because of its relatively easy application. To recall, PCR applies whitening to uncorrelate the data then applies multiple linear regression. SI-ICR however, applies whitening, followed by ICA, sorting, and then regression. These extra steps involved in SI-ICR make it a less pragmatic approach than PCR for this application. This coupled with PCR's performance advantage over SI-ICR would deem the latter as an inferior approach for this application.

Further studies with regards to predicting the response of nonlinear base isolated structures should utilize a more comprehensive approach. The developed models should be tested to see whether they predict the responses of an unsymmetrical nonlinear base isolated structure with a good degree of confidence. A multi-degree of freedom base isolated structure such as the benchmark building presented by Narasimhan *et al.* (2005) would be ideal for this task. However, this multi-degree of freedom structure incorporates torsion as well as directional movements. Recall that the structural models used to develop and test the regression models were symmetric and only included displacements at each degree of freedom. As a result, future models should be developed with torsion in mind as well as unsymmetrical irregularities. This would make the prediction models a lot more comprehensive and realistic, as it would give the added effect of rotation due to unsymmetric geometries.

REFERENCES

Abdi, H., 2007. The eigen-decomposition: eigenvalues and eigenvectors, *Encyclopedia of Measurement and Statistics,* http://www.utd.edu/~herve.

Arias, A., 1970. A measure of earthquake intensity, in *Seismic Design for Nuclear Power Plants*, R. J. Hansen, Editor, M.I.T. Press, Cambridge, 438-483.

Azizi, N., Khoshnoudian, F., 2007, Nonlinear response of a torsionally coupled base-isolated structure, *Structures and Buildings,* 160 SB4, 207-219.

Chopra, A. K., 2007. *Dynamics of Structures: Theory and Applications to Earthquake Engineering*, Third Edition, Prentice-Hall, Inc., Upper Saddle River, N.J.

Comon, P., 1994. Independent component analysis – an new concept? *Signal Processing*, 36:287-906.

Datta, T.K., Jangid, R.S., 1995. Performance of base isolated systems for asymmetric buildings subjected to random excitation, *Engineering Structures*, 17, 443-454.

Deb, S.K., Paul, D.K., 1995. Thakkar, S.K., Simplified non-linear dynamic analysis of base isolated buildings subjected to general plane motion, *Engineering Computations,* 14, 542-557.

Deoskar, H.S., Makris, N., 1996. Prediction of observed response of base-isolated structure, *Journal of Structural Engineering,* 122, 485-493.

Erkus, B., Johnson, E.A., 2005. MATLAB implementation of smart base isolated benchmark building part III.

Fuladgar, A., Shakib, H., 2003. Effect of vertical component of earthquake on the response of pure-friction base-isolated asymmetric buildings, *Engineering Structures,* 1841-1850.

Gavin, H.P, Johnson, E.A, Nagarajaiah, S., Narasimhan, S., 2005. Smart base-isolated benchmark building. Part I: problem definition, *Struct. Control Health Monit.* 13:573–588.

Gustafsson, M.G., 2005. Independent component analysis yields chemically interpretable latent variables in multivariate regression, *J. Chem. Inf. Model,* 45, 1244-1255.

Hocking, R. R., Speed, F. M., and Lynn, M. J., 1976. A class of biased estimators in linear regression, *Technometrics*, 18, 425–437.

Housner, G. W., 1970. *Stong ground motion, in Earthquake Engineering,* R. L. Wiegel, Editor, Prentice-Hall, Inc., Englewood Cliffs, N.J., 75-91.

Housner, G. W., 1975. Measures of severity of earthquake ground shaking, in *Proceedings of*

U.S. National Conference on Earthquake Engineering, Earthquake Engineering Research Institute, June 18-20, 1975, Ann Arbor, Michigan, 25-33.

Housner, G. W., and Jennings, P. C., 1964. Generation of artificial earthquakes, *Journal of the Engineering Mechanics Division*, *ASCE*, 90(EM1), 113-150.

Hyvarinen, A., Karhunen, J., and Oja, E., 2001. *Independent Component Analysis,* Editor, John Wiley & Sons, New York.

Hyvarinen, and Oja, E., 2000. Independent Component Analysis: Algorithms and Applications*, Neural Networks,* 13(4-5): 411-430.

International Code Council (ICC). (2000). *International building code*, Falls Church, Va.

Jackson, J.E., 1991. *A User's Guide to Principal Components,* John Wiley & Sons, Inc., New York.

Jutten, C., and Hetault, J., 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1-10.

MATLAB (2007b), The Math Works, Inc., Natick, Massachusetts.

Nagarajaiah, S. and Reinhorn, A. M. and Constantinou, M. C., 1991. Nonlinear dynamic analysis of 3-D-base-isolated structures, *J. of Str. Engrg.* ASCE, 117, No 7, 2035-2054.

Nagarajaiah, S., Xiaohong, S., 2000. Response of base-isolated USC Hospital building in Northridge Earthquake, *Journal of Structural Engineering,* 126, 1177-1185.

Nagarajaiah, S., Xiaohong, S., 2001. Base-isolated FCC building: impact response in Northridge Earthquake, *Journal of Structural Engineering,* 127, 1063-1075.

Narasimhan, S., Linear Systems, Control and Identification, *Course Notes, University of Waterloo, Department of Civil and Environmental Engineering,* 2008.

Narasimhan, S., Pandey, M., Wang, M., Principal Component Analysis for Predicting the Response of Nonlinear Base-isolated Buildings, *Earthquake Spectra* (accepted, in press, 2008).

Nau, J. M., and Hall, W. J., 1984. Scaling methods for earthquake response spectra, *Journal of Structural Engineering*, *ASCE*, 110, 1533-1548.

Park, Y.-J., Ang, A. H.-S., and Wen, Y. K., 1985. Seismic damage analysis of reinforced concrete buildings, *Journal of Structural Engineering*, *ASCE*, 111, 740-757.

Park, Y.J., Wen, Y.K., and Ang, A.H.S., 1986, Random vibration of hysteretic systems under bi-directional ground motions, *Earthquake Engrg. Struct. Dyn.*, Vol. 14(4), pp. 543-557.

Ramallo, J. C., Johnson, E. A., and Spencer, B. J., 2002. "Smart" base isolation system, *Journal of Engineering Mechanics*, *ASCE*, 128, 1088-1099.

Riddell, R., 2007. On ground motion intensity indices, *Earthquake Spectra*, 23, 147-173.

Riddell, R., and Garcia, J. E., 2001. Hysteretic energy spectrum and damage control, *Earthquake Engineering and Structural Dynamics*, 30, 1791-1816.

Ryan, K. L. and Anil K. Chopra, 2004a. Estimation of seismic demands on isolators based on nonlinear analysis, *Journal of Structural Engineering*, ASCE, 130(3):392-402.

Ryan, K. L. And Anil K. Chopra, 2004b. Estimating the seismic displacement of friction pendulum isolators based on nonlinear response history analysis, *Earthquake Engineering and Structural Dynamics*, 33(3): 359-373.

Symans, M.D., Wongprasert, N., 2005. Numerical evaluation of adaptive base-isolated structures subjected to earthquake ground motions, *Journal of Engineering Mechanics,* 131, 109-119.

Trifunac, M. D., and Brady, A. G., 1975. A study on the duration of strong earthquake ground motion, *Bulletin of the Seismological Society of America, 65, 581-626.*

Weisberg, S., 1980. *Applied Linear Regression*, John Wiley & Sons, Inc., New York.

Weisberg, S., 2005. *Applied Linear Regression*, Third edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

APPENDIX

Table A.1. Ground Motion Database

Event	Earthquake	Date, Time	M	Lat. (°)	Long. (°)	Depth (km)	# sites
1	Parkfield	19660628, 0426	6.1	35.9550	-120.4980	10	2
2	San Fernando	19710209, 1400	6.6	34.4400	-118.4100	13	3
3	Coyote Lake	19790806, 1705	5.7	37.0845	-121.5050	9.6	5
4	Imperial Valley-06	19791015, 2316	6.5	32.6435	-115.3090	10	14
5	Coalinga-01	19830502, 2342	6.4	36.2330	-120.3100	4.6	16
6	Morgan Hill	19840424, 2115	6.2	37.3060	-121.6950	8.5	4
7	Nahanni, Canada	19851223, 0000	6.8	62.1870	-124.2430	8	3
8	Taiwan SMART1(40)	19860520, 0525	6.4	24.0817	121.5915	15.8	8
9	N. Palm Springs	19860708, 0920	6.0	34.0000	-116.6120	11	3
10	Taiwan SMART1(45)	19861114, 2120	7.3	23.9918	121.8332	15	15
11	Whittier Narrows-01	19871001, 1442	6.0	34.0493	-118.0810	14.6	9
12	Superstition Hills-02	19871124, 1316	6.7	33.0222	-115.8314	9	1
13	Loma Prieta	19891018, 0005	6.9	37.0407	-121.8829	17.5	13
14	Cape Mendocino	19920425, 1806	7.1	40.3338	-124.2294	9.6	5
15	Landers	19920628, 1158	7.3	34.2000	-116.4300	7	2
16	Northridge-01	19940117, 1231	6.7	34.2057	-118.5539	17.5	6
17	Duzce, Turkey	19991112, 0000	7.1	40.7746	31.1870	10	3
18	Western Tottori Pref	20001006, 1330	6.8	35.27	133.35	9	24
19	Akinada Setonaikai	20010324, 1528	6.8	34.13	132.69	46	25
20	Southern Iwate Pref	20011202, 2202	6.3	39.40	141.26	122	14
21	Northern Miyagi Pref	20030526, 1824	7.0	38.82	141.65	72	25
22	Northern Miyagi Pref	20030726, 0713	6.1	38.41	141.17	12	13
23	SE Off Tokachi	20030926, 0450	8.0	41.78	144.08	45	24
24	SE Off Kii Peninsula	20040905, 2357	7.5	33.14	137.14	44	3
25	Mid Niigata Pref	20041023, 1756	6.7	37.29	138.87	13	25
26	Off Nemuro Peninsula	20041129, 0332	7.0	42.95	145.28	48	25
27	Rumoi Region	20041214, 1456	5.8	44.08	141.70	9	5
28	NW Off Kyushu	20050320, 1053	6.7	33.74	130.18	9	23
29	E Off Miyagi Pref	20050816, 1146	7.1	38.15	142.28	42	23
30	E Off Izu Peninsula	20060421, 0250	5.5	34.94	139.20	7	6
31	Off Noto Peninsula	20070325, 0942	6.6	37.22	136.69	11	17
32	Off S Niigata Pref	20070716, 1013	6.7	37.56	138.61	17	23

Table A.2. Correlation coefficient matrix of *IMs* for DS1 and d_b (T = 3.0 sec.)

ρ	PGA	I_A, E_a, a_{rs}	P_a, a_{rms}	I_c	I_a	PGV	E_v, v_{rs}	P_v, v_{rms}	P_D	I_F	I_v	PGD	E_d, d_{rs}	P_d, d_{rms}	I_d	t_d	d_b
PGA	1.00	0.88	0.88	0.91	0.94	0.58	0.44	0.47	0.55	0.55	0.51	0.30	0.24	0.25	0.28	-0.04	0.30
I_A, E_a, a_{rs}		1.00	0.87	0.98	0.91	0.59	0.57	0.52	0.69	0.62	0.62	0.36	0.33	0.30	0.38	0.21	0.35
P_a, a_{rms}			1.00	0.94	0.73	0.55	0.42	0.56	0.56	0.48	0.39	0.27	0.18	0.27	0.21	-0.31	0.28
I_c				1.00	0.88	0.60	0.54	0.55	0.67	0.59	0.56	0.34	0.29	0.30	0.33	0.04	0.33
I_a					1.00	0.56	0.51	0.41	0.60	0.61	0.63	0.34	0.32	0.26	0.38	0.31	0.32
PGV						1.00	0.93	0.94	0.84	0.98	0.94	0.87	0.78	0.81	0.84	0.05	0.89
E_v, v_{rs}							1.00	0.93	0.87	0.95	0.94	0.91	0.89	0.87	0.92	0.26	0.91
P_v, v_{rms}								1.00	0.82	0.89	0.82	0.87	0.79	0.87	0.82	-0.11	0.90
P_D									1.00	0.86	0.85	0.69	0.65	0.63	0.70	0.23	0.72
I_F										1.00	0.98	0.87	0.81	0.80	0.88	0.24	0.89
I_v											1.00	0.85	0.81	0.76	0.88	0.41	0.86
PGD												1.00	0.96	0.97	0.99	0.15	0.86
E_d, d_{rs}													1.00	0.96	0.97	0.28	0.92
P_d, d_{rms}														1.00	0.94	0.04	0.85
I_d															1.00	0.32	0.87
t_d																1.00	0.11
d_b																	1.00

Table A.3. Correlation coefficient matrix of *IMs* for DS2 and d_b (T = 3.0 sec.)

ρ	PGA	I_A, E_a, a_{rs}	P_a, a_{rms}	I_c	I_a	PGV	E_v, v_{rs}	P_v, v_{rms}	P_D	I_F	I_v	PGD	E_d, d_{rs}	P_d, d_{rms}	I_d	t_d	d_b
PGA	1.00	0.89	0.90	0.92	0.95	0.58	0.44	0.49	0.55	0.55	0.51	0.35	0.24	0.28	0.33	-0.09	0.25
I_A, E_a, a_{rs}		1.00	0.88	0.99	0.92	0.61	0.58	0.55	0.70	0.63	0.63	0.44	0.37	0.35	0.45	0.18	0.33
P_a, a_{rms}			1.00	0.95	0.78	0.58	0.45	0.58	0.60	0.52	0.45	0.35	0.23	0.32	0.29	-0.30	0.25
I_c				1.00	0.90	0.62	0.55	0.58	0.69	0.61	0.58	0.42	0.33	0.35	0.41	0.01	0.31
I_a					1.00	0.57	0.51	0.45	0.59	0.60	0.61	0.39	0.33	0.29	0.42	0.24	0.29
PGV						1.00	0.92	0.95	0.85	0.98	0.94	0.89	0.78	0.82	0.86	0.02	0.81
E_v, v_{rs}							1.00	0.94	0.87	0.95	0.95	0.93	0.90	0.89	0.93	0.24	0.86
P_v, v_{rms}								1.00	0.84	0.92	0.86	0.90	0.81	0.88	0.85	-0.10	0.83
P_D									1.00	0.86	0.85	0.72	0.67	0.66	0.72	0.18	0.69
I_F										1.00	0.99	0.90	0.81	0.82	0.89	0.19	0.82
I_v											1.00	0.88	0.82	0.79	0.90	0.35	0.80
PGD												1.00	0.96	0.97	0.99	0.15	0.83
E_d, d_{rs}													1.00	0.96	0.96	0.28	0.80
P_d, d_{rms}														1.00	0.94	0.06	0.81
I_d															1.00	0.31	0.82
t_d																1.00	0.13
d_b																	1.00

Table A.4. Correlation coefficient matrix of *IMs* for DS3 and d_b (T = 3.0 sec.)

ρ	PGA	I_A, E_a, a_{rs}	P_a, a_{rms}	I_c	I_a	PGV	E_v, v_{rs}	P_v, v_{rms}	P_D	I_F	I_v	PGD	E_d, d_{rs}	P_d, d_{rms}	I_d	t_d	d_b
PGA	1.00	0.87	0.93	0.91	0.96	0.47	0.31	0.38	0.39	0.44	0.40	0.17	0.12	0.16	0.14	-0.26	0.19
I_A, E_a, a_{rs}		1.00	0.93	0.99	0.92	0.48	0.44	0.42	0.55	0.50	0.51	0.22	0.20	0.19	0.23	0.08	0.18
P_a, a_{rms}			1.00	0.97	0.88	0.52	0.39	0.48	0.50	0.48	0.44	0.21	0.15	0.20	0.17	-0.30	0.22
I_c				1.00	0.92	0.50	0.43	0.45	0.54	0.50	0.49	0.22	0.19	0.20	0.21	-0.05	0.20
I_a					1.00	0.44	0.34	0.34	0.43	0.45	0.45	0.18	0.16	0.15	0.18	0.01	0.16
PGV						1.00	0.94	0.97	0.80	0.99	0.96	0.86	0.80	0.83	0.83	-0.16	0.82
E_v, v_{rs}							1.00	0.97	0.85	0.96	0.96	0.91	0.89	0.88	0.91	0.06	0.84
P_v, v_{rms}								1.00	0.82	0.95	0.91	0.89	0.84	0.87	0.86	-0.19	0.86
P_D									1.00	0.82	0.83	0.65	0.62	0.60	0.65	0.07	0.65
I_F										1.00	0.99	0.87	0.83	0.84	0.86	-0.02	0.82
I_v											1.00	0.87	0.84	0.82	0.87	0.13	0.79
PGD												1.00	0.98	0.99	0.99	0.01	0.82
E_d, d_{rs}													1.00	0.98	0.98	0.10	0.78
P_d, d_{rms}														1.00	0.97	-0.05	0.80
I_d															1.00	0.13	0.80
t_d																1.00	-0.13
d_b																	1.00

Table A.5. Correlation coefficient matrix of *IMs* for DS4 and d_b (T = 3.0 sec.)

ρ	PGA	I_A, E_a, a_{rs}	P_a, a_{rms}	I_c	I_a	PGV	E_v, v_{rs}	P_v, v_{rms}	P_D	I_F	I_v	PGD	E_d, d_{rs}	P_d, d_{rms}	I_d	t_d	d_b
PGA	1.00	0.90	0.90	0.93	0.93	0.70	0.58	0.64	0.69	0.67	0.63	0.51	0.41	0.44	0.48	0.07	0.48
I_A, E_a, a_{rs}		1.00	0.87	0.99	0.92	0.71	0.69	0.67	0.80	0.72	0.71	0.58	0.51	0.50	0.58	0.27	0.56
P_a, a_{rms}			1.00	0.94	0.75	0.62	0.50	0.66	0.65	0.55	0.47	0.43	0.32	0.41	0.36	-0.18	0.41
I_c				1.00	0.89	0.70	0.65	0.69	0.77	0.68	0.65	0.54	0.46	0.49	0.52	0.12	0.52
I_a					1.00	0.72	0.67	0.61	0.74	0.75	0.76	0.58	0.52	0.48	0.61	0.42	0.56
PGV						1.00	0.92	0.94	0.89	0.98	0.94	0.89	0.79	0.82	0.86	0.25	0.89
E_v, v_{rs}							1.00	0.93	0.90	0.94	0.93	0.94	0.92	0.89	0.94	0.40	0.93
P_v, v_{rms}								1.00	0.88	0.90	0.84	0.89	0.81	0.88	0.83	0.10	0.89
P_D									1.00	0.90	0.88	0.78	0.71	0.71	0.77	0.33	0.80
I_F										1.00	0.99	0.90	0.82	0.81	0.90	0.40	0.90
I_v											1.00	0.89	0.82	0.78	0.92	0.54	0.89
PGD												1.00	0.95	0.95	0.99	0.34	0.92
E_d, d_{rs}													1.00	0.96	0.95	0.41	0.86
P_d, d_{rms}														1.00	0.92	0.22	0.87
I_d															1.00	0.49	0.91
t_d																1.00	0.35
d_b																	1.00

Table A.6. Regression Coefficients and Statistical Properties for DS1 (T = 3.0 sec.)

Coefficients	Estimation	Standard deviation	t-value	Pr (> | t |)
β_1	-0.376	0.007	-53.30	0.00
β_2	-0.521	0.019	-27.98	0.00
β_3	-0.459	0.033	-14.14	0.00
β_4	0.472	0.056	8.38	0.00
β_5	-0.263	0.120	-2.20	0.03

Table A.7. Regression Coefficients and Statistical Properties for DS3 (T = 3.0 sec.)

Coefficients	Estimation	Standard deviation	t-value	Pr (> | t |)
β_1	-0.395	0.012	-34.49	0.00
β_2	-0.440	0.026	-17.15	0.00
β_3	-0.441	0.043	-10.17	0.00
β_4	-0.380	0.091	-4.18	0.00
β_5	0.191	0.222	0.86	0.39

Table A.8. Regression coefficients using four PCs obtained from DS1

T	α_0	α_1	α_2	α_3	α_4	α_5	σ_ε
1.3	-2.619	0.375	-0.384	0.731	0.156	0.182	0.33
1.4	-2.659	0.365	-0.384	0.728	0.187	0.158	0.35
1.5	-2.628	0.345	-0.390	0.718	0.220	0.135	0.36
1.6	-2.730	0.324	-0.375	0.699	0.266	0.097	0.36
1.8	-2.738	0.322	-0.388	0.709	0.285	0.090	0.37
1.9	-2.792	0.313	-0.396	0.714	0.318	0.067	0.35
2.1	-2.925	0.307	-0.411	0.728	0.361	0.033	0.35
2.2	-2.947	0.292	-0.412	0.719	0.392	0.012	0.35
2.4	-3.078	0.266	-0.388	0.692	0.444	-0.031	0.35
2.6	-3.205	0.251	-0.367	0.674	0.479	-0.063	0.34
2.8	-3.249	0.228	-0.348	0.648	0.515	-0.092	0.33
3.0	-3.150	0.194	-0.343	0.615	0.550	-0.104	0.33
3.2	-3.100	0.166	-0.327	0.584	0.577	-0.111	0.32
3.5	-3.068	0.122	-0.291	0.526	0.621	-0.131	0.32
3.8	-3.054	0.091	-0.263	0.485	0.654	-0.150	0.32
4.1	-3.145	0.075	-0.238	0.460	0.682	-0.173	0.32
4.4	-3.213	0.065	-0.214	0.441	0.698	-0.194	0.32
4.7	-3.195	0.049	-0.185	0.409	0.709	-0.209	0.32
5.1	-3.193	0.031	-0.157	0.377	0.722	-0.227	0.32
5.5	-3.247	0.032	-0.147	0.372	0.728	-0.244	0.33

Table A.9. Regression coefficients using four PCs obtained from DS3

T	α_0	α_1	α_2	α_3	α_4	α_5	σ_ε
1.3	-3.196	0.401	-0.337	0.680	0.214	0.101	0.35
1.4	-2.819	0.373	-0.369	0.669	0.221	0.112	0.38
1.5	-2.761	0.378	-0.409	0.698	0.231	0.109	0.38
1.6	-2.958	0.384	-0.401	0.708	0.259	0.071	0.37
1.8	-2.979	0.376	-0.392	0.701	0.277	0.052	0.38
1.9	-2.969	0.363	-0.397	0.699	0.299	0.048	0.38
2.1	-2.903	0.339	-0.403	0.687	0.328	0.045	0.37
2.2	-2.865	0.297	-0.397	0.657	0.379	0.013	0.37
2.4	-2.908	0.294	-0.394	0.658	0.395	-0.005	0.35
2.6	-2.787	0.258	-0.377	0.620	0.422	-0.019	0.36
2.8	-2.665	0.215	-0.355	0.576	0.455	-0.031	0.36
3.0	-2.694	0.181	-0.332	0.544	0.497	-0.053	0.35
3.2	-2.752	0.167	-0.321	0.533	0.523	-0.063	0.34
3.5	-2.696	0.141	-0.301	0.504	0.546	-0.076	0.34
3.8	-2.614	0.104	-0.276	0.463	0.577	-0.092	0.34
4.1	-2.532	0.076	-0.252	0.429	0.596	-0.104	0.36
4.4	-2.403	0.043	-0.234	0.393	0.615	-0.108	0.37
4.7	-2.240	0.024	-0.233	0.374	0.620	-0.101	0.37
5.1	-2.150	0.010	-0.232	0.362	0.629	-0.097	0.37
5.5	-2.115	0.003	-0.228	0.356	0.633	-0.089	0.37

Table A.10. Regression coefficients using four ICs obtained from DS1

T	α_0	α_1	α_2	α_3	α_4	α_5	σ_ε
1.3	-2.658	0.389	-0.390	0.719	0.160	0.184	0.34
1.4	-2.700	0.379	-0.390	0.715	0.191	0.159	0.35
1.5	-2.673	0.360	-0.397	0.705	0.225	0.137	0.36
1.6	-2.777	0.340	-0.382	0.685	0.271	0.099	0.37
1.8	-2.787	0.339	-0.395	0.695	0.290	0.091	0.37
1.9	-2.843	0.331	-0.403	0.698	0.323	0.069	0.36
2.1	-2.981	0.326	-0.419	0.711	0.367	0.035	0.35
2.2	-3.007	0.312	-0.420	0.702	0.398	0.014	0.35
2.4	-3.141	0.287	-0.397	0.674	0.450	-0.029	0.35
2.6	-3.270	0.273	-0.376	0.655	0.485	-0.061	0.35
2.8	-3.315	0.250	-0.358	0.629	0.522	-0.090	0.33
3.0	-3.218	0.216	-0.352	0.596	0.557	-0.102	0.33
3.2	-3.167	0.189	-0.337	0.565	0.584	-0.110	0.33
3.5	-3.136	0.144	-0.300	0.507	0.628	-0.130	0.32
3.8	-3.122	0.113	-0.273	0.466	0.661	-0.148	0.32
4.1	-3.213	0.097	-0.247	0.441	0.689	-0.171	0.32
4.4	-3.281	0.087	-0.224	0.422	0.705	-0.192	0.32
4.7	-3.263	0.071	-0.195	0.390	0.716	-0.207	0.32
5.1	-3.260	0.053	-0.167	0.359	0.729	-0.225	0.32
5.5	-3.316	0.054	-0.157	0.354	0.735	-0.243	0.33

Table A.11. Regression coefficients using four ICs obtained from DS3

T	α_0	α_1	α_2	α_3	α_4	α_5	σ_ε
1.3	-3.371	0.459	-0.362	0.619	0.239	0.105	0.35
1.4	-2.995	0.431	-0.396	0.610	0.245	0.116	0.38
1.5	-2.947	0.440	-0.436	0.636	0.256	0.114	0.38
1.6	-3.138	0.444	-0.428	0.645	0.284	0.076	0.37
1.8	-3.151	0.433	-0.417	0.638	0.303	0.056	0.38
1.9	-3.136	0.419	-0.422	0.636	0.325	0.052	0.38
2.1	-3.066	0.393	-0.427	0.627	0.352	0.048	0.37
2.2	-3.018	0.348	-0.420	0.604	0.400	0.017	0.37
2.4	-3.062	0.346	-0.418	0.606	0.416	-0.001	0.35
2.6	-2.926	0.305	-0.397	0.574	0.441	-0.015	0.36
2.8	-2.785	0.254	-0.372	0.534	0.472	-0.028	0.36
3.0	-2.795	0.215	-0.347	0.506	0.512	-0.050	0.35
3.2	-2.846	0.198	-0.334	0.497	0.538	-0.061	0.34
3.5	-2.776	0.168	-0.312	0.472	0.559	-0.074	0.34
3.8	-2.679	0.125	-0.285	0.437	0.588	-0.091	0.35
4.1	-2.583	0.092	-0.259	0.406	0.606	-0.103	0.36
4.4	-2.442	0.056	-0.240	0.375	0.623	-0.107	0.37
4.7	-2.272	0.034	-0.237	0.358	0.627	-0.100	0.37
5.1	-2.178	0.019	-0.235	0.348	0.635	-0.097	0.37
5.5	-2.139	0.011	-0.231	0.342	0.639	-0.088	0.37

Table A.12. Regression coefficients using *PGV* model for DS1

T	α_0	α_1	σ_ε
1.3	-2.605	1.188	0.38
1.4	-2.571	1.200	0.40
1.5	-2.529	1.208	0.41
1.6	-2.512	1.224	0.42
1.8	-2.519	1.249	0.43
1.9	-2.523	1.274	0.43
2.1	-2.546	1.301	0.45
2.2	-2.533	1.314	0.46
2.4	-2.521	1.329	0.47
2.6	-2.516	1.341	0.47
2.8	-2.483	1.347	0.47
3.0	-2.465	1.357	0.49
3.2	-2.465	1.372	0.50
3.5	-2.459	1.384	0.51
3.8	-2.433	1.389	0.52
4.1	-2.438	1.400	0.52
4.4	-2.400	1.396	0.53
4.7	-2.324	1.377	0.53
5.1	-2.254	1.359	0.53
5.5	-2.199	1.344	0.54

Table A.13. Regression coefficients using *PGV* model for DS3

T	α_0	α_1	σ_ε
1.3	-2.684	1.193	0.38
1.4	-2.567	1.182	0.43
1.5	-2.575	1.204	0.44
1.6	-2.572	1.222	0.43
1.8	-2.529	1.230	0.43
1.9	-2.579	1.259	0.44
2.1	-2.626	1.287	0.45
2.2	-2.629	1.303	0.47
2.4	-2.596	1.309	0.46
2.6	-2.543	1.307	0.47
2.8	-2.520	1.312	0.48
3.0	-2.558	1.337	0.48
3.2	-2.615	1.368	0.48
3.5	-2.572	1.371	0.48
3.8	-2.533	1.373	0.49
4.1	-2.473	1.366	0.50
4.4	-2.432	1.364	0.52
4.7	-2.385	1.361	0.52
5.1	-2.375	1.369	0.53
5.5	-2.396	1.386	0.54

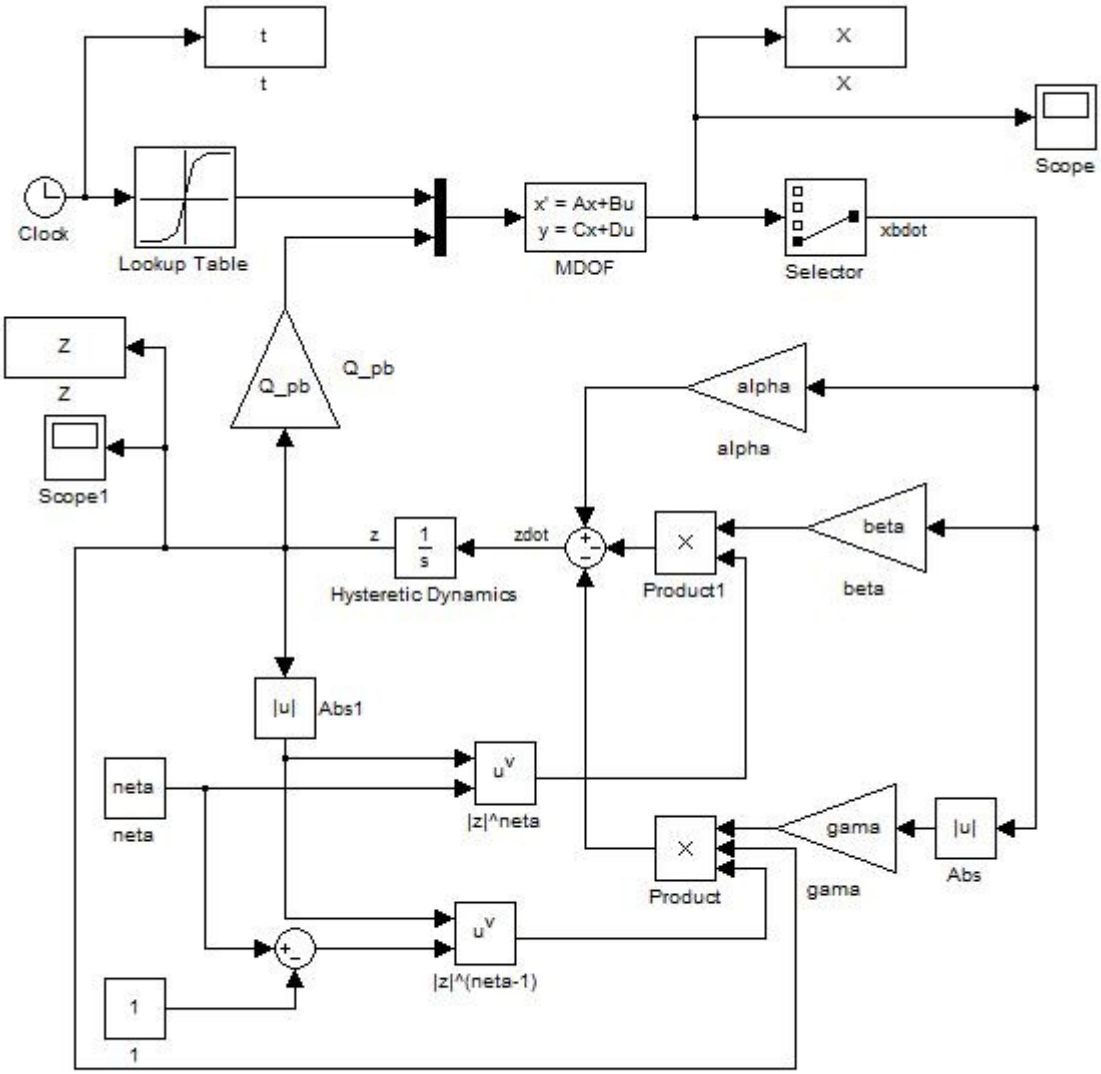

Figure A.1. SIMULINK representation of UBI system

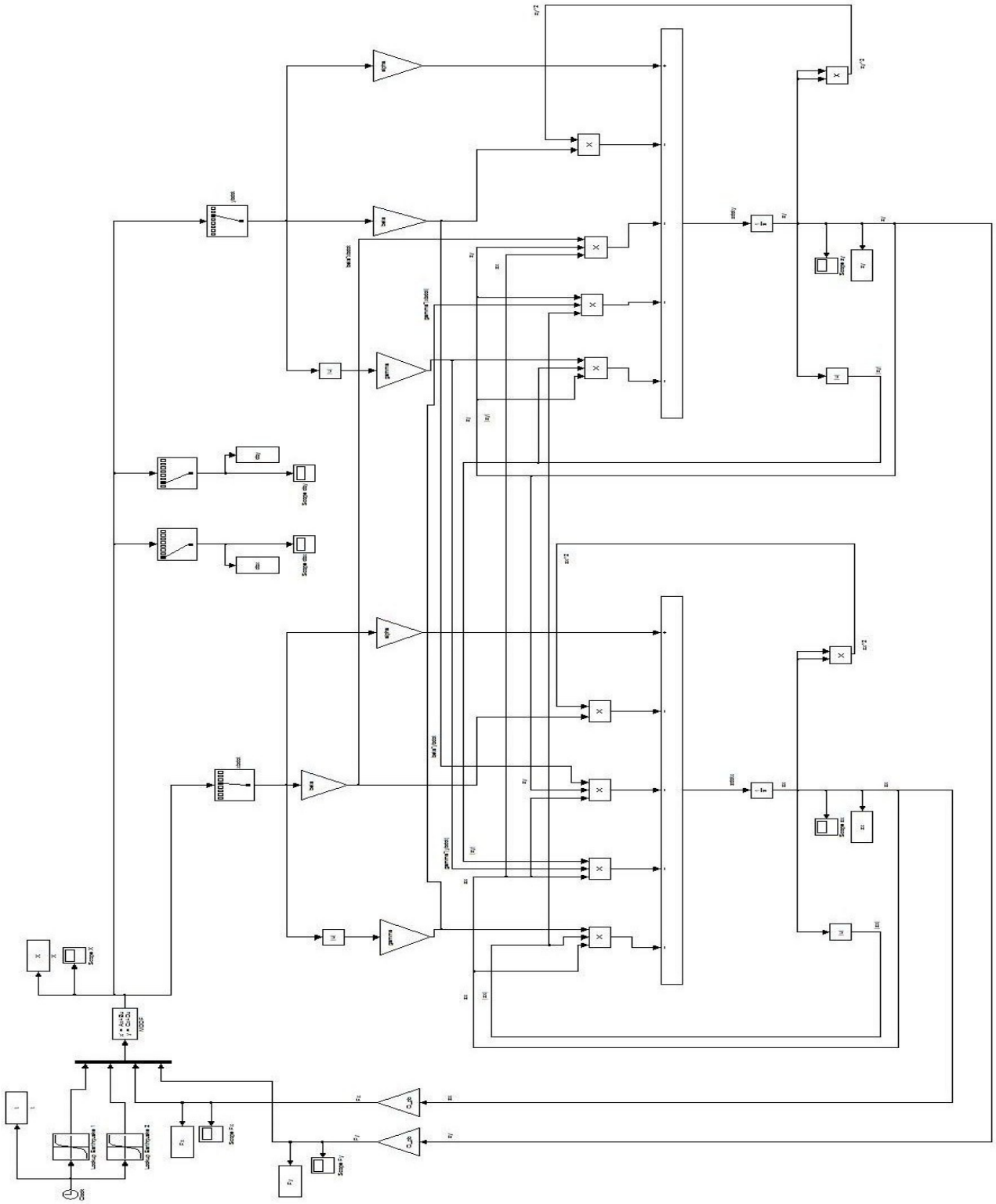

Figure A.2. SIMULINK representation of BBI system

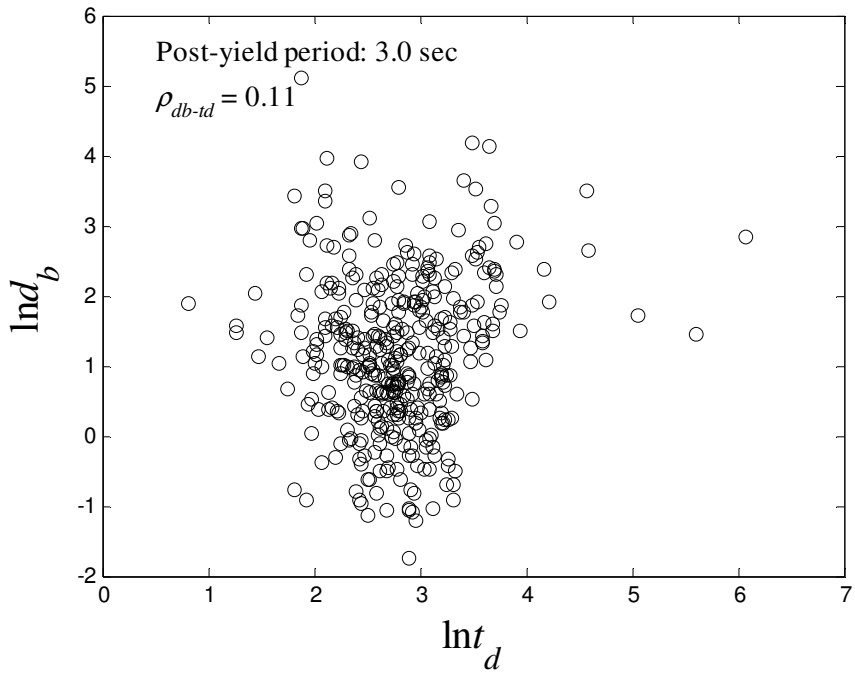

Figure A.3. Correlation between d_b and t_d (T = 3.0 sec., DS1)

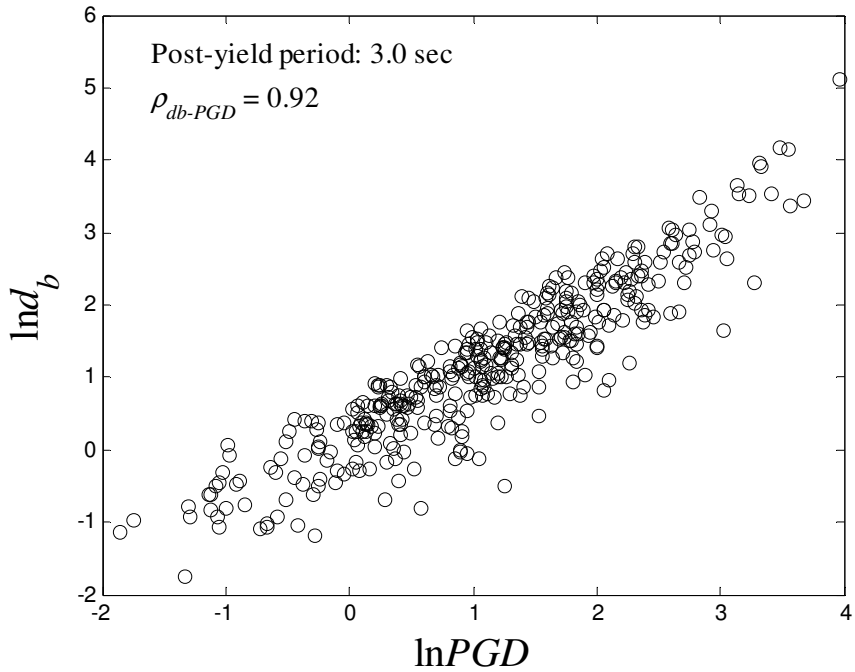

Figure A.4. Correlation between d_b and *PGD* (T = 3.0 sec., DS2)

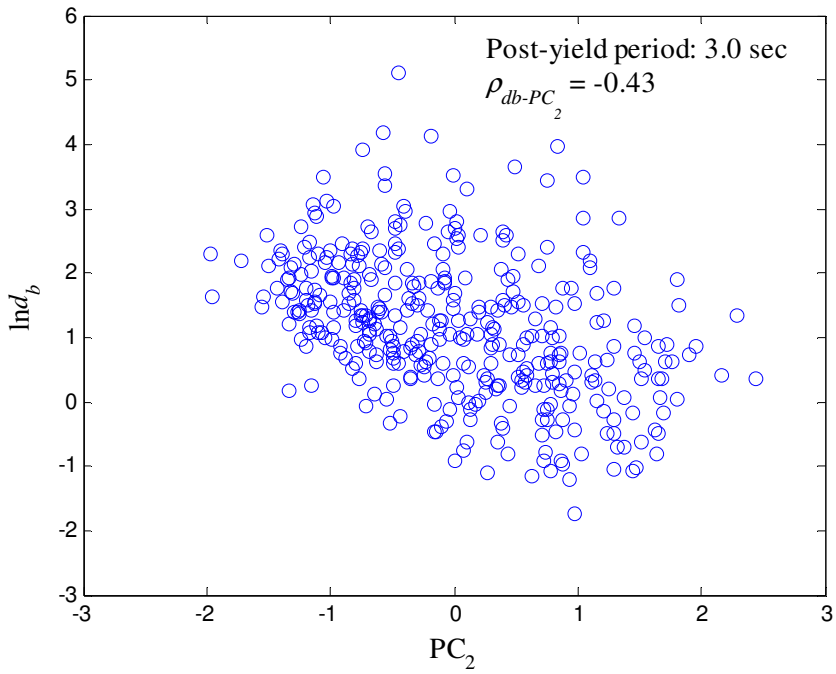

Figure A.5. Scatter plot of d_b to PC_2 for DS1

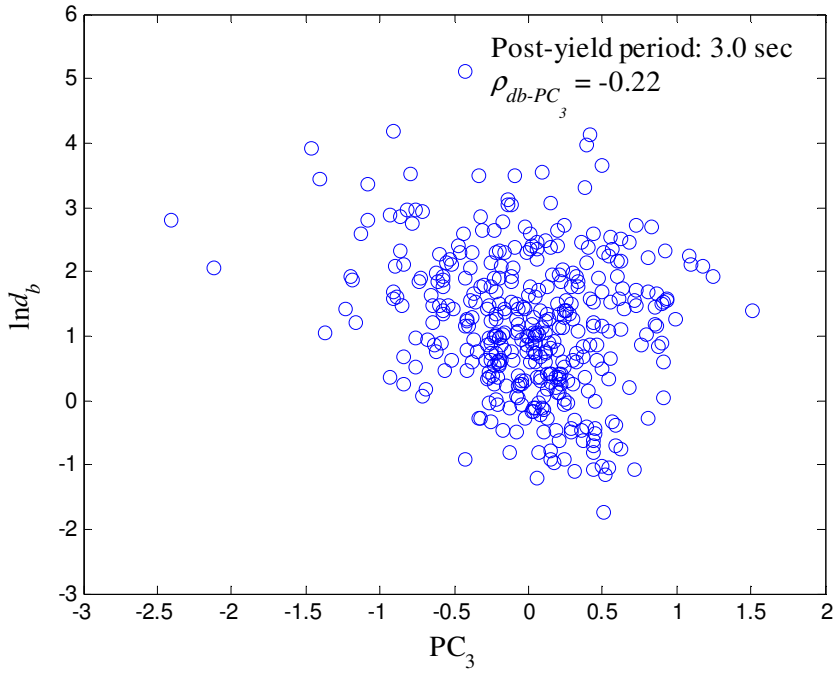

Figure A.6. Scatter plot of d_b to PC_3 for DS1

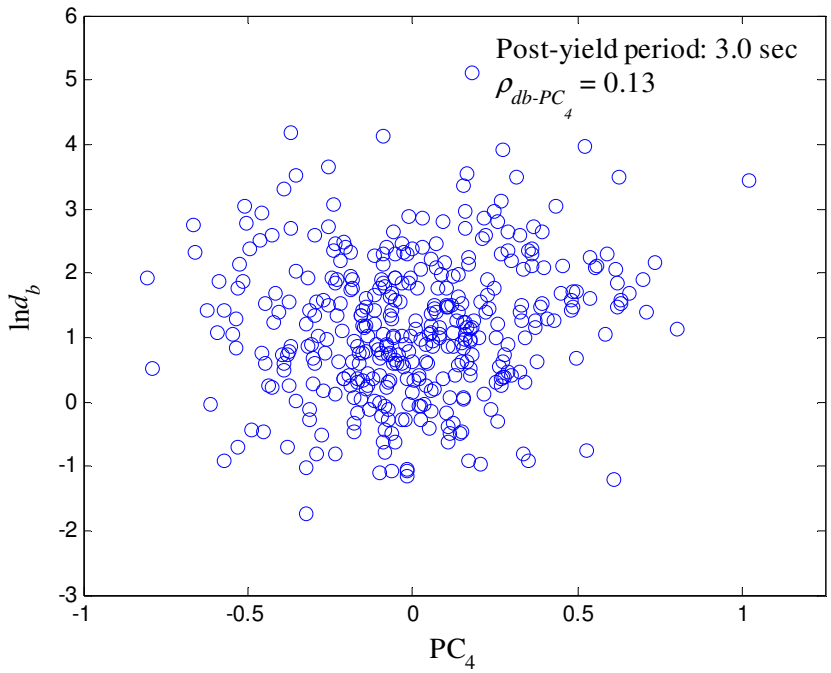

Figure A.7. Scatter plot of d_b to PC_4 for DS1

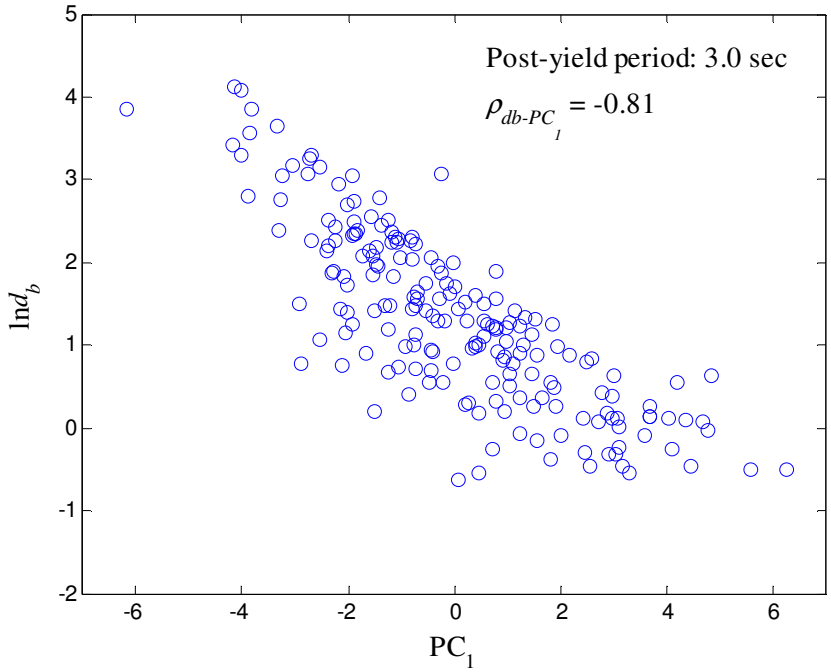

Figure A.8. Scatter plot of d_b to PC_1 for DS3

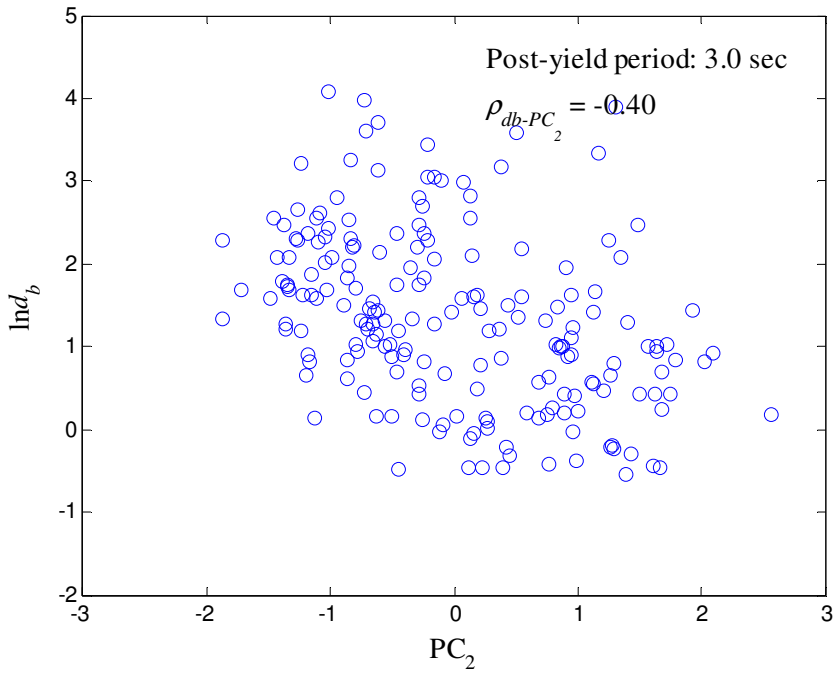

Figure A.9. Scatter plot of d_b to PC_2 for DS3

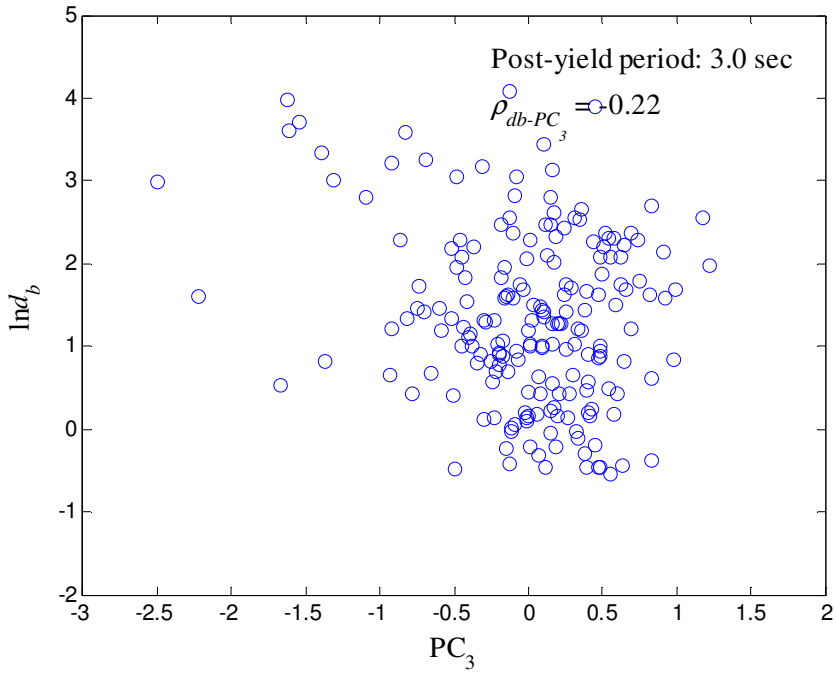

Figure A.10. Scatter plot of d_b to PC_3 for DS3

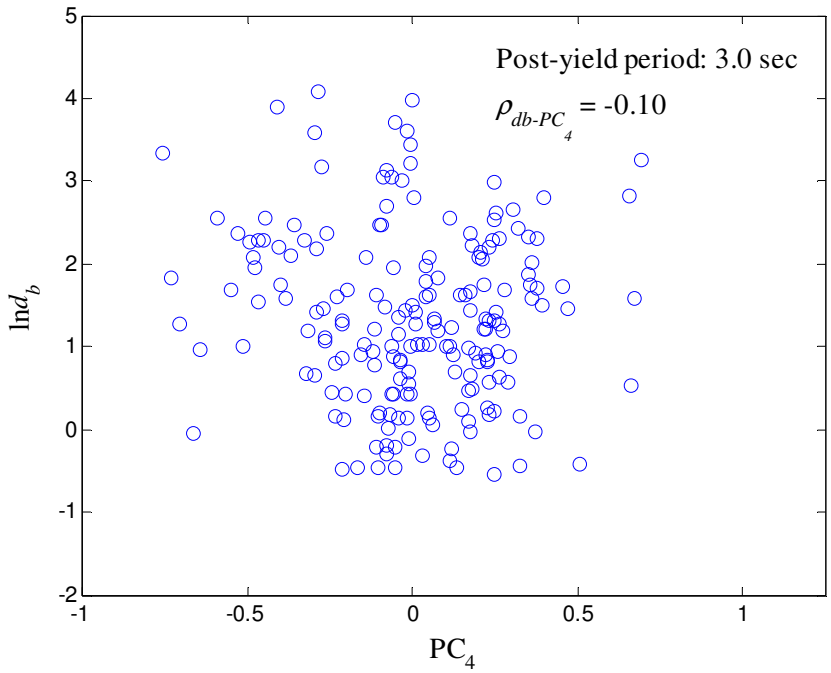

Figure A.11. Scatter plot of d_b to PC_4 for DS3

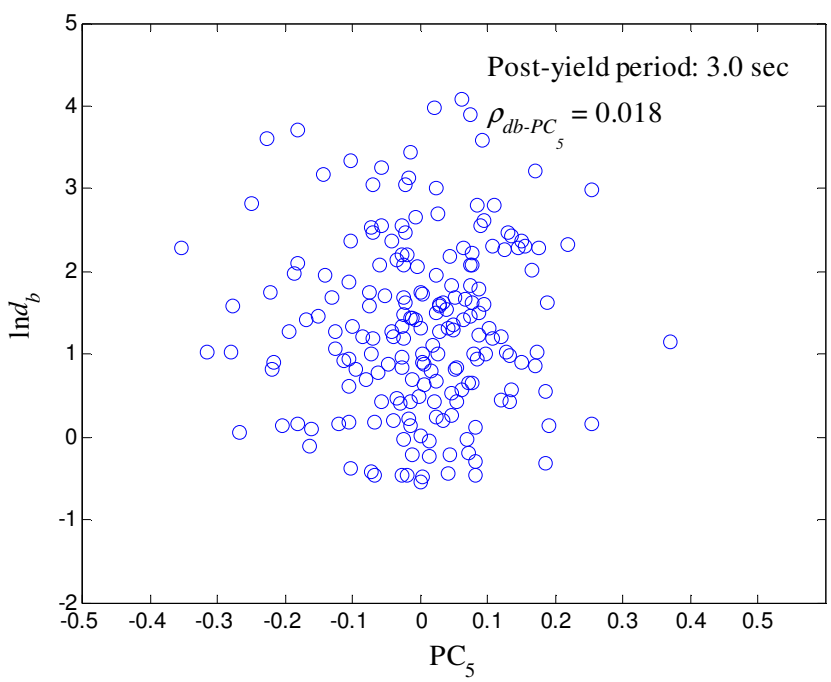

Figure A.12. Scatter plot of d_b to PC_5 for DS3

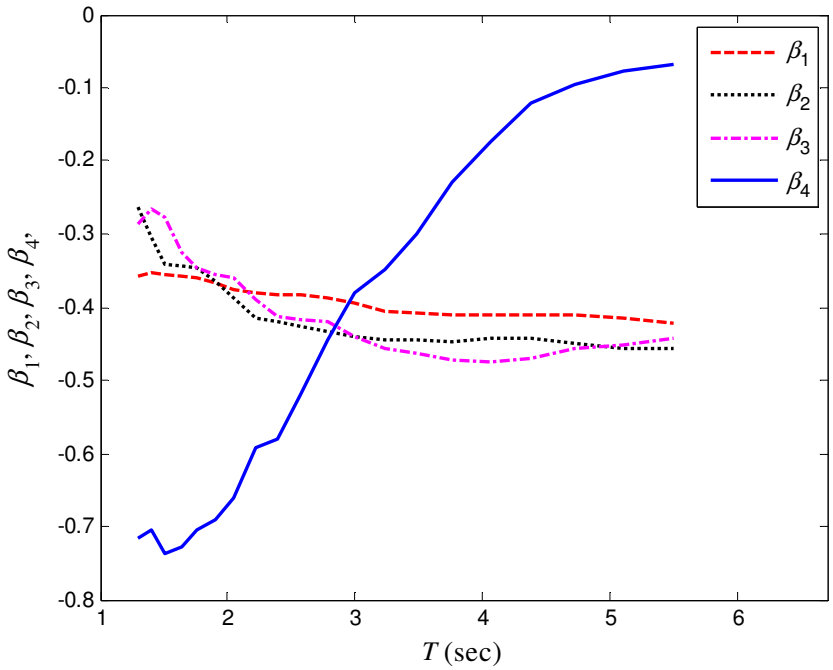

Figure A.13. β using four PCs for DS3

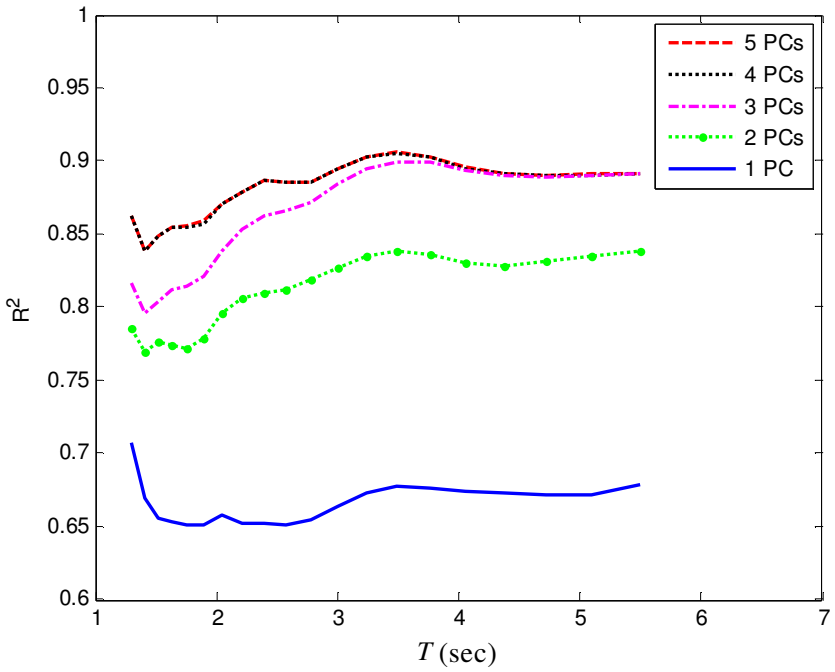

Figure A.14. R^2 values using various PCs for DS3

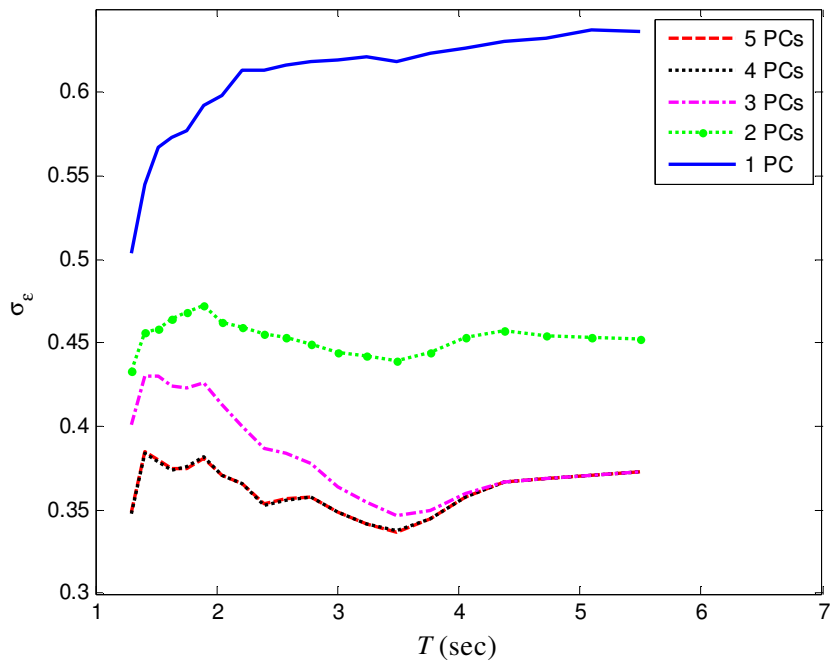

Figure A.15. σ_ε values using various PCs for DS3

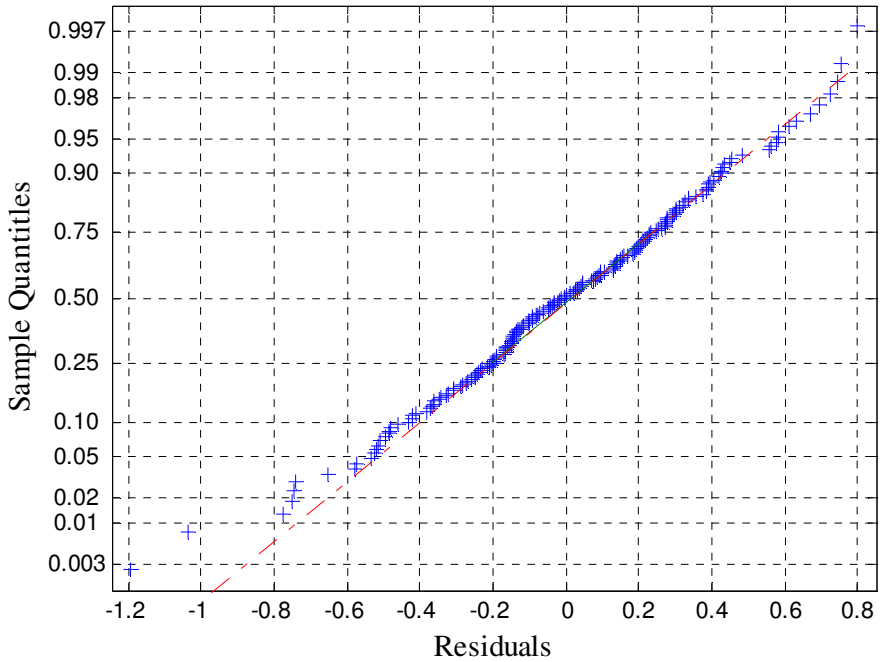

Figure A.16. Normal probability plot of residuals for DS3 using PCR (T = 3.0 sec.)

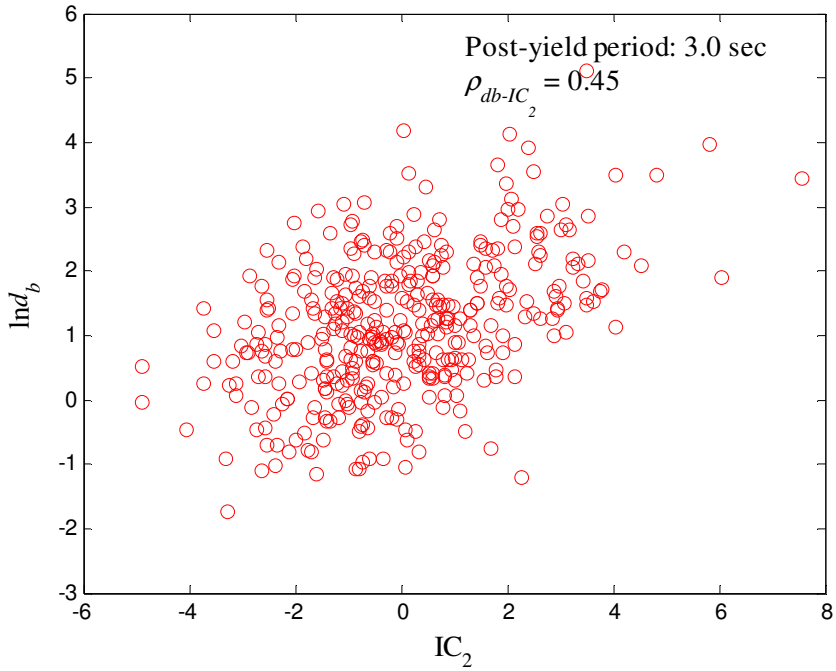

Figure A.17. Scatter plot of d_b to IC_2 for DS1

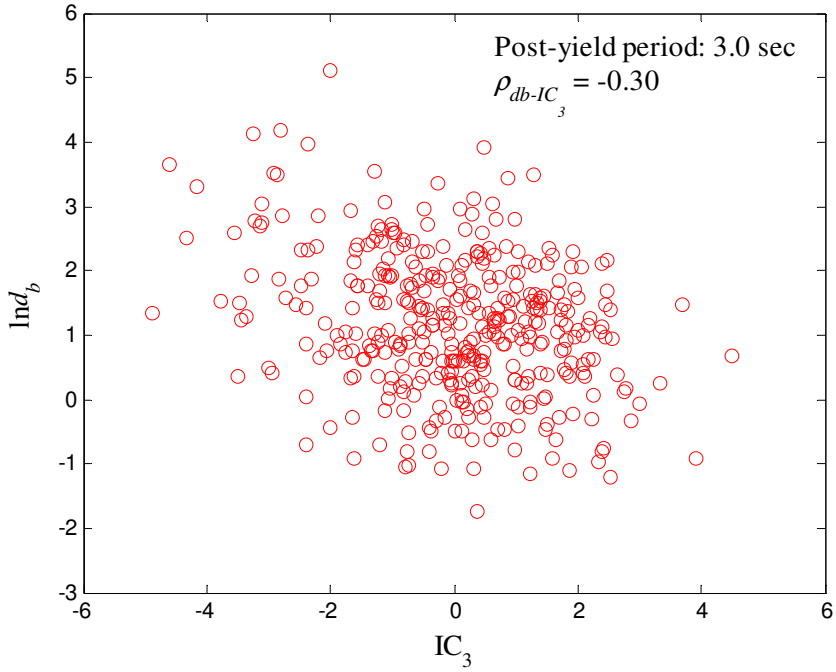

Figure A.18. Scatter plot of d_b to IC_3 for DS1

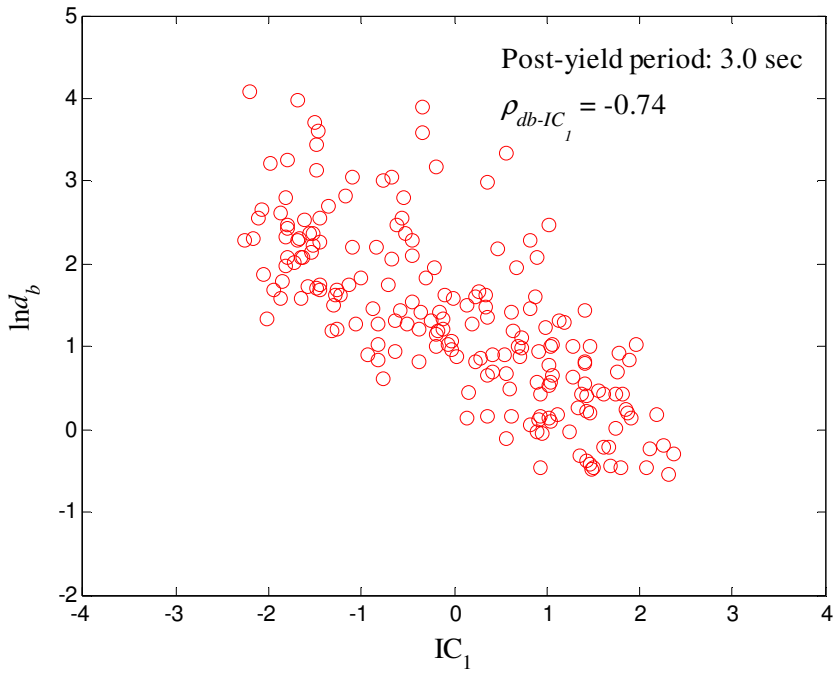

Figure A.19. Scatter plot of d_b to IC_1 for DS3

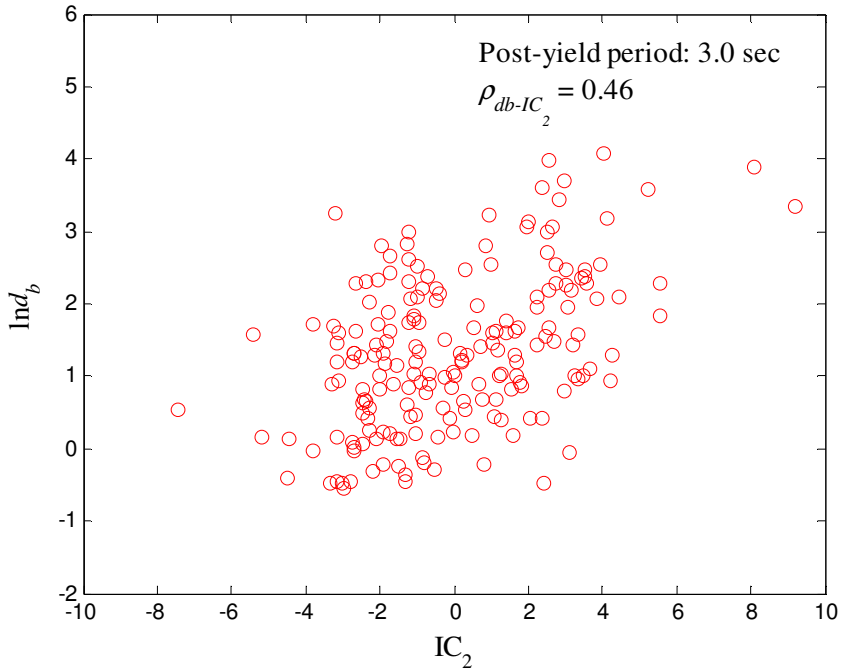

Figure A.20. Scatter plot of d_b to IC_2 for DS3

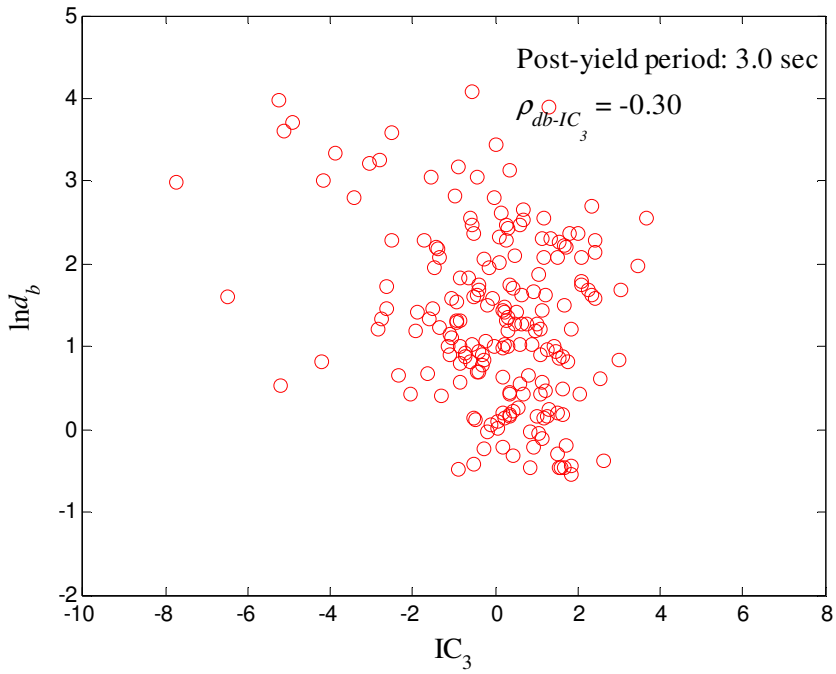

Figure A.21. Scatter plot of d_b to IC_3 for DS3

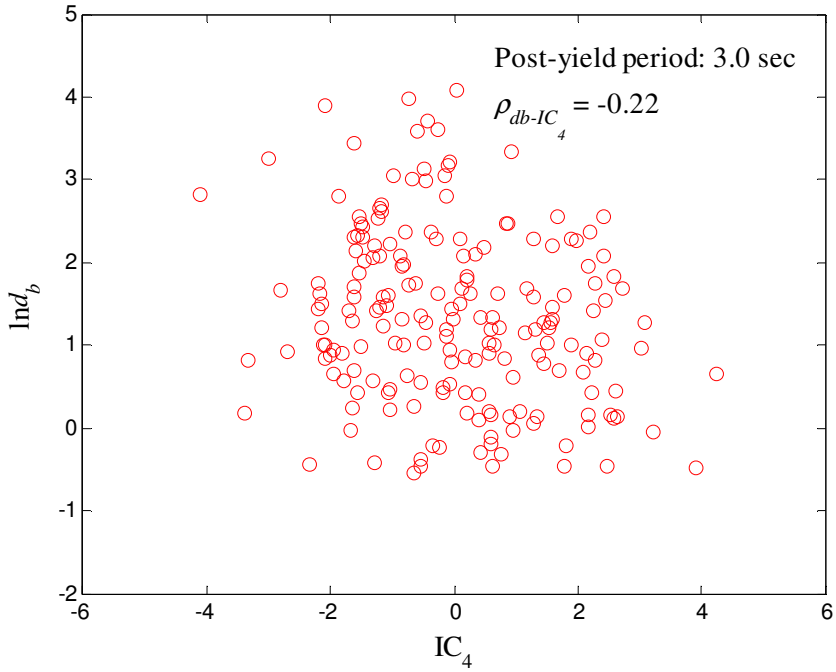

Figure A.22. Scatter plot of d_b to IC_4 for DS3

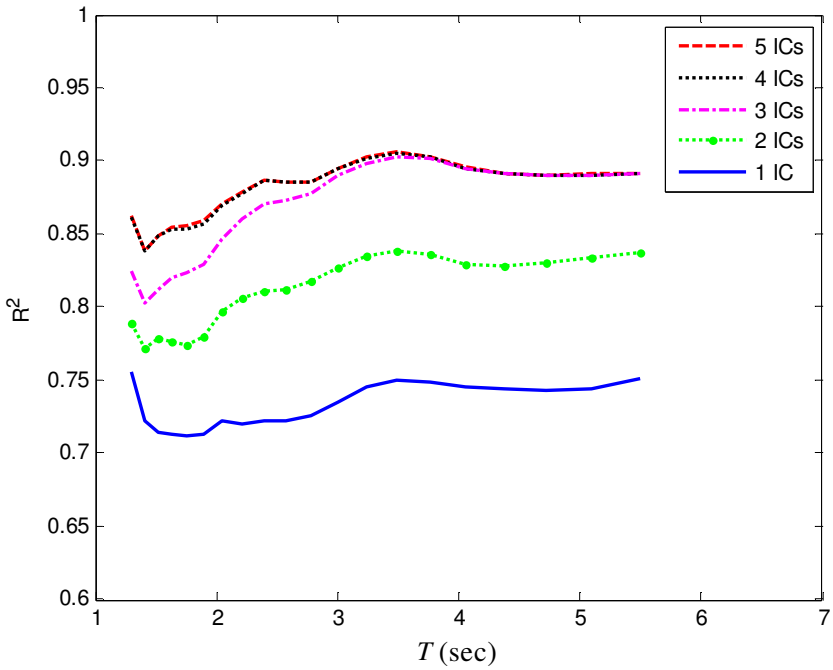

Figure A.23. R^2 values using various ICs for DS3

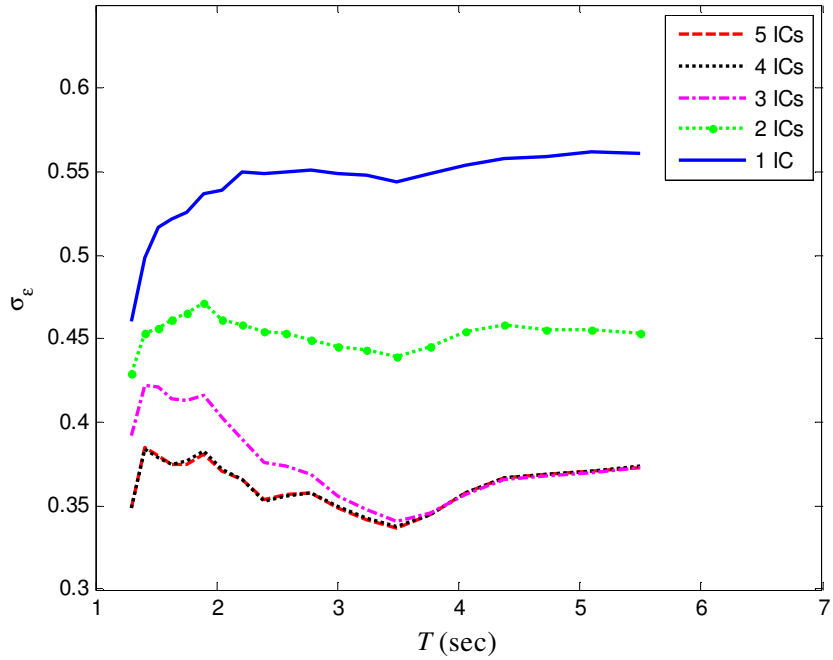

Figure A.24. σ_ε values using various ICs for DS3

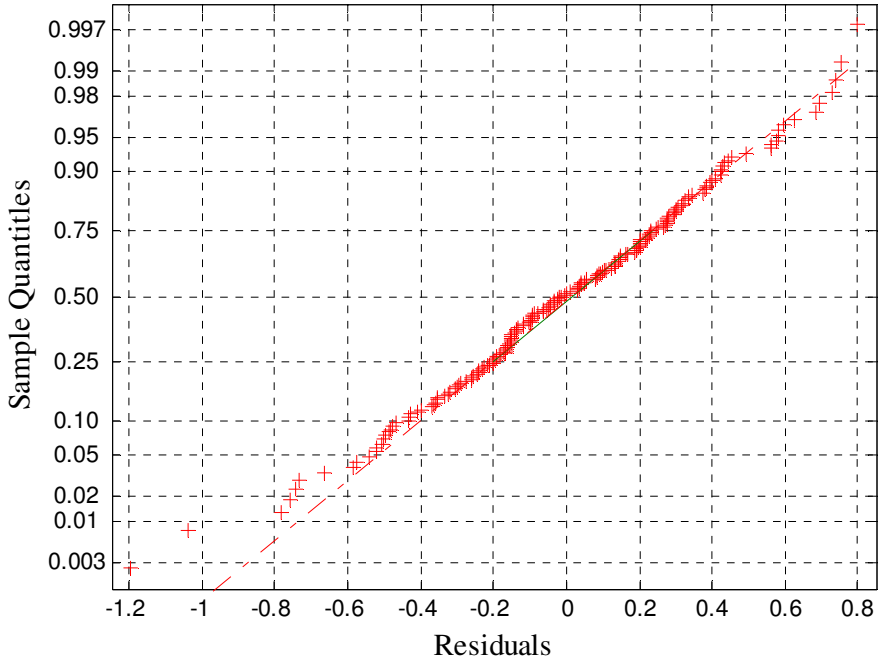

Figure A.25. Normal probability plot of residuals for DS3 using SI-ICR (T = 3.0 sec.)