International Journal of Concrete Structures and Materials

Prediction of Stress Increase at Ultimate in Unbonded Tendons Using Sparse Principal Component Analysis

Eric McKinney[1*], Minwoo Chang[3], Marc Maguire[2] and Yan Sun[1]

Abstract

While internal and external unbonded tendons are widely utilized in concrete structures, an analytical solution for the increase in unbonded tendon stress at ultimate strength, Δf_{ps}, is challenging due to the lack of bond between strand and concrete. Moreover, most analysis methods do not provide high correlation due to the limited available test data. The aim of this paper is to use advanced statistical techniques to develop a solution to the unbonded strand stress increase problem, which phenomenological models by themselves have done poorly. In this paper, Principal Component Analysis (PCA), and Sparse Principal Component Analysis (SPCA) are employed on different sets of candidate variables, amongst the material and sectional properties from a database of Continuous unbonded tendon reinforced members in the literature. Predictions of Δf_{ps} are made via Principal Component Regression models, and the method proposed, linear models using SPCA, are shown to improve over current models (best case R^2 of 0.27, measured-to-predicted ratio [λ] of 1.34) with linear equations. These models produced an R^2 of 0.54, 0.70 and λ of 1.03, and 0.99 for the internal and external datasets respectively.

Keywords: Principal Component Analysis, Sparse Principal Component Analysis, unbonded tendons, strand stress increase, LASSO

1 Introduction

The use of unbonded tendons, either internal or external, increases cost-efficiency, provides aesthetic satisfaction for users, and achieves fast and efficient construction (Cooke et al. 1981; Naaman 2005; Roberts-Wollmann et al. 2005). However, analysis of structures using unbonded tendons is exceptionally difficult and has been the subject of many international research projects, most of which attempt to simplify the problem considerably. Although numerous studies have been conducted to estimate the tendon stress increases at nominal strength, the analytic solution for the increase in unbonded tendon stress (Δf_{ps}) is challenging due to the lack of bond between strand and concrete, and most analysis methods

do not provide high correlation due to the limited available test data (Maguire et al. 2017).

Current design for unbonded tendon reinforced members in the United States uses American Concrete Institute 318 (ACI 318) (ACI 2008):

$$\Delta f_{ps} = 70 + \frac{f_c'}{\mu \rho_{ps}} \tag{1}$$

or American Association of State Highway and Transportation Officials Load and Resistance Factor Design (AASHTO LRFD) (AASHTO 2010) guidelines:

$$\Delta f_{ps} = 6200 \left(\frac{d_{ps} - c}{L} \right) \left(1 + \frac{N}{2} \right) \tag{2}$$

Both of the above methods are relatively easy for implementation in design. However, there are concerns with both. The ACI model is a curve fit to statistical data from only a handful of experimental data prior to 1978 (Mojtahedi and Gamble 1978; Mattock et al. 1971). The

*Correspondence: ericmckinney77@gmail.com
[1] Department of Mathematics and Statistics, Utah State University, Logan, UT 84322, USA
Full list of author information is available at the end of the article

McKinney *et al. Int J Concr Struct Mater* (2019) 13:20

Page 2 of 18

AASHTO method is not dependent on an experimental curve fit for Δf_{ps}, but is dependent on an estimation of the scaled plastic hinge length (ψ) from Tam and Pannell (1976). The ACI method especially is well liked by designers due to its simplicity for design.

There are considerably more prediction methods available in the literature as well as international design codes. Maguire et al. (2017) performed an in-depth review of various prediction methods based on the common mechanisms and empirical assumptions. The collapse mechanism model uses the relationship between strain, angle of rotation and applied load. The AASHTO LRFD method based on Roberts-Wollmann et al. (2005) and MacGregor (1989) is considered a collapse mechanism model. Other collapse mechanism models have been developed by the British Standard Institution (BSI 2001) and Harajli (2011) among others. Another category, called bond-reduction models, calculates a bond-reduction coefficient (Ω) to reduce the strength of a cross section unbonded reinforcement. Probably the most well-known bond reduction model was introduced by Naaman and Alkhairi (1991) and at one time was accepted in the 1994 AASHTO LRFD code, but later replaced in the 1998 AASHTO LRFD and also included statistical fitting to some degree. Alternatively, statistical analysis methods have been developed using the available experimental data of their time. The 1963 ACI code (ACI 1963) and European design codes, including German (DIN 1980) and Swiss (SIA 1979) codes, are widely accepted for design and real world application, and are statistically based. The 1963 and current ACI methods purposely under-predict strand stress increase in most cases and when compared to other methodologies provide closer to a lower bound prediction as opposed to an accurate prediction.

Maguire et al. (2014, 2017) indicated considerable phenomenological difference between Continuous unbonded tendon reinforced members, which are common, and simply supported members, which are uncommon in design. Interestingly, most methods from the literature compared prediction performance to a majority of simply supported members. In response, Maguire et al. (2017) compiled the largest known international database of 83 Continuous members, illustrating the dearth of data on this subject. This database only contains tests that have vetted and valid test setups and strand stress measurement. Considerable discussion was made to make clear the reasons for inclusion or exclusion of many test programs and even outlines future experimental needs. In order to consider multiple variables including internal and external tendons, Maguire et al. (2017) also suggested an update to the AASHTO LRFD collapse mechanism

model ($\psi = 14$ and $\psi = 18.5$ for internal and external tendons, respectively) based on statistical analysis and found nearly all types of prediction methods to have very low prediction accuracy with best case fit statistics R^2 of 0.27 and a best measured-to-prediction ratio (λ) of 1.34, neither of which indicates ideal prediction.

With the overall lack of available data and targeted research programs to drive improved phenomenological models for unbonded tendon reinforced structures, a statistical approach may provide the best prediction for Δf_{ps} (McKinney 2017). The advantages of a statistically based model are clear. Like the ACI equation, statistical models can be easily implemented, do not require excessive design time, and do not burden the engineer with several design iterations (e.g., bond reduction and collapse mechanism models). Furthermore, they can be optimized to fit the data and cross validation used to verify their accuracy.

The aim of this paper is to use advanced statistical techniques to develop a solution to the unbonded strand stress increase problem, which phenomenological models have done poorly (Maguire et al. 2017). While many engineers would prefer a phenomenological model, many also have affinity for the purely empirical ACI equation, which does not require complicated analysis, but has noted shortcomings. In this paper the authors present a novel approach to predict the increase in tensile strength in unbonded tendons using Principal Component Analysis (PCA), and Sparse Principal Component Analysis (SPCA). PCA is a statistical procedure to select significant variables by converting the variable information into the orthogonal base set (Jolliffe 2002). PCA has gained considerable popularity in structural engineering in recent years in combination with machine learning and structural health monitoring (Yan et al. 2008; Zhang et al. 2014) vibrations (Kuzniar and Waszczyszyn 2006; Hua et al. 2007; Kesavan and Kiremidjian 2012; Zolghadri et al. 2016; Zolghadri 2017) and image based crack detection (Abdel-Qader et al. 2004) because it is especially useful for analyzing large dataset with many variables. SPCA uses the Least Absolute Shrinkage and Selection Operator (LASSO) to reduce the contribution of relatively insignificant principal coefficients in the proposed statistical model, which simplifies the model further (Zou et al. 2006; Chang et al. 2017). Ultimately, the LASSO technique identifies the most important variables from a larger set in order to develop the most effective prediction equation with limited human influence.

The experimental and analytical literature is somewhat mixed on what the most important variables are for predicting tendon stress increase. Hemakom (1970) and Gebre-Michael (1970) tested five Continuous, one-way, slabs varying concrete strength the level of prestress, prestressing reinforcing ratio and pattern loading. They

found the percentage of prestressed reinforcement varied inversely with Δf_{ps}, while concrete strength varied directly with Δf_{ps}, while the level of effective prestress had no effect. Chen (1971) performed similar tests on two, one-way, slabs and found the distribution of cracks and moment capacity of the member were increased by including bonded reinforcement.

Trost et al. (1984) found the main factors influencing their experiments were compressive strength of the concrete and the level of prestress, and that Δf_{ps} was proportional to the sum of the deflections at the critical sections, while span-to-depth ratio was insignificant. Harajli and Kanj (1991) tested 26 simply supported beams with internal unbonded tendons. Beams varied span to depth ratio, loading, mild and prestressing reinforcement. This study found that as the mild reinforcing ratio decreased, the Δf_{ps} increased. Additional observations were that the type of loading (single point load or third point loads) and the span-to-depth ratio (ranging from 8 to 20) did not affect tendon stress increases, contradicting many analytical and experimental studies (Mojtahedi and Gamble 1978; Naaman and Alkhairi 1991; Lee et al. 1999).

Harajli et al. (2002) performed tests on nine, two-span Continuous, externally pretressed beam members and found that the geometry of load within a span, area of external prestressing steel and second order effects reduce Δf_{ps}. A reduction in steel stress with increase of span-to-depth ratio was also noticed and attributed to its influence on plastic hinge length and rotation capacity.

Lou and Xiang (2006), validated a finite element model on the Harajli and Kanj (1991) dataset. This numerical investigation found that a significant increase in Δf_{ps} can be found with an increase of yield stress of the bonded reinforcement. Furthermore, the stress increase was shown to decrease significantly with an increase of the combined reinforcing index, but this was attributed to the change in mild steel reinforcing index, verifying similar behavior from Du and Tao (1985).

Ozkul et al. (2008) performed an experimental investigation of 25 simply supported members with internal unbonded tendons. The experimental results showed effective prestressing and area of prestressed reinforcement, but mild steel and concrete strength were not important even though plastic hinge lengths were affected by the mild steel provided. There was an inverse relationship noted between Δf_{ps} and the prestressed reinforcement indices that was attributed to sharing of tensile force between prestressed and nonprestressed reinforcement. Lou et al. (2013) in a numerical investigation, calibrated a FEM to two-span members tested by Harajli et al. (2002) indicated that Δf_{ps} in external tendons of Continuous beams is most strongly related to rotational capacity and non-prestressed reinforcement.

The above summary of experimental and analytical literature conflicts on nearly every investigated variable. The reason for this is likely the relatively focused nature of their investigations. In order to identify the variables that are most important, this paper uses the LASSO technique with SPCA to identify the variables of most importance from a large dataset.

This paper focuses on improving the accuracy of Δf_{ps} predictions for internally and externally reinforced unbonded tendons separately. Sets of candidate variables, amongst the material and geometric properties from the database compiled by Maguire et al. (2017), are considered to analyze the significant factors in the database for prediction of Δf_{ps}, and to construct models. It is acknowledged that variables like deviator type and location are important to the prediction of design, but since this information is not present in the database, for the purposes of this investigation, second order effects are neglected. The performance of all of the PCA models are compared against a benchmark PCA model involving all of the variables. Likewise, the authors compare the SPCA models to a SPCA benchmark. Additionally, these predictions are compared to other prediction methods from the literature on the same database. The results show that improvements in predictions can be made with a simplified SPCA regression model.

2 Principal Component Analysis (PCA) and Sparse PCA (SPCA)

PCA is a widely used statistical technique for dimension reduction. It takes linear combinations of all of the variables to create a reduced number of uncorrelated variables (called principal components, or PC's) that still express a majority of the information from the original data (Lattin et al. 2003). The number of principal components selected, which is usually much smaller than the number of original variables, is determined by considering how much information is retained at the cost of simplifying the data. In addition to dimension reduction, another typical scenario where PCA works well is when a level of collinearity exists in the data, i.e., some or all of the predictor variables are correlated. After applying PCA, the resulting principal components are uncorrelated, and hence the replication of information in the original variables is removed.

Let $X = [x_{ij}]$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, be the $n \times p$ data matrix of n observations on the p-dimensional random vector $X = [X_1, X_2, \ldots, X_p]^T$. Define the $1 \times p$ mean vector $\bar{x}$ as

$$\bar{x} = \left[\frac{1}{n} \sum_{i=1}^{n} x_{i1}, \ldots, \frac{1}{n} \sum_{i=1}^{n} x_{ip} \right] \tag{3}$$

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 4 of 18

That is, the jth element of $\bar{x}$ is the sample mean of the jth variable. The $p \times p$ sample covariance matrix S is computed as

$$S = \frac{1}{n-1}(X - \mathbf{1}_n \bar{x})^T (X - \mathbf{1}_n \bar{x}), \qquad (4)$$

where $\mathbf{1}_n$ is an $n \times 1$ column vector of ones. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the eigenvalues of S in descending order, and let $u_1, u_2, \ldots, u_p$ be the corresponding eigenvectors. The first principal component Y_1 is defined as a linear combination of X_j's such that it has the largest variance under the constraint that the coefficient vector has unit norm. It turns out that the coefficient vector, which is called the loading of Y_1, is estimated by u_1, the eigenvector of S corresponding to the largest eigenvalue λ_1. The second principal component Y_2 is the linear combination of X_j's with the second largest variance under the unit norm constraint uncorrelated with Y_1, and the loading of Y_2 is estimated by u_2. In general, the kth principal component is computed as

$$\widehat{Y_k} = u_k^T X, \quad k = 1, \ldots, q. \qquad (5)$$

Subsequent analyses are usually performed based on these q uncorrelated principal components (as opposed to the original p variables), whose observed values are given by the principal component score matrix

$$Z = U^T X. \qquad (6)$$

here, $U = [u_1, u_2, \ldots, u_q]$ is the $p \times q$ loading matrix. To mitigate the effect of scaling, it is a common practice to standardize the variances before performing a PCA. In such a situation, the sample correlation matrix ρ

$$\rho = \sqrt{D}^{-1} S \sqrt{D}^{-1} \qquad (7)$$

is used in replacement of the sample covariance matrix S, where D is the diagonal matrix of the diagonal entries of S, i.e.

$$D = diag\{S(1,1), S(2,2), \ldots, S(p,p)\}. \qquad (8)$$

It is equivalent to using the sample covariance matrix when the variances of all variables are standardized to be 1.

One major drawback of PCA is that each principle component is a linear combination of all of the predictor variables, which often makes the results difficult to interpret. To address this problem, Zou et al. (2006) proposed the Sparse Principal Components Analysis (SPCA) as an alternative shrink some of the coefficients to 0 by producing a sparse estimate of the loading matrix via the technique of penalized regression. Technically, this is done by expressing PCA as a regression problem with a quadratic penalty, which essentially forms the ridge regression:

$$\left\{ \widehat{A}, \widehat{B} \right\} = \arg \min_{A,B} \left\{ \left\| X - XBA^T \right\|^2 + \lambda \sum_{k=1}^{q} \|\beta_k\|^2 \right\},$$
$$\text{subject to } A^T A = I. \qquad (9)$$

here, $A = [\alpha_1, \alpha_2, \ldots, \alpha_q]$ and $B = [\beta_1, \beta_2, \ldots, \beta_q]$ are two $p \times q$ coefficient matrices, and $\|\cdot\|$ denotes the Euclidean norm. The normalized vector of β_k gives the approximation to the loadings of the kth principal component, i.e.,

$$\widehat{u_k} = \frac{\widehat{\beta}_k}{\left\| \widehat{\beta}_k \right\|}, \quad k = 1, \ldots, q. \qquad (10)$$

Then, an L_1 or Lasso penalty (Tibshirani 1996) is added to the optimization criterion to induce sparsity, i.e., shrink some of the coefficients to 0. Thus, the sparse PCA is formulated as

$$\left\{ \widehat{A}, \widehat{B} \right\} = \arg \min_{A,B} \left\{ \left\| X - XBA^T \right\|^2 + \lambda \sum_{k=1}^{q} \|\beta_k\|^2 \right.$$
$$\left. + \sum_{k=1}^{q} \lambda_k \|\beta_k\|_1 \right\}, \quad A^T A = I, \qquad (11)$$

where $\|\cdot\|_1$ denotes the L_1 norm, i.e., summation of the absolute values of the elements. The constants λ and λ_k, $k = 1, \ldots, q$ are tuning parameters, of which λ_k's are associated with the Lasso penalty and control the amount of shrinkage, i.e., how many coefficients are shrunk to 0. Smaller values of λ_k induce more 0's in $\widehat{\beta}_k$. Fitting of SPCA can be carried out in the software R using the package *elasticnet* (see Zou and Hastie 2005). As a remark, due to the induced sparsity in SPCA, the resulting loadings deviate from being orthogonal, and consequently, the corresponding sparse PCs are no longer guaranteed to be uncorrelated (Zou et al. 2006). However, engineers will likely willingly trade off PCs being uncorrelated for improvements in simplicity and predictive accuracy.

3 Principal Component Analysis Application

The unbonded tendon data are split into internally reinforced (internal) and externally reinforced (external) subsets each possessing 17 predictor variables and the response variable, Δf_{ps}. The 15 predictors contained in the database are included in the analysis as well as two additional variables, v_{ACI} and v_{AASHTO}, which are the variable parts to the ACI and AASHTO prediction equations (ACI 2008; AASHTO 2010). These are included in the analysis in an attempt to build upon any already discovered explained variation in the data. The ACI

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 5 of 18

variable part is well known for being inaccurate, whereas the AASHTO variable part is highly phenomenological and some variation is included in many design codes around the world.

$$v_{\text{ACI}} = \frac{f_c'}{\mu \rho_{ps}} \quad (12)$$

$$v_{\text{AASHTO}} = \left(\frac{d_{ps} - c}{L} \right) \left(1 + \frac{N}{2} \right) \quad (13)$$

The internal data has 182 observations, and the external data has 71. The variable names and type, as they are typically defined for statistical analyses (Nowak and Collins 2012), are found in Table 1. The only Categorical data type is the LT variable, which is 1, 2 or 3 for single point loading, third point loading or uniform loading. Both data subsets exhibit multicollinearity among predictors in their respective sample covariance matrices suggesting repetition of information. Due to the wide variation in scale of the different variables, the correlation matrix is chosen over the covariance matrix for the PCA.

Because variable selection is not handled by the LASSO operator as it is with SPCA, multiple approaches were used in selecting important variables for the PCA. The initial approach consisted of merely assuming that all 17 variables were important. An Eigen-decomposition was applied to the correlation matrix using Eq. (7) to calculate the PCs. Figure 1 consists of scree-plots showing the proportion of variation and cumulative proportion of variation explained by each principal component for their respective data subset.

An 'elbow', or change in slope between PCs (Jolliffe 2002), in the scree-plot suggest good choices for the number of PCs that express the most information while keeping the model simple, e.g. the elbow seen at three PCs in Fig. 1a. However, five principal components are selected for both the internal and external data as a means to compare models, and since five PCs capture a majority of proportion of variation in the data, while keeping the models relatively simple. The cumulative proportion of variation for 5 PC's is 0.80 for the internal tendons, and 0.84 for the external tendons.

From the five selected principal components, linear combinations of the 17 variables can now be expressed as five new uncorrelated variables. Then with tenfold cross validation, linear models are then fit to the data using the five new variables. As criterion of how well the models are fitting the

Table 1 Variable names and descriptions for the statistical analysis.

Variable name	Notation	Type
Variable part of the ACI prediction equation	v_{ACI}	Continuous
Variable part of the AASHTO prediction equation	v_{AASHTO}	Continuous
Loading type	LT	Categorical
Total span length	L	Continuous
Beam height	h	Continuous
Beam width	b	Continuous
Depth to prestressing reinforcement	d_{ps}	Continuous
Area of prestressing reinforcement	A_{ps}	Continuous
Ultimate tendon strength	f_{pu}	Continuous
Concrete strength	f_c'	Continuous
Area of mild reinforcing steel on tension face	A_s	Continuous
Yield strength of mild reinforcing steel	f_y	Continuous
Depth to tension mild reinforcing steel from compression face	d_s	Continuous
Area of mild reinforcing steel on compression face	A_s'	Continuous
Depth to compression mild reinforcing steel from compression face	d_s'	Continuous
Effective stress in the prestressing reinforcement at time of testing	f_{pe}	Continuous
Modulus of elasticity of the prestressing reinforcement	E_{ps}	Continuous
Stress increase in unbonded tendons	Δf_{ps}	Continuous

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20
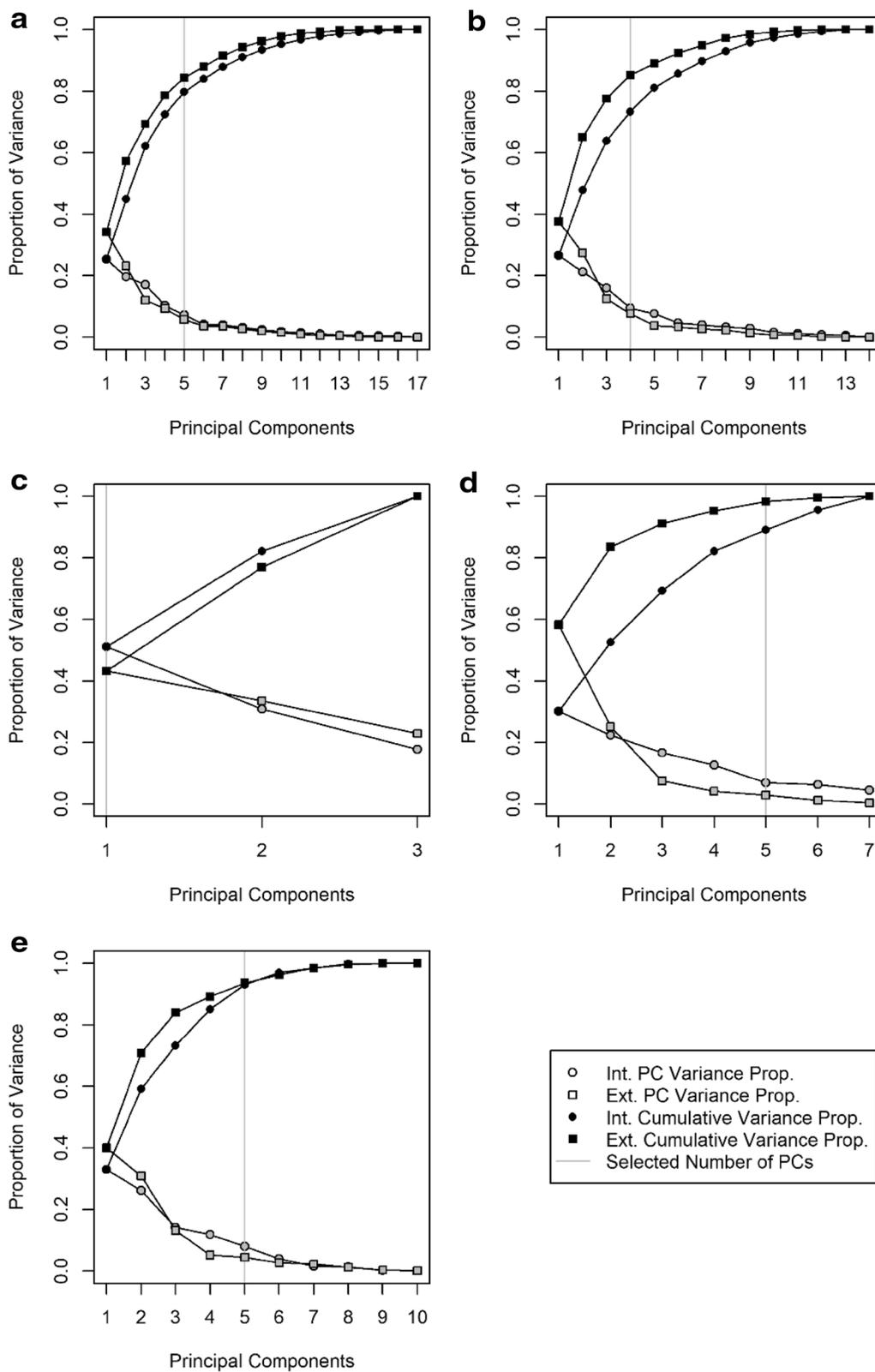
Page 6 of 18

Fig. 1 Individual and cumulative explained proportion of variation for each Principal Component for **a** all variables and the **b** Continuous, **c** Categorical, **d** Self-Selected, and **e** Correlation Cutoff variable subsets.

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 7 of 18

data, the coefficient of determination (R^2), adjusted (R_a^2), average ratio of measured vs. predicted responses λ, root mean squared error (RMSE), and the mean absolute error (MAE) are calculated for each model (Kutner et al. 2004). R^2 is the ratio of the explained variation made by the model over the total variation in the data, defined as:

$$R^2 = \frac{\sum_{i=1}^n \left(\widehat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^n \left(y_i - \bar{y}\right)^2} \tag{14}$$

where $\widehat{y}_i$ is the ith predicted Δf_{ps}, y_i is the ith Δf_{ps}, and $\bar{y}$ is the sample average of Δf_{ps}. Adjusted R^2 is similar to R^2 but it is penalized for more complicated models that involve more predictors. It is calculated as follows:

$$R_a^2 = 1 - \frac{\sum_{i=1}^n \left(y_i - \widehat{y}_i\right)^2 \big/ (n - p - 1)}{\sum_{i=1}^n \left(y_i - \bar{y}\right)^2 \big/ (n - 1)} \tag{15}$$

where p is the number of predictor variables used in the model plus one. λ is calculated as the mean of all of the ratios of Δf_{ps} values and their corresponding linear model predicted values, $\widehat{\Delta f_{ps}}$, i.e.

$$\lambda = \frac{1}{n} \cdot \sum_{i=1}^n \frac{\Delta f_{ps_i}}{\widehat{\Delta f_{ps_i}}} \tag{16}$$

A visualization related to λ is seen in Fig. 2 as plots of the Δf_{ps} values against the linear model's predicted values, $\widehat{\Delta f_{ps}}$. While R^2, R_a^2, and λ values closer to 1 indicate better fitting models, RMSE and MAE are best minimized. RMSE gives greater emphasis on extreme values, whereas MAE treats all data points with equal importance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \left(y_i - \widehat{y}_i\right)^2}{n}} \tag{17}$$

$$MAE = \frac{\sum_{i=1}^n \left|y_i - \widehat{y}_i\right|}{n} \tag{18}$$

The initial models, referred to as PCA-All-Int and PCA-All-Ext, have R^2, R_a^2, λ, RMSE, and MAE values of 0.43, 0.43, 1.00, 110.56, 87.38, and 0.64, 0.63, 1.01, 166.91, 128.16 respectfully (as listed in the first row of Table 2).

A second approach was attempted by handling the Continuous and Categorical variables separately. While all of the variables are continuous except *LT*, the variables E_{ps} and d_s' behaved as Categorical in the data and are treated as such (see Table 1). A separate PCA was computed for the 14 Continuous variables and the 3 Categorical variables within each data set. In order to keep the same number of

overall PC's in the final models, four PC's are chosen for the Continuous variables, and one is chosen for the Categorical variables as seen in Fig. 1b and c. The results were then combined into linear models called PCA-ContCate-Int and PCA-ContCate-Ext, and their criteria are $R^2 = 0.36$, 0.66, $R_a^2 = 0.35$, 0.65, $\lambda = 1.02$, 1.02, RMSE = 117.38, 162.08, and MAE = 95.88, 124.60 as shown in Table 2. Plots for measured vs. predicted Δf_{ps} are also included in Fig. 2b.

Again, the four previously calculated PCA linear models suffer due to the fact that each principal component is a linear combination of all predictor variables, which is not ideal for structural design. Variable selection restricting only important variables into the PCA would allow for simpler linear models with possibly better predictive power. Two additional subsets of the original variables are considered and a model selection technique was employed and compared to the initial analysis. The first set of selected important variables is decided through professional suggestion. The authors call this set the "Self-Selected" set. The second set, called the "Correlation Cutoff" set, was selected by a test of minimum linear correlation with Δf_{ps}. Subsequent PCA linear models are then computed for all possible combinations of PC's as predictors, statistical significance is assessed on the coefficients via t-tests, and the final models chosen are those which achieve the highest R_a^2.

The Self-Selected important variables are L, h, A_{ps}, f_c', A_s, A_s', f_{pe}, and Δf_{ps} based on the literature and experience. After a PCA is applied to these variables the data is reduced from only seven predictor variables to five. While this is not a gain of much more simplicity to the models, the correlation between the predictors is removed. The scree plots in Fig. 1d again show that most of the information is expressed in the first five PC's chosen.

While there is a noticeable gain in cumulative proportion of variance explained by these 5 PC's in both data sets (0.89 for the internal data, and 0.98 for the external data), the final models, called PCA-SS-Int and PCA-SS-Ext, do not make similar gains in modeling the data, as seen by their respective $R^2 = 0.26$, 0.49, $R_a^2 = 0.25$, 0.48, $\lambda = 1.04$, 1.04, RMSE = 126.34, 198.51, and MAE = 103.43, 160.44 values. A lack of fit to the data is seen in Fig. 2c by the models tendency to over predict for lower values of Δf_{ps} and to under predict for higher values of Δf_{ps}.

This process is repeated for the Correlation Cutoff set as well. However, these variables were selected by first examining their respective linear correlations with Δf_{ps}. While simply selecting predictors with a significant amount of correlation with the response does not consider collinearity among predictors, the subsequent PCA handles this by producing uncorrelated PC's, likewise for SPCA. A Pearson's product-moment correlation test is applied with a level of significance set at 0.05. Table 3 contains the correlations and p-values for both internal and external data.
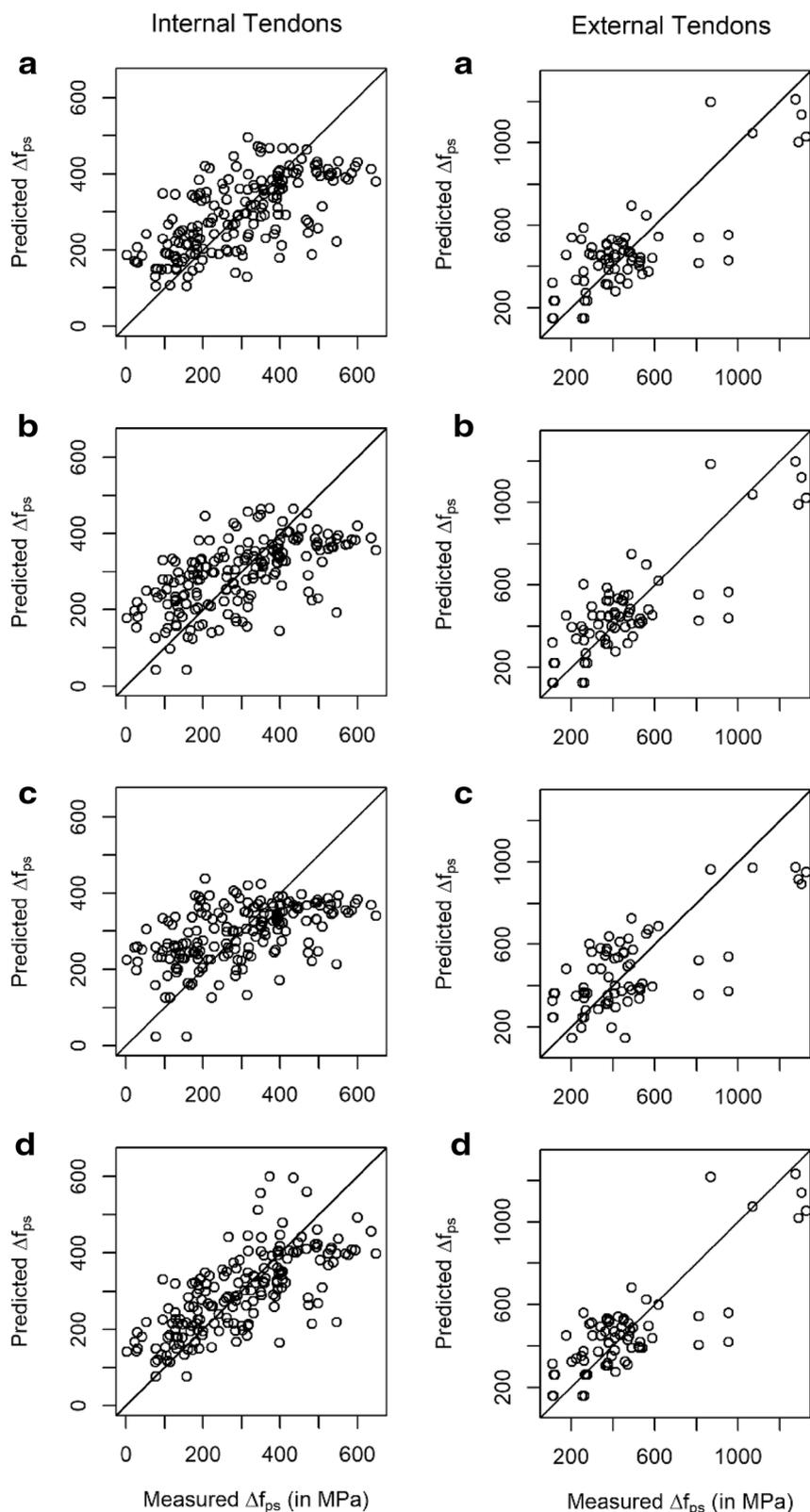
Fig. 2 Measured Δf_{ps} vs. predicted Δf_{ps} (in MPa) from the PCA models using **a** all variables, and the **b** combined Continuous and Categorical, **c** Self-Selected, and **d** Correlation Cutoff variable subsets.

McKinney *et al. Int J Concr Struct Mater* (2019) 13:20

Page 9 of 18

Table 2 PCA models' R^2, R_a^2, λ, RMSE, and MAE values.

Variables	Internal data					External data				
	R^2	R_a^2	λ	RMSE	MAE	R^2	R_a^2	λ	RMSE	MAE
All variables	0.43	0.43	1.00	110.56	87.38	0.64	0.63	1.01	166.91	128.16
Cont. and Cate.	0.36	0.35	1.02	117.38	95.88	0.66	0.65	1.02	162.08	124.60
Self-Selected	0.26	0.25	1.04	126.34	103.43	0.49	0.48	1.04	198.51	160.44
Corr. Cutoff	0.52	0.50	0.99	102.36	81.09	0.67	0.66	1.01	160.93	123.49

Table 3 Correlation Cutoff important variables for the internal and external data.

Variable	Internal data			External data		
	Correlation with Δf_{ps}	p-value	Important	Correlation with Δf_{ps}	p-value	Important
v_{ACI}	0.42	< 0.001	TRUE	0.77	< 0.001	TRUE
v_{AASHTO}	0.48	< 0.001	TRUE	0.57	< 0.001	TRUE
LT	0.51	< 0.001	TRUE	0.25	0.04	TRUE
L	− 0.06	0.45	FALSE	− 0.27	0.02	TRUE
h	0.28	< 0.001	TRUE	− 0.57	< 0.001	TRUE
b	− 0.17	0.02	TRUE	− 0.03	0.78	FALSE
d_{ps}	0.29	< 0.001	TRUE	0.24	0.048	TRUE
A_{ps}	− 0.51	< 0.001	TRUE	− 0.49	< 0.001	TRUE
f_{pu}	0.33	< 0.001	TRUE	− 0.22	0.07	FALSE
f_c'	− 0.01	0.89	FALSE	0.52	< 0.001	TRUE
A_s	0.01	0.87	FALSE	− 0.53	< 0.001	TRUE
f_y	0.22	0.002	TRUE	− 0.23	0.054	FALSE
d_s	− 0.04	0.60	FALSE	− 0.14	0.26	FALSE
A_s'	− 0.05	0.55	FALSE	− 0.35	0.003	TRUE
d_s'	− 0.08	0.29	FALSE	− 0.14	0.25	FALSE
f_{pe}	0.06	0.46	FALSE	0.35	0.003	TRUE
E_{ps}	− 0.08	0.27	FALSE	0.09	0.44	FALSE
Δf_{ps}	1.00	0.00	TRUE	1.00	0.00	TRUE

Interestingly, Table 3 indicates that for internally bonded tendons, the length is not important, which Mojtahedi and Gamble (1978), among others, indicate is important. Concrete strength is not considered important, although it shows up in the ACI code, and several, and the current ACI code. The variables b, d_{ps} and A_{ps} are considered important and are also considered in the ACI code as the prestressing reinforcing ratio (ρ_{ps}). Interestingly, f_y is considered important although it is not included in any known prediction model, and conversely, A_s is not considered important contradicting several experimental studies.

Additionally, Table 3 indicates that there are considerable differences in the significance of many variables. Most notably is the 0.77 correlation between v_{ACI} and Δf_{ps}, as compared to the 0.42 correlation for the internally bonded tendons. There is agreement on several variables, for instance, the loading type, depth of section (h and d_{ps}) and A_{ps} are considered important while d_s, d_s' and E_{ps} are not considered important in both sets. However, the remaining variables are in contention. For instance, length is considered important in the external dataset as is concrete strength, f_{pe} and A_s, but not f_y. Interestingly, A_s' is considered important in the external dataset. Furthermore, h, f_{pu}, A_s and f_y were found to have opposite effect (see difference in signs in Table 3) on the behaviour, indicating either very different phenomenological effects or shortcomings in the dataset.

The dataset itself is made of all of the available experimental data, but the dataset is also shaped by the experimental needs. Externally reinforced members tend to be larger bridge girders with higher reinforcing ratios and, often, A_s'. The make-up of the externally reinforced dataset reflects this and contains more beam-like members (higher d_{ps}, h, A_{ps}, A_s' etc.), many

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 10 of 18

Table 4 SPCA models' R^2, R_a^2, λ, RMSE, and MAE values.

Variables	Internal data					External data				
	R^2	R_a^2	λ	RMSE	MAE	R^2	R_a^2	λ	RMSE	MAE
All variables	0.46	0.46	0.99	107.93	85.83	*0.70*	*0.69*	*0.99*	*152.79*	*110.93*
Cont. and Cate.	0.44	0.43	1.02	109.72	88.58	0.70	0.69	0.99	152.79	110.93
Self-Selected	0.31	0.30	0.95	121.96	101.16	0.54	0.52	1.05	189.29	151.38
Corr. Cutoff	*0.54*	*0.53*	*1.03*	*99.53*	*78.04*	0.68	0.68	1.00	156.88	113.58

of them simulating bridge girders. The internally reinforced dataset is made up of many more slab like members that do not contain compression steel and are smaller, some of which are scaled (Burns et al. 1978; Six 2015). Regardless, one should be aware that the dataset, while the largest available, does contain limited numbers and limited variations for many variables. From this analysis, it is unclear if the difference in variable importance is due to the dataset or phenomenological differences. The analysis does seem to dispute the use of the same equation for internal and external members (like ACI and AASHTO) and indicates that predictions that somehow account for the difference may be better (like Maguire et al. 2017; Harajli 2011).

If a variable exhibited significant correlation (p-value less than 0.05) with Δf_{ps} it was kept for subsequent analysis. The correlation cutoff variables for the internal data are ν_{ACI}, ν_{AASHTO}, LT, h, b, d_{ps}, A_{ps}, f_{pu}, and f_y, and the correlation cutoff variables for the external data are ν_{ACI}, ν_{AASHTO}, LT, L, h, d_{ps}, A_{ps}, f_{pu}, f_c', A_s, f_y, A_s', and f_{pe}. The scree plots in Fig. 1e show a cumulative proportion of variation for the internal data is 0.93, and 0.94 for the external data.

By using Pearson's product-moment correlation test to remove variables that exhibit low correlations with Δf_{ps}, applying a PCA on the remaining predictors, and then using model selection the linear models tend to model the data better as seen in their respective $R^2 = 0.52$, 0.67, $R_a^2 = 0.50$, 0.66, $\lambda = 0.99$, 1.01, RMSE $= 102.36$, 160.93, and MAE $= 81.09$, 123.49 values (see Table 2). Due to the PCA predictions resulting in very long and cumbersome equations, even when simplified (as they load all 15 of the explanatory variables), they are not presented here. However, they can be constructed using the PC loadings presented above in the PCA section.

4 Sparse Principal Components Application
SPCA was applied to both internal and external data sets on all of the subsets of variables producing eight additional linear models called SPCA-All-Int, SPCA-All-Ext, SPCA-ContCate-Int, SPCA-ContCate-Ext, SPCA-SS-Int, SPCA-SS-Ext, SPCA-CC-Int, and SPCA-CC-Ext (see Table 4). In Table 4, the italic numbers indicate the

selected models for the respective data datasets. In all of these cases, a decision must be made about how much sparsity is desirable. Again, sparsity in the Principal Components is the reduction of some of the coefficients, or loadings, for the linear combinations of the predictor variables to zero.

In applying SPCA to all of the variables, Fig. 3 reveals optimal choices for the number of sparse coefficients per PC by maximization of R_a^2. Note that the variation in the external subset is being explained significantly better by the data than the internal subset as seen by the consistently higher R_a^2 (Fig. 3a, b, d, e). However, Fig. 3c shows little variation in data being explained by the variables that were treated as Categorical variables. More specifically, Fig. 3a suggests 2 and 1 non-zero loadings (for each SPC) for the internal and external data respectively. The sparse loadings for all of the SPCA models are represented by heat maps found in Fig. 4. The two initial SPCA models achieve R^2 values of 0.46, 0.70, maximum R_a^2 values of 0.46, 0.69, λ values of 0.99, 0.99, RMSE values of 107.93, 152.79, and MAE values of 85.83, and 110.93 (see Table 4). Lastly, as in the PCA comparisons are made between measured and predicted Δf_{ps} as seen in Fig. 5a.

Due to the PCA predictions resulting in very long and cumbersome equations, even when simplified (as they load all 17 of the explanatory variables), they are not presented here. However, their SPCA versions are produced and explicitly listed in the following section.

Furthermore, the following results of applying SPCA to the Continuous and Categorical, Self-Selected, and Correlation Cutoff subsets are similarly recorded and compared to the previous analysis. For each the number of non-zero loadings per SPC are calculated (see Fig. 3), model selection is evaluated, heat maps of the sparse loadings are produced (see Fig. 4), and the R^2, R_a^2, λ, RMSE, and MAE values are recorded (see Table 4). These linear models are listed explicitly with their respective linear combinations for each SPC. While the models are shown here with their respective PCs, with some algebraic manipulation alternative versions of the final suggested models are presented in the following section. It should be noted when SPCA is applied to the Correlation Cutoff variables that ten variables were retained for the external data, while only nine
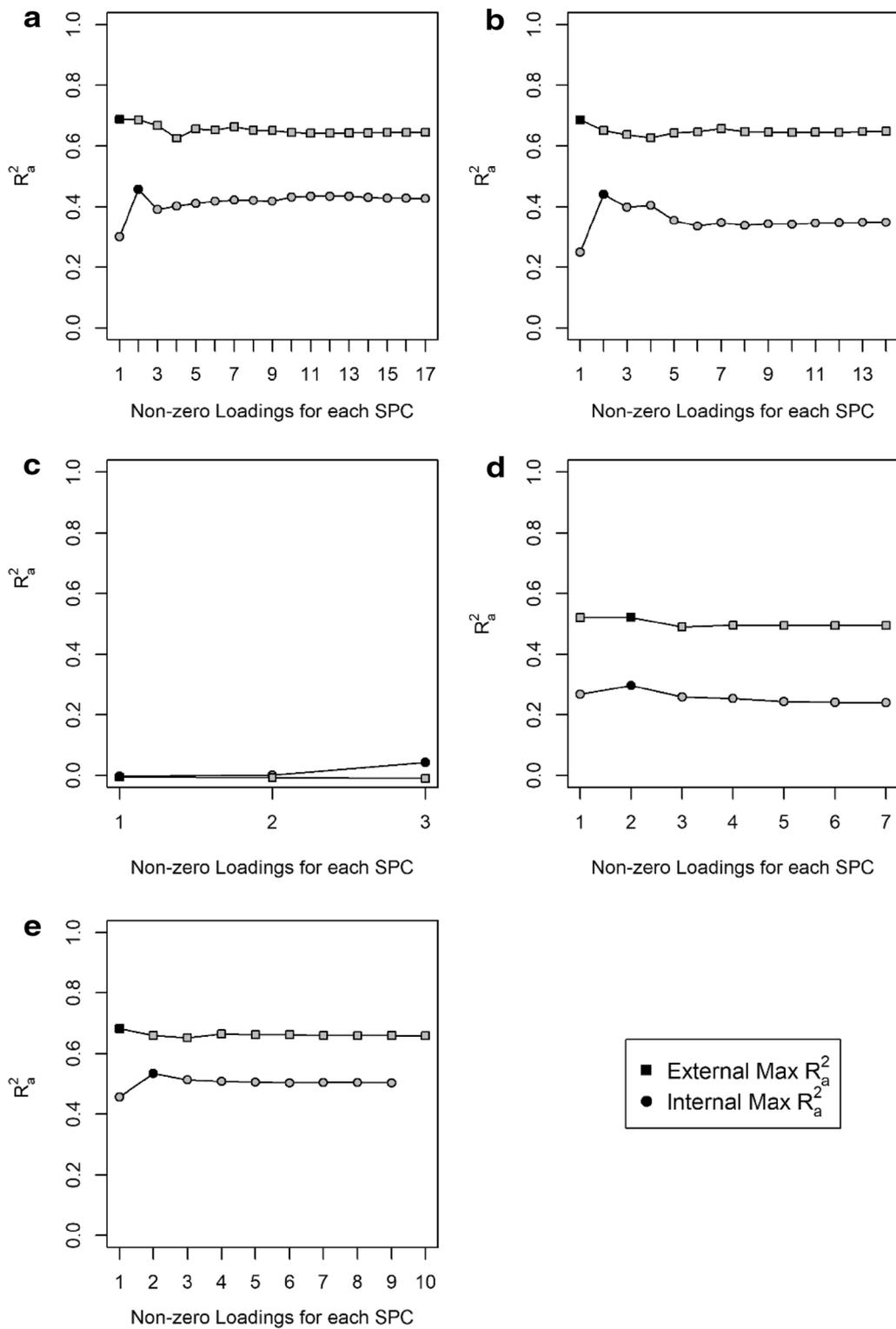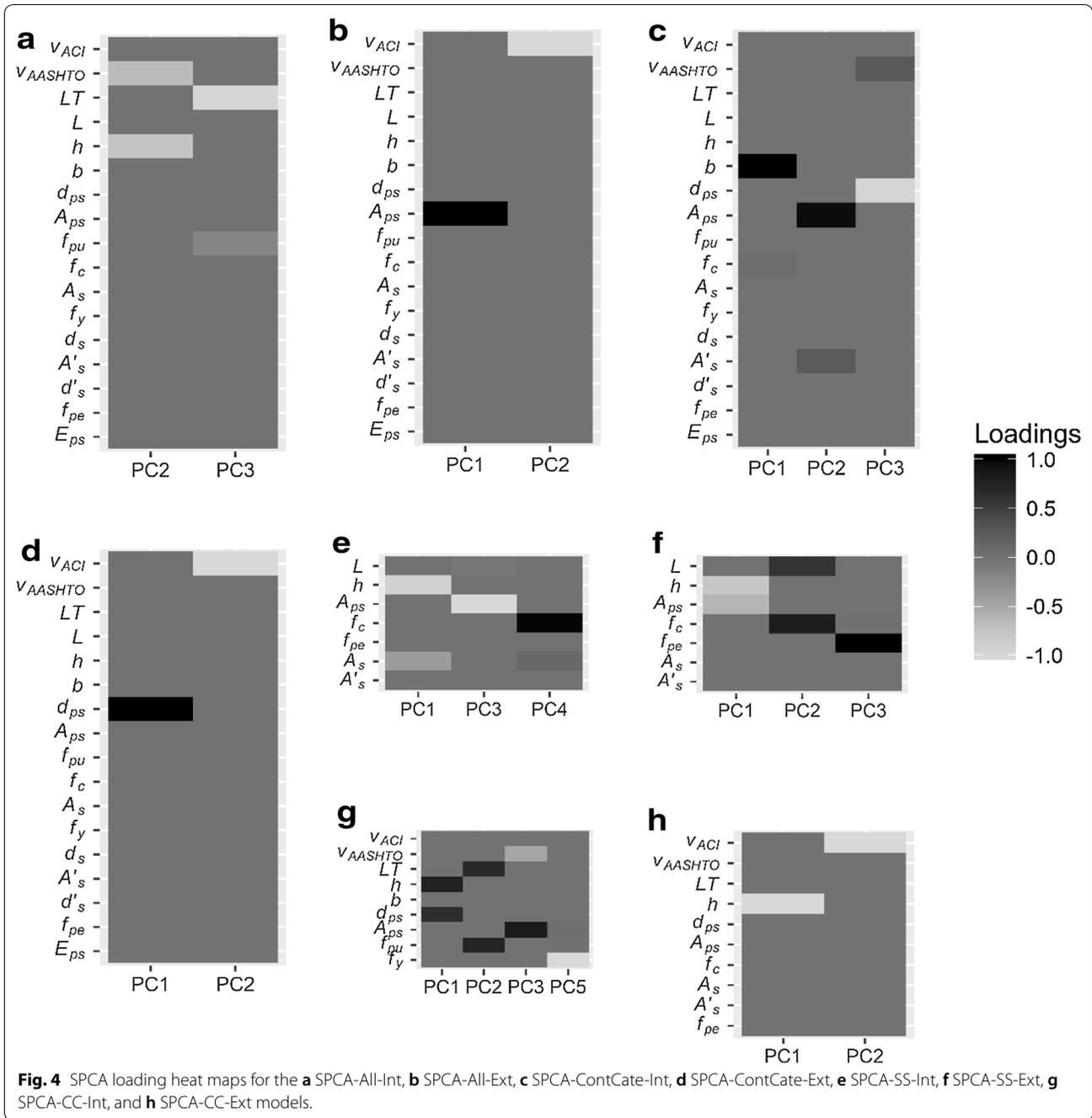
Fig. 3 Plot of SPCA R_a^2 against number of non-zero loadings used per PC for **a** all variables and the **b** Continuous, **c** Categorical, **d** Self-Selected, and **e** Correlation Cutoff variable subsets. The maximum R_a^2 for each model is highlighted with a filled marker.

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 12 of 18

Fig. 4 SPCA loading heat maps for the **a** SPCA-All-Int, **b** SPCA-All-Ext, **c** SPCA-ContCate-Int, **d** SPCA-ContCate-Ext, **e** SPCA-SS-Int, **f** SPCA-SS-Ext, **g** SPCA-CC-Int, and **h** SPCA-CC-Ext models.

were kept for the internal data. Hence, the number of non-zero loadings for each SPC for the internal data extends to only nine in Fig. 3e.

4.1 Prediction Equation for Internal all Variables SPCA (SPCA-All-Int)

$$\widehat{\Delta f_{ps}} = 295.06 - 54.22PC_{1,2} - 69.91PC_{1,3}$$
$$PC_{1,2} = -0.66v_{AASHTO} - 0.75h$$
$$PC_{1,3} = -0.98LT - 0.18f_{pu}$$

4.2 Prediction Equation for External all Variables SPCA (SPCA-All-Ext)

$$\widehat{\Delta f_{ps}} = 470.19 - 95.79PC_{2,1} - 195.77PC_{2,2}$$
$$PC_{2,1} = 1A_{ps}$$
$$PC_{2,2} = -1v_{ACI}$$

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 13 of 18

Internal Tendons

External Tendons
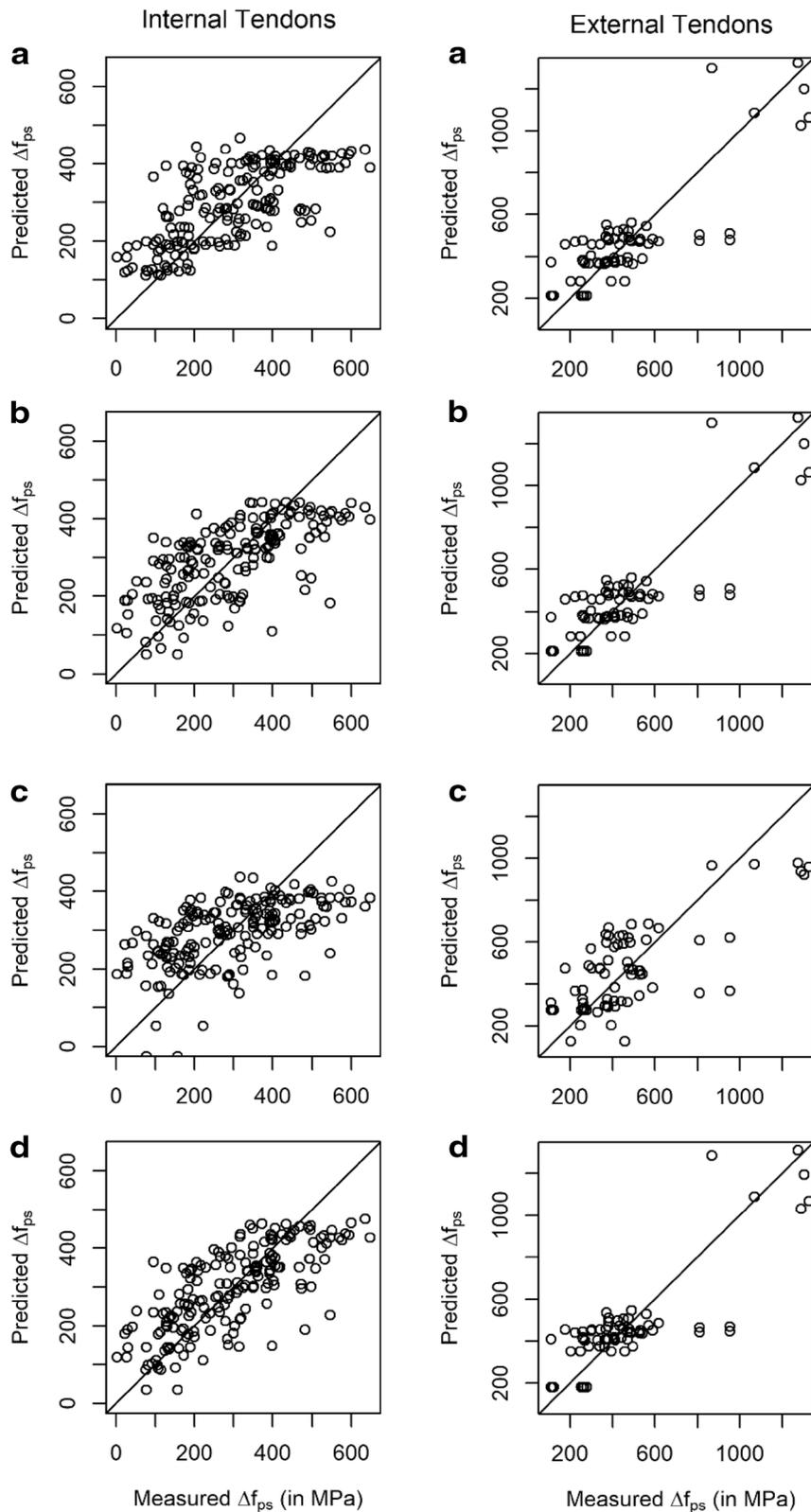
Fig. 5 Measured Δf_{ps} vs. predicted Δf_{ps} (in MPa) from the SPCA models using **a** all variables, and the **b** combined Continuous and Categorical, **c** Self-Selected, and **d** Correlation Cutoff variable subsets.

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 14 of 18

4.3 Prediction Equation for Internal Continuous and Categorical SPCA (SPCA-ContCate-Int)

$$\widehat{\Delta f_{ps}} = 295.06 + 30.45PC_{3,1} + 41.68PC_{3,2} + 69.51PC_{3,3}$$
$$PC_{3,1} = 1d_{ps} + 0.06A_s$$
$$PC_{3,2} = 0.97f_{pu} + 0.24f_{pe}$$
$$PC_{3,3} = 0.27v_{AASHTO} - 0.96A_{ps}$$

4.4 Prediction Equation for External Continuous and Categorical SPCA (SPCA-ContCate-Ext)

$$\widehat{\Delta f_{ps}} = 470.19 - 95.79PC_{4,1} - 195.77PC_{4,2}$$
$$PC_{4,1} = 1A_{ps}$$
$$PC_{4,2} = -1v_{ACI}$$

4.5 Prediction Equation for Internal Self-Selected SPCA (SPCA-SS-Int)

$$\widehat{\Delta f_{ps}} = 295.06 - 29.26PC_{5,1} + 74.59PC_{5,3} - 20.84PC_{5,4}$$
$$PC_{5,1} = -0.92h - 0.39A_s$$
$$PC_{5,3} = -0.05L - 1A_{ps}$$
$$PC_{5,4} = 0.99f_c' + 0.11A_s$$

4.6 Prediction Equation for External Self-Selected SPCA (SPCA-SS-Ext)

$$\widehat{\Delta f_{ps}} = 470.19 + 135.20PC_{6,1} + 60.81PC_{6,2} + 89.02PC_{6,3}$$
$$PC_{6,1} = -0.79h - 0.61A_{ps}$$
$$PC_{6,2} = 0.57L + 0.82f_c'$$
$$PC_{6,3} = 0.01f_c' + 1f_{pe}$$

4.7 Prediction Equation for Internal Correlation Cutoff SPCA (SPCA-CC-Int)

$$\widehat{\Delta f_{ps}} = 295.06 + 22.44PC_{7,1} + 53.13PC_{7,2}$$
$$\qquad\qquad - 68.24PC_{7,3} + 26.69PC_{7,5}$$
$$PC_{7,1} = 0.78h + 0.63d_{ps}$$
$$PC_{7,2} = 0.68LT + 0.73f_{pu}$$
$$PC_{7,3} = -0.49v_{AASHTO} + 0.87A_{ps}$$
$$PC_{7,5} = 0.04A_{ps} - 1f_y$$

4.8 Prediction Equation for External Correlation Cutoff SPCA (SPCA-CC-Ext)

$$\widehat{\Delta f_{ps}} = 470.19 + 92.94PC_{8,1} - 182.04PC_{8,2}$$
$$PC_{8,1} = -1h$$
$$PC_{8,2} = -1v_{ACI}$$

5 Discussion

From Table 2, the R^2, R_a^2, λ, RMSE, and MAE values for the initial models involving all 17 variables are 0.43, 0.43, 1.00, 110.56, 87.38 for the internal data, and 0.64, 0.63, 1.01, 166.91, and 128.16 for the external data. Comparatively, these initial PCA linear models improve significantly over previous methods (Maguire et al. 2017), where $\lambda = 1.85$ and $R^2 = 0.16$ for the AASHTO, being the most accurate and precise of the available American codified methods, as well as $\lambda = 1.34$ and $R_a^2 = 0.27$ for the previously proposed method modification to the AASHTO prediction.

Also, notice the linear equations for the initial SPCA models are much simpler when compared to their corresponding PCA models since each of the five PCs are required to have 17 loadings, whereas each SPC only produce 1 or 2 (Fig. 3). This gain in simplicity is paired with gains in R^2, and R_a^2, λ values close to one, and smaller RMSE and MAE values (compare the first row in Table 2 to the first row in Table 4).

The PCA models handling the Continuous and Categorical variables separately did not perform better than the initial model involving all 17 variables for the internal tendons, but did for the external (Table 2). This may be due to the unaccounted covariances between the Continuous and Categorical variables along with the significant contribution of explained variability by v_{ACI} in the external data (see the first row of Table 3). A similar behavior is seen in the SPCA models (compare first and second rows of Table 4). Note that after model selection the final SPCA models for both all variables and the Continuous and Categorical subsets resulted in identical coefficients. This suggests that handling the variables separately does not differ from handling the variables collectively when applying SPCA with model selection to the external data.

Notice only one loading for each PC in is suggested for the external models using all of the variables, the Continuous subset, and the correlation cutoff subset to maximize R_a^2. This suggests that a linear model is sufficient in modeling the variation in the stress increase Δf_{ps} for these cases.

However, while the PCA and SPCA models for the Self-Selected variables did improve over the AASHTO and proposed modified AASHTO predictions, they performed poorer than the initial PCA and SPCA on all of the variables (compare first and third rows in Table 4). This suggests that variables that engineers and the literature commonly associate with Δf_{ps}, may not be as impactful as thought, underscoring the necessity for further experimental and phenomenological study.

Additionally, it should be noted that the predicted stress increase, $\widehat{\Delta f_{ps}}$, is consistently under predicting for higher measured values of Δf_{ps} in the internal data

McKinney *et al. Int J Concr Struct Mater*   (2019) 13:20

Page 15 of 18

(Figs. 2, 5). Some of this is also exhibited in the external data though not as strongly. This suggests that an underlying non-linear relationship may be present in the data, and suggests further analysis possibly involving more advanced models.

Most notably, the R^2, R_a^2, λ, RMSE and MAE values are 0.54, 0.53, 1.03, 99.53, and 78.04 for the internal correlation cutoff SPCA model, and 0.70, 0.69, 0.99, 152.79, and 110.93 for the external model involving all of the variables (see italic values in Table 4). Notice that while the difference in increased R^2 and R_a^2 for the SPCA-CC-Int model is 0.08 and 0.07, a noticeable amount, the SPCA-CC-Ext model does not improve over the initial SPCA for all external variables (compare first and forth rows of Table 4). Furthermore, after the model selection process only two terms remain in both the SPCA-All-Ext and SPCA-All-Int models (Fig. 4b, h). Hence, while not as reduced as the external model, the most predictive accuracy for the internal data is in the suggested SPCA-CC-Int model. Whereas for the external data, the SPCA-All-Ext model is recommended, achieving both the highest predictive accuracy while producing a simplistic design. Many of the under and over predictions made by the ACI and AASHTO models are handled better by the SPCA-CC-Int and SPCA-All-Ext models (compare Fig. 5a External to Fig. 6a, b External. Also, compare Fig. 5d Internal to Fig. 6a, b Internal).

It should be noted that while the SPCA-All-Ext and SPCA-CC-Ext models both have two variables, with ν_{ACI} being in common, the other two variables (A_{ps} in SPCA-All-Ext and h in SPCA-CC-Ext) are not the same (Fig. 4). The reasoning for the difference is likely the fact that both A_{ps} and h are highly correlated (specifically 0.93 correlation), and similar information is being expressed in each model through collinearity (Table 5).

5.1 Simplified Prediction Equation for Internal Data on the Correlation Cutoff Subset (SPCA-CC-Int)

$$\widehat{\Delta f_{ps}} = 295.06 + 17.45h + 14.11d_{ps} + 36.12LT + 38.96f_{pu} + 33.57\nu_{AASHTO} - 58.40A_{ps} - 26.68f_y$$

5.2 Simplified Prediction Equation for External data on all of the Variables (SPCA-All-Ext)

$$\widehat{\Delta f_{ps}} = 470.19 - 95.79A_{ps} + 195.77\nu_{ACI}$$

Interestingly, ν_{ACI} was found by the SPCA technique to be beneficial to the external prediction equations, whereas the highly phenomenological ν_{AASHTO}, which takes into account hinging location, was found to be important to the internal model. This is not surprising since Maguire et al. (2017) found a calibrated version of the internal equation was most accurate, and the ν_{ACI} equation, while not intended when developed, predicts external members better than most other methods. Interestingly, the final SPCA prediction for external tendons relies only on the ν_{ACI} and A_{ps} variables, of which the latter was often found as important by experimental studies.

Conversely, even after efforts to simplify through model selection, the final SPCA prediction for internal tendons contains seven variables including LT, which lends some phenomenological influence. Furthermore, ν_{AASHTO} is also present, which lends significant phenomenological influence. However, the other variables are several of those disputed by the literature.

6 Summary and Conclusions

The PCA and SPCA linear modeling is applied to study the relationship between Δf_{ps} and a collection of variables. The method consists of two consecutive steps: creation of uncorrelated (sparse) principal components and linear regression with the principal components. Due to the uncorrelatedness of the PC's, variable selection for the linear regression is simple and straightforward. In fact, the PCA/SPCA is an important alternative to perform model selection, compared to the celebrated penalized regression, which requires intensive tuning to achieve optimal performances. Furthermore, the PC's also provide an insightful understanding of the relationship between the outcome and the original variables.

The data in Maguire et al. (2017) were separated into two data sets determined by internal or external tendons. Stochastic linear models based on PCA and SPCA were constructed as prediction equations for Δf_{ps}. Eight resulting linear models involved all the available explanatory variables, of which four handled the Continuous and Categorical variables separately. The remaining eight models used only subsets of important variables, which were the Self-Selected, or Correlation Cutoff important variable subsets. Upon comparison, the linear models using SPCA on the Correlation Cutoff variables performed notably for internal tendons, and SPCA on all the variables performed significantly for the external tendons (see italic values in Table 4).

The following conclusions can be made from the above work:

- External and internal members show different levels of importance for the variables within the dataset. For instance, only A_{ps} was considered important to both internal and external predictions in the final SPCA equations. However, h, d_{ps}, LT, f_{pu}, ν_{AASHTO} and f_y were all considered important to internal ten-
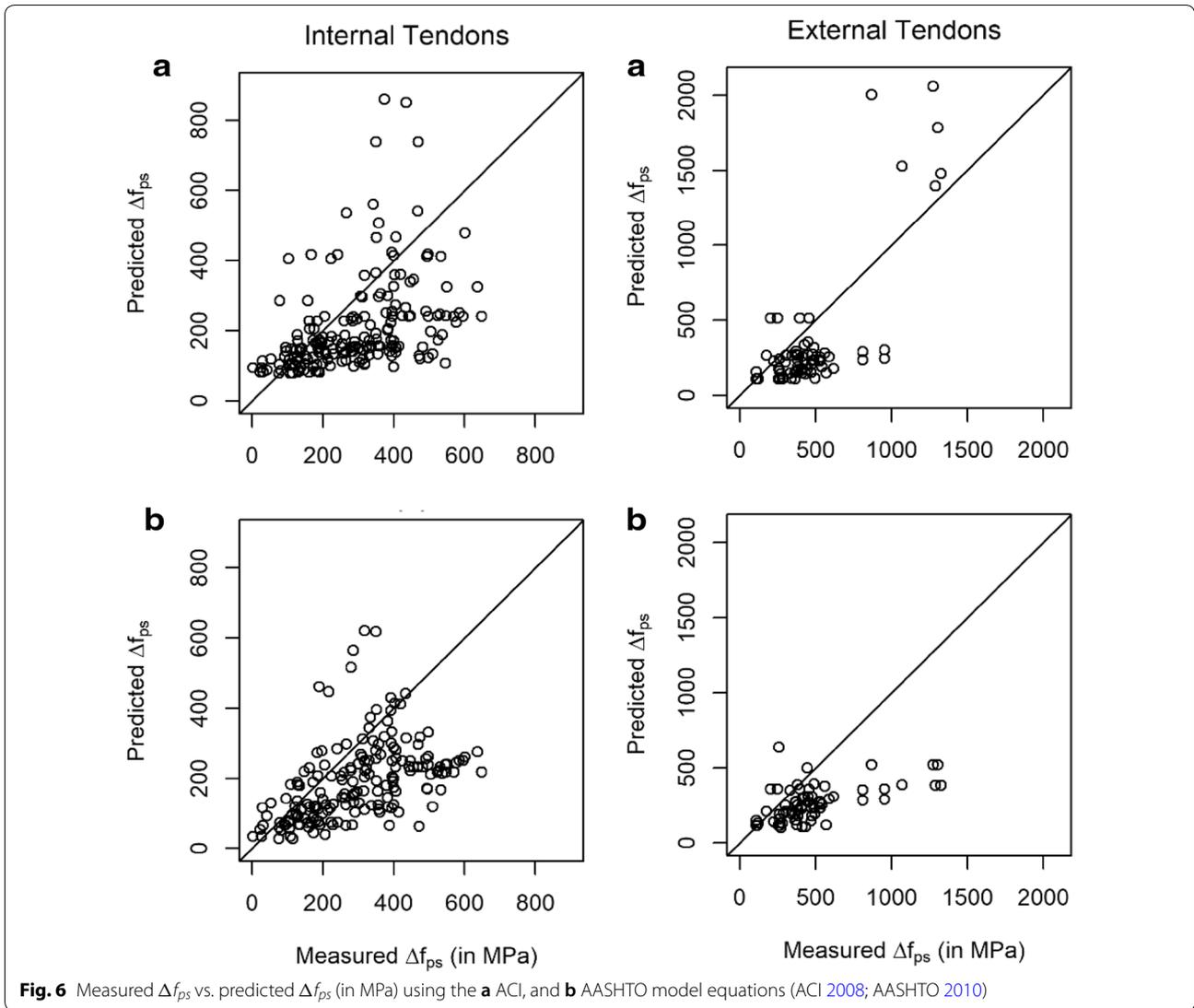
Fig. 6 Measured Δf_{ps} vs. predicted Δf_{ps} (in MPa) using the **a** ACI, and **b** AASHTO model equations (ACI 2008; AASHTO 2010)

Table 5 Cross tabulated R^2 and λ model values for simply supported and Continuous tendons

Variables	Simply supported				Continuous			
	Internal		External		Internal		External	
	R^2	λ	R^2	λ	R^2	λ	R^2	λ
AASHTO	0.30	1.71	0.02	2.42	0.18	1.82	0.11	1.95
ACI	0.47	1.90	0.12	2.73	0.04	1.31	0.06	2.50
Maguire et al. (2017)	0.27	1.34	0.06	1.48	0.18	1.34	0.17	1.25

dons, but none were important to external tendons. The reason for this is unclear, but is likely due to the differences in data contained in the dataset and phenomenological differences between the two structural systems. Interestingly, the influence of A_{ps} is a near consensus from the literature, but the other variables are disputed.

- Based on the above conclusion and the surveyed experimental and analytical literature, there is a significant need for more data in order to obtain better understanding, statistically and phenomenologically, of unbonded tendon reinforced members. This is ideally accomplished through additional testing, as

McKinney *et al. Int J Concr Struct Mater* (2019) 13:20

Page 17 of 18

the available database is relatively small compared to other member databases (e.g., Reineck et al. 2013).

- The SPCA-CC-Int model produced an $R^2 = 0.54$, $R_a^2 = 0.53$, $\lambda = 1.03$, RMSE $= 99.53$, and MAE $= 78.04$.
- The SPCA-All-Ext model produced an $R^2 = 0.70$, $R_a^2 = 0.69$, $\lambda = 0.99$, RMSE $= 152.79$, and MAE $= 110.93$.
- While the PCA and SPCA models performed similarly, according to the R^2 and λ metrics, SPCA combined with model selection techniques results in considerably shorter equations and produced better fit statistics.
- The PCA and SPCA analysis predicted significantly better than codified methods on the same dataset ($R^2 = 0.16$ and 0.08, $\lambda = 1.85$ and 2.01 for AASHTO and ACI respectively) and the optimized semi-empirical model presented by Maguire et al. (2017) ($R^2 = 0.27$ and $\lambda = 1.34$).
- The predicted stress increase, Δf_{ps}, is consistently under predicted for higher measured values of Δf_{ps} in the internal data (see Figs. 2, 5). Some of this is also exhibited in the external data though not as strongly. This suggests that an underlying non-linear relationship may be present in the data, and suggests further analysis possibly involving more advanced models.

List of symbols

A_{ps}: area of prestressing reinforcement (mm²); A_s: area of mild reinforcing steel on tension face (mm²); A_s': area of mild reinforcing steel on compression face (mm²); E_{ps}: modulus of elasticity of the prestressing reinforcement (MPa); L: total span length (m); LT: loading type (1.0 for single point load, 2.0 for third point loading, 3.0 for uniform loading); b: beam width (mm); c: depth from compression fiber to neutral axis (mm); d_{ps}: depth to prestressing reinforcement (mm); d_s: depth to tension mild reinforcing steel from compression face (mm); d_s': depth to compression mild reinforcing steel from compression face (mm); f_c': concrete strength (MPa); f_{pe}: effective stress in the prestressing reinforcement (MPa); Δf_{ps}: stress increase in unbonded tendons (MPa); $\overline{\Delta f_{ps}}$: predicted stress increase in unbonded tendons (MPa); f_{pu}: ultimate tendon strength (MPa); f_y: yield strength of mild reinforcing steel (MPa); h: beam height (mm); N: number of internal supports crossed by the tendon; v_{ACI}: variable part of the ACI prediction equation (MPa); v_{AASHTO}: variable part of the AASHTO prediction equation; μ: 100 if $L/d_{ps} \leq 35$, and 300 if $L/d_{ps} > 35$; ρ_{ps}: prestressed reinforcing ratio; ψ: scaled plastic hinge length.

Authors' contributions
All authors contributed significantly to the drafting of the manuscript. EM contributed the statistical programming, and interpretation of results. MC conducted the data curation, and participated in the sequence alignment. MM provided data access, context, and subject matter expertise. YS contributed statistical theory and application expertise. All authors read and approved the final manuscript.

Author details
[1] Department of Mathematics and Statistics, Utah State University, Logan, UT 84322, USA. [2] Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322, USA. [3] New Transportation Innovative Research Center, Korea Railroad Research Institute, 176, Cheoldo bangmulgwan-ro, Uiwang-si, Gyeonggi-do 16105, Republic of Korea.

References
AASHTO. (2010). *AASHTO load and resistance factor and design (LRFD) bridge design specifications*. Washington, DC: AASHTO.

Abdel-Qader, K. A. & Abudayyeh, O. (2004). Linear structure modeling and PCA algorithm for bridge crack detection. In *2004 IEEE electro/information technology conference*, Milwaukee, WI, pp. 138–142. https://doi.org/10.1109/eit.2004.4569378.

ACI (American Concrete Institute). (1963). Building code requirements for structural concrete and commentary. ACI 318-63, Farmington Hills, MI.

ACI (American Concrete Institute). (2008). Building code requirements for structural concrete and commentary. ACI 318-08, Farmington Hills, MI.

BSI (British Standards Institution). (2001). Structural use of concrete. BS8110, London.

Burns, N. H., Charney, F. A., & Vines, W. R. (1978). Tests of one-way post-tensioned slabs with unbonded tendons. *PCI Journal, 23*(5), 66–83.

Chang, M., Maguire, M., & Sun, Y. (2017). Framework for mitigating human bias in selection of explanatory variables for bridge deterioration modeling. *ASCE Journal of Infrastructure Systems*. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000352.

Chen, R. (1971). *The strength and behavior of post-tensioned prestressed concrete slabs with unbonded tendons*. MS Thesis, University of Texas, Austin, TX.

Cooke, N., Park, R., & Yong, P. (1981). Flexural strength of prestressed concrete members with unbonded tendons. *PCI Journal, 26*(6), 52–80.

DIN (Deutsches Institut für Normung). (1980). Spannbeton, dauteile mitvorspannung ohne verbund. DIN 4227, Teil 6, Berlin **(in German)**.

Du, G., & Tao, X. (1985). Ultimate stress of unbonded tendons in partially prestressed concrete beams. *PCI Journal, 31*(6), 72–91.

Gebre-Michael, Z. (1970). *Behavior of post-tensioned concrete slabs with unbonded tendon reinforcement*. MS Thesis. University of Texas, Austin, TX.

Harajli, M. (2011). Proposed modification of AASHTO-LRFD for computing stress in unbonded tendons at ultimate. *Journal of Bridge Engineering*. https://doi.org/10.1061/(asce)be.1943-5592.0000183.

Harajli, M., & Kanj, M. Y. (1991). Ultimate flexural strength of concrete members prestressed with unbonded tendons. *ACI Structural Journal, 88*(6), 663–673.

Harajli, M., Mabsout, M., & Al-Hajj, J. (2002). Response of externally post-tensioned continuous members. *ACI Structural Journal,, 99*(5), 671–680.

Hemakom, R. (1970). *Behavior of post-tensioned concrete slabs with unbonded tendons*. MS Thesis, University of Texas, Austin, TX.

Hua, X. G., Ni, Y. Q., Ko, J. M., & Wong, K. Y. (2007). Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *ASCE Journal of Computing in Civil Engineering*. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:2(122).

Jolliffe, I. (2002). *Principal component analysis* (2nd ed.). New York: Springer.

Kesavan, K. N., & Kiremidjian, A. S. (2012). A wavelet-based damage diagnosis algorithm using principal component analysis. *Structural Control and Health Monitoring, 19*(8), 672–685. https://doi.org/10.1002/stc.462.

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. New York: McGraw-Hill/Irwin.

Kuzniar, K., & Waszczyszyn, Z. (2006). Neural networks and principal component analysis for identification of building natural periods. *ASCE Journal of Computing in Civil Engineering, 20,* 6.

Lattin, J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data* (1st ed.). California, USA: Brooks/Cole–Thomson Learning, Pacific Grove.

Lee, L.-H., Moon, J.-H., & Lim, J.-H. (1999). Proposed methodology for computing of unbonded tendon stress at flexural failure. *ACI Structural Journal, 96*(6), 1040–1048.

Lou, T., Lopes, S., & Adelino, V. (2013). Flexural response of continuous concrete beams prestressed with external tendons. *ASCE Journal of Bridge Engineering., 18*(6), 525–537.

Lou, T. J., & Xiang, Y. Q. (2006). Finite element modeling of concrete beams prestressed with external tendons. *Engineering Structures, 28*(14), 1919–1926.

MacGregor, R. J. G. (1989). Strength and ductility of externally posttensioned segmental box girders. Doctoral dissertation, Univ. of Texas at Austin, Austin, TX.

Maguire, M., Chang, M., Collins, W. N., & Sun, Y. (2017). Stress increase of unbonded tendons in continuous posttensioned members. *Journal of Bridge Engineering, 22,* 04016115.

Maguire, M., Moen, C. D., Roberts-Wollmann, C., & Cousins, T. (2014). Field verification of simplified analysis procedures for segmental concrete bridges. *Journal of Structural Engineering, 141*(1), D4014007.

Mattock, A. H., Yamazaki, J., & Kattula, B. T. (1971). Comparative study of prestressed concrete beams, with and without bond. *Journal Proceedings, 68*(2), 116–125.

McKinney, E. (2017). Prediction of stress increase in unbonded tendons using sparse principal component analysis. All Graduate Plan B and other Reports. 1034.

Mojtahedi, S., & Gamble, W. L. (1978). Ultimate steel stresses in unbonded prestressed concrete. *Journal of the Structural Division, 104*(7), 1159–1164.

Naaman, A. E. (2005). PC Beams with unbonded tendons: Analysis in cracked, uncracked and ultimate state. In *Proceedings of Ned H. Burns on Historical Innovations in Prestressed Concrete*, ACI SP231. Edited by B.W. Russell and S.P. Gross. American Concrete Institute, Farmington Hills, Mich., pp 105–127.

Naaman, A. E., & Alkhairi, F. M. (1991). Stress at ultimate in unbonded post-tensioning tendons: Part 2-proposed methodology. *ACI Structural Journal., 88*(6), 683–692.

Nowak, A., & Collins, K. (2012). *Reliability of Structures* (2nd ed.). New York: CRC Press.

Ozkul, O., Nassif, H., Tanchan, P., & Harajli, M. (2008). Rational approach for predicting stress in beams with Unbonded Tendons. *ACI Structural Journal, 105*(3), 338–347.

Reineck, K. H., Bentz, E., Fitik, B., Kuchma, D., & Bayrak, O. (2013). ACI-DAfStb database of shear tests on slender reinforced concrete beams without stirrups. *ACI Structural Journal., 110*(5), 867–876.

Roberts-Wollmann, C. L., Kreger, M. E., Rogowsky, D. M., & Breen, J. E. (2005). Stresses in external tendons at ultimate. *ACI Structural Journal, 102*(2), 206.

SIA (Swiss Society of Engineers and Architects). (1979). *Ultimate load behavior of slabs*. SIA 162, Zurich, Switzerland.

Six, P.D. (2015) *Continuous unbonded post-tensioned members: Quantifying strand stress increase*. Master's Thesis. Utah State University.

Tam, A., & Pannell, F. N. (1976). The ultimate moment of resistance of unbonded partially prestressed reinforced concrete beams. *Magazine of Concrete Research, 28*(97), 203–208.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B, 58,* 267–288.

Trost, H., Cordens, H., Weller, B. (1984). Undersuchungen zur Vonspannung ohne Verbund. In *Deutscher Ausschu fur Stahlbeton*, Heft 355, Berlin, Germany.

Yan, L., Fraser, M., Elgamal, A., Fountain, T., & Oliver, K. (2008). Neural networks and principal component analysis for strain-based vehicle classification. *ASCE Journal of Computing in Civil Engineering*. https://doi.org/10.1061/(ASCE)0887-3801(2008)22:2(123).

Zhang, Y., McDaniel, J. G., & Wang, M. L. (2014). Estimation of pavement macrotexture by principal component analysis of acoustic measurements. *ASCE Journal of Transportation Engineering, 140,* 2. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000617.

Zolghadri, N. (2017). Short- and Long-Term Structural Health Monitoring of Highway Bridges. PhD Dissertation, Utah State University, Logan, Utah.

Zolghadri, N., Halling, M., & Barr, P. (2016). Effects of temperature variations on structural vibration properties. *Geotechnical and Structural Engineering Congress, 2016,* 1032–1043.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B, 67,* 301–320.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics, 15*(2), 265–286.