1. Edward KOZŁOWSKI[3] , 2. Krzysztof KRÓL[1], 3. Tomasz RYMARCZYK[1,2]

Research and Development Center, Netrix S.A (1), University of Economics and Innovation, Lublin (2), Lublin University of Technology (3)
ORCID: 1. 0000-0002-3524-9151; 2. 0000-0002-0114-2794; 3. 0000-0002-7147-4903

doi:10.15199/48.2022.04.27

Application of PCA with logistic regression in embankment drainage

Abstract*. The article presents a method using deep-sea probes, which were used to collect measurements in electrical tomography on the leakage of flood embankments. For this purpose, the main components analysis and elasticnet in logistic regression were used. The results of research on the method of spatial analysis of object moisture are presented. Research focused on the development and comparison of algorithms and models for data analysis and reconstruction using electrical tomography. The presented algorithms were used in the process of converting the input electrical values into the conductance represented by the pixels of the output image. The article presents PCA methods in logistic regression and elastic network in logistic regression to identify leakages in shafts. Deep probes were used to collect data in electrical impedance tomography.*

Streszczenie. *W artykule została zaprezentowana metoda wykorzystująca sondy głębinowe, które posłużyły do zbierania pomiarów w tomografii elektrycznej na temat przesiąkania wałów przeciwpowodziowych. W tym celu została wykorzystana analiza głównych składowych oraz elasticnet w regresji logistycznej. Przedstawiono wyniki badań nad metodą przestrzennej analizy zawilgocenia obiektów. Badania koncentrowały się na opracowaniu i porównaniu algorytmów i modeli do analizy i rekonstrukcji danych z wykorzystaniem tomografii elektrycznej. Przedstawione algorytmy zostały wykorzystane w procesie konwersji wejściowych wartości elektrycznych na konduktancję reprezentowaną przez piksele obrazu wyjściowego. W artykule przedstawiono metody PCA w regresji logistycznej oraz sieci elastycznej w regresji logistycznej do identyfikacji wycieków w szybach. Do zbierania danych w tomografii impedancji elektrycznej wykorzystano sondy głębinowe. (****Zastosowanie PCA z regresją logistyczną w odwadnianiu wałów przeciwpowodziowych****)*

Keywords: logistic regresion,tomography,seepage of embarkment
Słowa kluczowe: regresja logistyczna, tomografia, przesiąkanie nasypów

Introduction

In order to identify infiltration in flood banks, the imaging domain has been modeled as a mesh that consists of a set of finite elements. For each finite element, a model was created that shows the relationship between the signal obtained from the electrodes and the conductivity. Whenever cases related to leakage are analyzed, there is a variable conductivity within the test object. It is related to the concentration of water in a given area. It is important to recognize the place of flooding or excessive moisture. For this purpose, a model was established for each finite element. These models allowed the computation of non-background conductivity probabilities for each of the finite elements. Based on this approach, the resolution in the imaging domain was determined and thus the mesh was reconstructed. Many different methods are used to solve optimization problems [1-11]. The presented solution was based on electrical impedance tomography [12-18]. In order to identify the areas correctly, the ROC analysis [23,24] should be performed for each finite element, thus the classification level was determined.

Methods

Principal Component Analysis is based on the identification of factors (components) occurring in a data set $X \epsilon R^{n \times m}$, eg [19-22,25]. The data set consists of n observations for m variables (in R^m it is a certain cloud of n points). The goal of PCA is to rotate the coordinate system in such a way as to maximize first the variability of the first coordinate, then the variability of the second coordinate, etc. The coordinates of the new system are called the loads of the generated principal components. In the new space, the initial factors explain most of the variability. PCA is often used to reduce the size of a statistical dataset by discarding recent factors [20].

For X matrix we applied Singular Value Decomposition. Each real matrix X can be represented as:

$$X = UDV^T + \varepsilon_x \tag{1}$$

where $U \epsilon R^{n \times m}$ is the left matrix of orthogonal vectors (matrix of left singular vectors), $D \epsilon R^{m \times m}$ is the diagonal matrix containing the eigenvalues. - the diagonal matrix of singular values, and $V \epsilon R^{m \times m}$ - the right orthogonal vector matrix (matrix of right singular vectors). The orthogonal matrix V satisfies the following property $V^T=V^{-1}$. The decomposition of the matrix X can be presented in the form

$$X = TP^T + \varepsilon_x \tag{2}$$

where T=UD denote coordinates in the orthogonal system and V=P denote loadings. The coordinate matrix T in the new space is determined by multiplying both sides of the equality $X=TP^T$ by P and obtaining

$$XP = TP^T P = T \tag{3}$$

the columns V of the matrix contain the weights used in the linear combination to form the new dimensions. The variances of the coordinates in the new space are expressed by the formula

$$\lambda_i = \frac{d_i^2}{n-1} \tag{4}$$

where d_i, $0 \le i \le k$ are singular values from the diagonal of the matrix D.

In EIT, for each case, the reading from the electrodes taking into account the polarization and projection angles (according to (3) correspond to the values $x_{(j)}$ P in the new space. For each finite element we determine logistic regression, except that we analyze the linear dependence of the log of odds on the value in new coordinates. For each finite element, based on the $x_{(j)} \epsilon R^m$ measurements obtained from electrodes, we determine the probability of inclusion $P(Y=1| x_{(j)})$. Second application is based on applying elasticnet to logistic regression [13, 20, 22, 23]. Reconstructions for each approach were compared.

The basic terminology and coefficients describing the recognition of inclusions in the visual field are presented below. Below, we assume the absence of an inclusion in the place of a finite element as a negative case (N), and the occurrence of an inclusion as a positive case (P). The confusion matrix should be determined as follows: TP (True Positive) means the number of finite elements for which

inclusions were correctly identified, TN (True Negative) - the number of finite elements for which the absence of inclusions was correctly recognized, FP (False Positive) - the number of elements finite elements without inclusions for which it has been recognized that they have inclusions (false alarm), FN (False Negative) - the number of finite elements with inclusions for which it has been recognized that they do not have inclusions.

Therefore, the Accuracy size represents the portion of the visual field that has been correctly recognized by the model. On the other hand, it is one of the measures that directly shows the correctness of the diagnosis.

In the EIT, the possibilities of finding inclusions in the visual field should also be described during image reconstruction. To determine the ability of a classifier based on the use of logistic regression (see eg [26,27]), we determine the Receiver Operating Characteristic (ROC curve) curve. This curve shows the relationship between sensitivity and specificity during the reconstruction. The diagonal in the ROC drawing describes a strategy based on guessing the inclusions during the reconstruction. If the ROC is above the diagonal, it means that the recognition technique is much better than guesswork. The area under the ROC curve in the literature is called AUC (Area under ROC curve) and is a measure of predictivity (predictability). This figure is also included in the tables describing the reconstructions.

Examples

The estimation of structural parameters for each of the finite elements was performed in the R programming language.

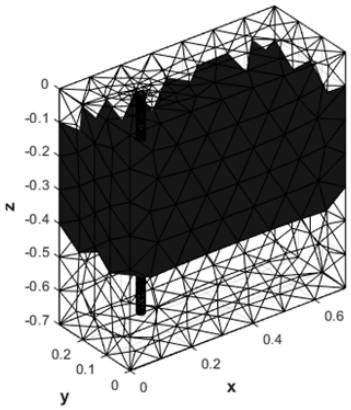

Fig.1. Example 1 pattern

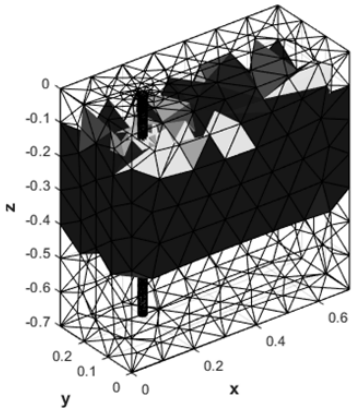

Fig.2. Example 1 reconstruction based on logistic regression with PCA

The glmnet function from the glmnet package was used to determine the coefficients in the model using Elasticnet. The prcomp function was used to determine the main components (when using PCA). The number of components used for the models with 16 electrodes is 25, while for the models with 32 electrodes it is 30. Figures 1-12 show the test object model and image reconstruction.

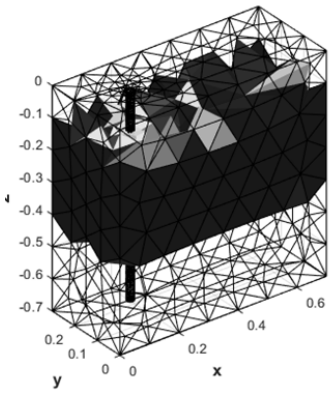

Fig.3. Example 1 reconstruction based on logistic regression with Elasticnet

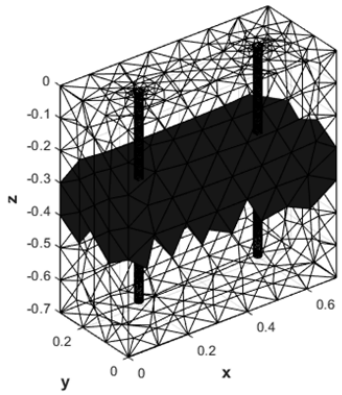

Fig.4. Example 2 pattern

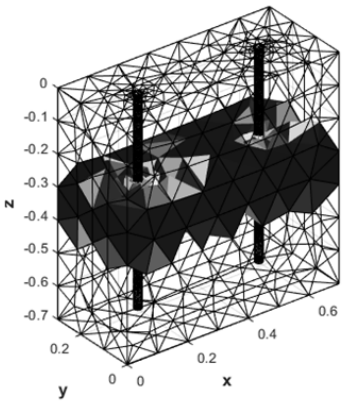

Fig.5. Example 2 reconstruction based on logistic regression with PCA

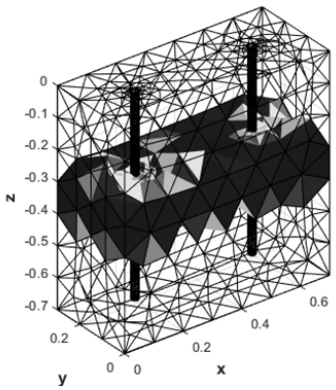

Fig.6. Example 2 reconstruction based on logistic regression with Elasticnet

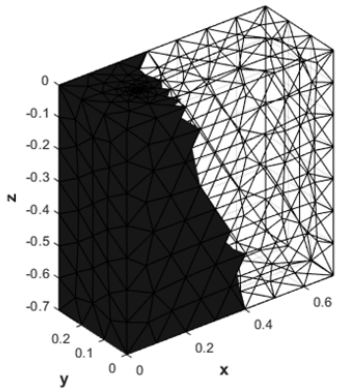

Fig.7. Example 3 pattern

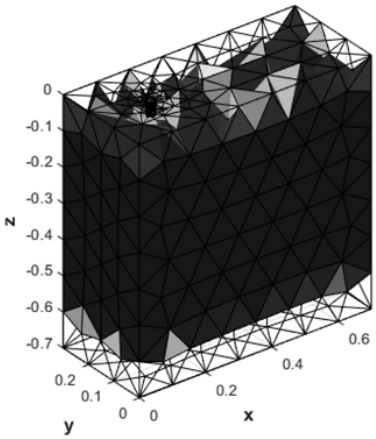

Fig.8. Example 3 reconstruction based on logistic regression with PCA

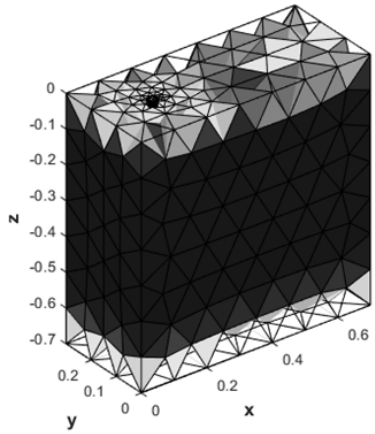

Fig.9. Example 3 reconstruction based on logistic regression with Elasticnet

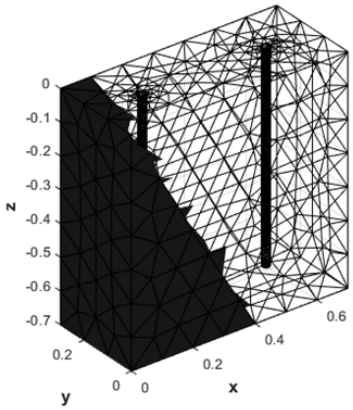

Fig.10. Example 4 pattern

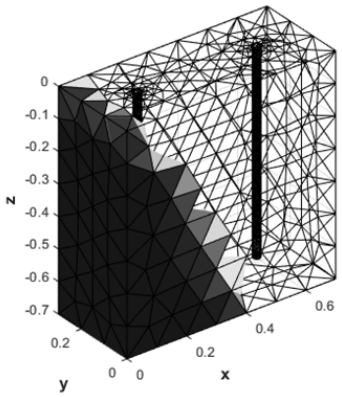

Fig.11. Example 4 reconstruction based on logistic regression with PCA

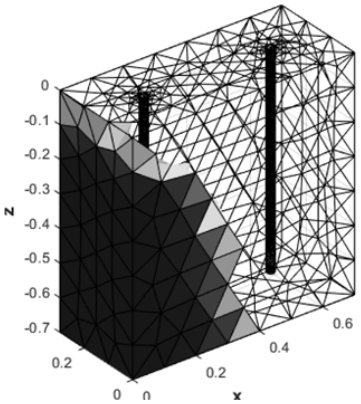

Fig.12. Example 4 reconstruction based on logistic regression with Elasticnet

Conclusion

The presented solution allows for a better understanding and monitoring of the properties of the tested object. The main challenge in this field is to design precise measurement devices and image reconstruction algorithms. The article presents PCA methods in logistic regression and an elastic net in logistic regression to identify leakages in shafts. Deep-electrodes were used to collect data in electrical impedance tomography.

Table 1. Evaluation metrices for examples

	Elasticnet example3	PCA example3	Elasticnet example4	PCA example4
Sensitivity	0.8761783	0.8462420	0.9809559	0.9831533
Specificity	0.2569170	0.2845850	0.9775489	0.9849708
Precision	0.9481665	0.9483226	0.9567780	0.9707105
Recall	0.8761783	0.8462420	0.9809559	0.9831533
Prevalence	0.9394447	0.9394447	0.3362685	0.3362685
Detection Rate	0.8231211	0.7949976	0.3298645	0.3306034
Detection Prevalence	0.8681187	0.8383198	0.3447660	0.3405788
Balanced Accuracy	0.5665477	0.5654135	0.9792524	0.9840620

Authors: *Tomasz Rymarczyk, D.Sc, Ph.D. Eng., University of Economics and Innovation, Projektowa 4, Lublin,. E-mail: tomasz@rymarczyk.com; Krzysztof Król, Research&Development Centre Netrix S.A., Email: krzysztof.krol@netrix.com.pl; Edward Kozłowski Ph.D.Eng.,Lublin University of Technology, Nadbystrzycka 38, Lublin, E-mail: e.kozlowski@pollub.pl;*

REFERENCES

[1] Miłak, M., Leszczyńska, A., Grudzień, K., Romanowski, A., & Sankowski, D. Slug flow velocity estimation during pneumatic conveying of bulk solid materials based on image processing

techniques. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska*, 9 (2019), No. 1, 11-14
[2] Kryszyn, J., Wanta, D., Smolik, W. T., Evaluation of the electrical capacitance tomography system for measurement using 3d sensor. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska*, 9 (2019), No. 94, 52-59
[3] Kłosowski G., Rymarczyk T., Wójcik D., Skowron S., Adamkiewicz P., The Use of Time-Frequency Moments as Inputs of LSTM Network for ECG Signal Classification, *Electronics*, 9 (2020), No. 9, 1452
[4] Korzeniewska E., Krawczyk A., Stando J., Torsion field - an example of pseudo-scientific concept in physics, *Przeglad Elektrotechniczny*, 97 (2021), No.1,196-199
[5] Korzeniewska, E; Szczesny, A; Lipinski, P; Drozdz, T; Kielbasa, P; Miernik, A, Prototype of a Textronic Sensor Created with a Physical Vacuum Deposition Process for Staphylococcus aureus Detection, *Sensors*, 21 (2021), No. 1, 183
[6] Wajman, R; Banasiak, R; Babout, L, On the Use of a Rotatable ECT Sensor to Investigate Dense Phase Flow: A Feasibility Study, *Sensors*, 20 (2020), No. 17, 4854
[7] Banasiak, R.; Wajman, R.; Jaworski, T.; Fiderek, P.; Fidos, H.; Nowakowski, J.; Sankowski, D. Study on two-phase flow regime visualization and identification using 3D electrical capacitance tomography and fuzzy-logic classification. *Int. J. Multiph. Flow*, 58 (2014), 1–14
[8] Dusek, J.; Mikulka J., Measurement-Based Domain Parameter Optimization in Electrical Impedance Tomography Imaging, *Sensors*, 21 (2021), No. 7, 2507
[9] Daniewski K., Kosicka E., Mazurkiewicz D., Analysis of the correctness of determination of the effectiveness of maintenance service actions. *Management and Production Engineering Review*, 9 (2018); No. 2, 20-25
[10] Romanowski, A. Contextual Processing of Electrical Capacitance Tomography Measurement Data for Temporal Modeling of Pneumatic Conveying Process. *In Proceedings of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Poznan, Poland, 9–12 September (2018); 283–286
[11] Chen, B.; Abascal, J.; Soleimani, M. Extended Joint Sparsity Reconstruction for Spatial and Temporal ERT Imaging. *Sensors* 18 (2018), 4014
[12] Rymarczyk T., Kłosowski G., Hoła A., Sikora J., Wołowiec T., Tchórzewski P., Skowron S., Comparison of Machine Learning Methods in Electrical Tomography for Detecting Moisture in Building Walls, *Energies*, 14 (2021), No. 10, 2777
[13] Rymarczyk T., Kozłowski E., Kłosowski G., Electrical impedance tomography in 3D flood embankments testing – elastic net approach, *Transactions of the Institute of Measurement and Control*, 42 (2020), No. 4, 680-690
[14] Rymarczyk T., Nita P., Vejar A., Woś M., Stefaniak B., Adamkiewicz P.: Wearable mobile measuring device based on electrical tomography, *Przegląd Elektrotechniczny*, 95 (2019), No. 4, 211-214
[15] Rymarczyk T., Kłosowski G., Tchórzewski P., Cieplak T., Kozłowski E., Area monitoring using the ERT method with multisensor electrodes, *Przegląd Elektrotechniczny*, 95 (2019), No. 1, 153-156
[16] Koulountzios P., Rymarczyk T., Soleimani M., A quantitative ultrasonic travel-time tomography system for investigation of liquid compounds elaborations in industrial processes, *Sensors*, 19 (2019), No. 23, 5117
[17] Kłosowski G., Rymarczyk T., Kania K., Świć A., Cieplak T., Maintenance of industrial reactors based on deep learning driven ultrasound tomography, *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, 22 (2020), No. 1, 138–147
[18] Kłosowski G., Rymarczyk T., Cieplak T., Niderla K., Skowron Ł., Quality Assessment of the Neural Algorithms on the Example of EIT-UST Hybrid Tomography, *Sensors*, 20 (2020), No. 11 3324
[19] Wehrens, R., Chemometrics with r, *Springer-Verlag GmbH*, 2011
[20] Hastie,T.; Tibshirani,R.; Friedman, J., The elements of statistical learning, *Springer-Verlag New York Inc*., 2009
[21] James, G.; Witten, D.; Hastie, T.; Tibshirani, R., An introduction to statistical learning, *Springer-Verlag GmbH*, 2013
[22] Zou, H.; Hastie, T., Regularization and variable selection via the elastic net, Journal of the Royal *Statistical Society: Series B (Statistical Methodology)*, 67 (2005), 301–320
[23] Tibshirani, R., Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*.. 58 (1994) 267–288
[24] Friedman, J.; Hastie, T.; Tibshirani, R., Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, 33 (2010), 1-22
[25] Xin Yan, X.G.S., Linear regression analysis, World Scientific Publishing Company, 2009
[26] Fawcett, T., An introduction to ROC analysis, *Pattern Recognition Letters*, 27 (2006) 861–874
[27] Hand, D.J.; Till, R.J., A simple generalisation of the area under the roc curve for multiple class classification problems, *Machine Learning*. 45 (2001), 171–186