A Regularization Approach for Identifying Cumulative Lagged Effects in Smart Health Applications

Karthik Srinivasan
Faiz Currim
Sudha Ram
INSITE: Center for Business Intelligence and Analytics
Department of Management Information Systems
University of Arizona, Tucson
{karthiks, currim, sram}@email.arizona.edu

Casey Lindberg
Esther Sternberg
Perry Skeath
Institute on Place and Wellbeing
University of Arizona, Tucson
{caseylindberg, esternberg, perryskeath}@email.arizona.edu

Bijan Najafi
Javad Razjouyan
Hyo-Ki Lee
Interdisciplinary Consortium on Ambulatory Motion Performance
Baylor College of Medicine, Houston
{bijan.najafi, javad.razjouyan, Hyoki.Lee}@bcm.edu

Matthias R. Mehl
Department of Psychology
University of Arizona, Tucson
mehl@email.arizona.edu

Davida Herzl
Reuben Herzl
Melissa Lunden
Nicole Goebel
Scott Andrews
Aclima, Inc., San Francisco
{davida.herzl, reuben.herzl, melissa.lunden, nicole.goebel, scott.andrews}@aclima.io

Brian Gilligan
Judith Heerwagen
Kevin Kampschroer
Kelli Canada
U. S. General Services Administration, Washington DC
{brian.gilligan, judith.heerwagen, kevin.kampschroer, kelli.canada}@gsa.gov

ABSTRACT

Recent development of wearable sensor technologies have made it possible to capture concurrent data streams for ambient environment and instantaneous physiological stress response at a fine granularity. Characterizing the delay in physiological stress response time to each environment stimulus is as important as capturing the magnitude of the effect. In this paper, we discuss and evaluate a new regularization-based statistical method to determine the ideal lagged effect of five environmental factors—carbon dioxide, temperature, relative humidity, atmospheric pressure and noise levels on instantaneous stress response. Using this method, we infer that the first four environment variables have a cumulative lagged effect, of approximately 60 minutes, on stress response whereas noise level has an instantaneous effect on stress response. The proposed transformations to inputs result in models with better fit and predictive performance. This study not only informs the field of environment-wellbeing research about the cumulative lagged effects of the specified environmental factors, but also proposes a new method for determining optimal feature transformation in similar smart health studies.

DH '17, July 02-05, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5249-9/17/07…$15.00
http://dx.doi.org/10.1145/3079452.3079503

KEYWORDS

Indoor environment quality; heart rate variability; cumulative lag; environment-wellbeing studies; smart health

1 INTRODUCTION

With rapid development of sensor technologies and the internet of things, the scope of smart health applications has considerably widened. Understanding the effects of indoor environment on individual wellbeing in office workplaces is one such application that has gained importance recently [7, 21]. Environment-wellbeing studies can now be conducted by measuring instantaneous effects of an environment of interest on observable health artifacts, instead of traditional experience surveys and controlled experiments that suffer from problems of internal and external validity [4]. Environmental factors such as ambient noise, temperature, air quality and humidity can be measured in fixed (indoor or outdoor) sensors as well as mobile (wearable) sensors in real time. Simultaneously, heart rate, sleep, activity, cortisol monitors and one click mobile surveys can measure individual physiological and psychological wellbeing accurately. Modeling the environment wellbeing relationship at minute-level granularity has become possible due to such smart sensors, thus enabling researchers to ask more complex questions.

A cumulative lagged effect is a special case of functional transformations of the input (e.g., lags, logarithms, exponential), commonly encountered in smart health applications that have multilevel streaming (sensor) data. We therefore propose a new method for determining optimal feature transformation for multilevel streaming data in this study. This method selects one among k transformations of an input, based on its weighted sum of penalized coefficients over grouped data. It is a regularization-based ranking method consisting of three steps: (a) fitting an ensemble of lasso models (b) generating features importance score as a weighted sum of lasso coefficients, and (c) selecting the feature transformation with highest score.

We apply this method on real data used to model effects of ambient environment quality factors—CO2, noise level, temperature, atmospheric pressure and relative humidity on physiological stress. Our method determines that ambient noise has an instantaneous effect and environmental factors—temperature, relative humidity, pressure and CO2 have a cumulative lagged effect of one hour on physiological stress of an individual. Predictive performance of regression models fitted with the proposed input transformations are compared with the performance of models with other input sets. Models with inputs following the proposed cumulative lagged effects are observed to perform better than other models.

The remainder of this paper is structured as follows. In section 2, we discuss related work in methods to determine optimal cumulative lagged effects. In section 3, we describe the proposed regularization based ranking method to identify optimal feature transformations in multilevel data. Sections 4 and 5 compare and evaluate the validity of the method. Section 6 presents a discussion of the findings along with study limitations. Conclusion and areas for future work are in Section 7.

2 RELATED WORK

There are several studies that analyze effects of environment factors such as temperature, air quality and noise level on individual wellbeing [5, 10, 14, 20]. Previous studies either do not model instantaneous effects measured at less than a one-hour granularity [2, 14] or have assumed instantaneous effects [11, 15]. MacNaughton et al. [9] address the problem of cumulative lagged effect identification for CO2 on heart rate by comparing the estimated coefficients and *p*-values for CO2 with different cumulative lags. They suggest a cumulative effect of an hour on heart rate variability and other health outcomes. To the best of our knowledge, there is no prior theory or evidence regarding the presence or absence of cumulative lagged effects of other environmental factors such as temperature, humidity or ambient noise.

The approach used by MacNaughton et al. is heuristic and depends on the analyst's subjective inputs while inspecting the coefficients. On the other hand, traditional stepwise feature selection procedures are ridden with challenges such as sensitivity to changes in data and low external validity [6]. The problem of selecting the right cumulative lag for an input is different from feature selection and hence, standard feature selection methods [22] cannot be directly used. We therefore require a validated method that can be applied across multilevel data scenarios such as the ones used in previous environment-wellbeing studies. With such a method, we can then identify cumulative lagged effects for smart wearable systems (SWS) that measure vital signs such as body/skin temperature, heart rate variability, arterial blood pressure, respiration and activity [3].

3 METHODOLOGY

We propose a new regularization based ranking method called *mixed lasso* to identify optimal feature transformations in multilevel data. In this section, we first formulate cumulative lagged effects as set of candidate feature transformations, and then describe the mixed lasso method as a three-step process for optimal feature transformation identification.

A cumulative lagged effect of an input is the effect of sum of the values of inputs measured from current state to a finite interval p. The cumulative lag is also known as area under the curve (AUC) [1] or window of exposure [16] effect. The problem of identification of the right cumulative lagged effect of an input in multilevel streaming data can be presented as a feature engineering problem. That is, we generate functional transformations of an input that are cumulative lagged effects with intervals $K = \{k_i \mid k_i \subseteq [0, p]\}$. The K cumulative lags can be represented as the set $\{x'_1 = x_t, x'_2 = x_t + x_{x-1} + \cdots + x_{x-k_1}, x'_3 = x_t + x_{x-1} + \cdots + x_{x-k_2}, \dots, x'_K = x_t + x_{x-1} + \cdots + x_{x-k_{K-1}}\}$. For the limiting case, p+1 cumulative lagged effects of the input x_t are given by the set $\{x_t, x_t + x_{x-1}, x_t + x_{x-1} + x_{x-2}, \dots, x_t + x_{x-1} + \cdots + x_{x-p}\}$; which is the distributed lag representations of finite lags $k \in$ {0,...,p}. For streaming data, the intervals k are often interpretable multiples of temporal units, such as multiples of 15 minutes (e.g., instantaneous value of temperature, AUC of temperature for past 15 minutes, AUC of temperature for past 30 minutes, AUC of temperature for past 45 minutes and so on).

Streaming datasets in smart health applications typically have a two-level structure of repeated measures of sensor information (level-1) across participants (level-2). For such a two-level data, we propose a three-step procedure to determine the optimal cumulative lag for each input. First, fit lasso-regularized linear regression model to each of the m participants. Secondly, combine the coefficients using a weighted pooling strategy to get an overall importance score of each cumulative lagged effect with interval p for the input. The third step is to rank the cumulative lagged effects in decreasing order of their score and select the one with the highest score as the predictor in the model.
Mathematically, we can represent the steps as follows:

Step I: Fit lasso models for each of the M group data

$$L_m : \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(y_i^{(m)} - \beta_0 - \sum_{j=1}^{d} x'^{(m)}_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{d} |\beta_j| \right\} \tag{1}$$

The coefficients for each element in the feature subsets $\{x_1^{'}, x_2^{'}, \ldots, x_k^{'}\}^r \subseteq X$ for inputs $r \in R$ is therefore determined for each group m.

Step II: Generate the feature importance score as a weighted sum of each lasso coefficient

$$\beta_j = \frac{1}{n} \sum_{m=1}^{M} \beta_j^m W^m \quad (2)$$

$$W^{(m)} = \frac{S(m)}{\sqrt{L\left(\hat{Y}(m)\right)}}$$

$S(m)$ and $L\left(\hat{Y}(m)\right)$ are size and loss functions for linear model fit for group m respectively.

Step III: Determine the optimal feature corresponding to each input r as

$$\max_{x_j^r \subset X^r} \beta_j^r \quad (3)$$

Step I is fitting standard lasso models for M groups or clusters in the multilevel data. Here, we circumvent the need to account for within group dependence [18]. For multilevel data with more than 2 levels, we propose an $M \; x \; N$ factoring approach (e.g., with four groups in level 2 and 10 groups in level 3, we can implement step one for 40 factored groups). In Step II, we propose the weights $W^{(m)}$ as a ratio of cardinality of each group $S(m)$ and square root of the loss function $L\left(\hat{Y}(m)\right)$ of the model for the corresponding group. This ensures that the $\beta^{(m)}$ for model corresponding to group m is penalized for smaller representation as well as inferior fit, compared to coefficients of models for other groups in the dataset. Step III simply selects the best feature from set of feasible cumulative lagged effects for input X^r, as the final step of the feature extraction process.

4 EXPERIMENTAL EVALUATION

We applied the mixed lasso method to determine cumulative lagged effects in a real data used to model effects of ambient environment quality factors—CO2, noise level, temperature, atmospheric pressure and relative humidity on physiological stress. The data was generated by a field experiment conducted as part of a multidisciplinary research program supported by the General Services Administration (GSA) to study the impact of the workplace environment on individual wellbeing. The experimental setup consisted of participants wearing two sensors for three days while carrying out their day-to-day activities: (a) A heart rate monitor and (b) A personal environment quality sensor-based device[2]. Two hundred and thirty-one participants across multiple locations participated in the experiment during July 2015 through November 2016. The data for this study was hierarchical, where participants are the secondary level of data abstraction. Short term RMSSD (*Root mean square of successive differences*), SDNN (*Standard deviations of NN interval*), normalized HF (*normalized high frequency component*) and LF/HF (*low frequency to high frequency ratio*) measured at every 5 minutes are four heart rate variability (HRV) indicators that measure the instantaneous physiological stress response of individuals [8, 23]. HRV has been considered as a proxy measure for the wellbeing of a person, i.e., higher its value, higher the wellbeing of the subject [13, 14, 21]. One of the primary objectives of this study was to characterize the effects of ambient environment quality factors—CO2, noise level, temperature, atmospheric pressure and relative humidity on physiological stress.

The input variables for investigation of cumulative lagged effects are temperature, noise level, CO2, relative humidity and pressure measured by the personal environment quality sensor-based device. The outcome variables are SDNN, RMSSD, normalized HF and LF/HF which are heart rate variability measures of physiological stress response. Other covariates such as participant demographics (e.g., Body mass index (BMI), Age, Gender), temporal indicators (Time of day, Day of the week) and activity-level (actigraph measure that gauges movement of participants) are included in the model. After data integration, preprocessing and cleaning, our final dataset contained approximately 200,000 minutes of heart rate monitor and environment quality data streams. The dataset was randomly split into training and test datasets (75:25 split) for analysis.

Different values of cumulative lags for each of the inputs were considered, e.g., every 5 minutes, every 10 minutes, every 15 minutes, and so on. A spacing of 30 minutes was determined as optimal for interpretability and generalizability. Three versions of each input (instantaneous value, 30-minute cumulative lag and 60-minute cumulative lag) are presented in this study. Note: we analyzed higher granularities (e.g., 90-minute cumulative lag, 120-minute cumulative lag), but our results did not change.

We compared the performance of a model with the proposed feature set (denoted as *Mixed lasso*) with other models having following feature sets: (a) Only instantaneous inputs (b) Only 30 minutes cumulative lagged inputs, (c) Only 60 minutes cumulative lagged inputs, (d) Instantaneous, 30 minutes and 60 minutes cumulative lagged versions of inputs (denoted as *all cumulative lag*) (e) Supervised stepwise feature selection using AIC (denoted as *MinAIC*). The fixed effects model was used as a baseline, denoting the performance when only fixed effects of instantaneous inputs are considered in the multilevel model.

Table 1: Model fit comparison using Pseudo R Squared

Pseudo R squared	RMSSD	SDNN	Norm. HF	LF/HF
Fixed effects only (baseline)	0.6724	0.6644	0.5152	0.4544
Instantaneous	0.7736	**0.6988**	0.5968	0.4994
30 mins cumulative lag	0.7802	0.6936	0.5896	0.4971
60 mins cumulative lag	**0.7883**	0.6970	0.5938	0.4981
All cumulative lag	0.7734	0.6975	0.5987	0.5007
MinAIC	0.7860	0.6970	0.5952	**0.5042**
Mixed lasso	0.7856	0.6987	**0.5995**	0.5011

[2] Aclima, Inc. is the research partner that provides the fixed and mobile environmental sensor data.

Table 2: Prediction accuracy comparison using RMSE

RMSE	RMSSD	SDNN	Norm. HF	LF/HF
Fixed effects only (baseline)	8.5419	17.9597	9.1037	6.5198
Instantaneous	7.6826	17.4308	8.6555	6.4416
30 mins cumulative lag	7.5697	17.4366	8.7345	6.4455
60 mins cumulative lag	7.5226	17.3872	8.7065	6.4509
All cumulative lag	7.6769	17.4307	8.6442	6.4414
MinAIC	7.5293	17.4307	8.7071	**6.4372**
Mixed lasso	**7.5131**	**17.3448**	**8.6364**	6.4420

Table 3: Prediction accuracy comparison using MAE

MAE	RMSSD	SDNN	Norm. HF	LF/HF
Fixed effects only (baseline)	5.8293	12.4376	6.3599	2.4439
Instantaneous	5.1534	11.9572	6.0014	2.3441
30 mins cumulative lag	5.0853	11.9575	6.0631	2.3475
60 mins cumulative lag	5.0664	11.9049	6.0488	2.3437
All cumulative lag	5.1534	11.9535	**5.9953**	2.3424
MinAIC	**5.0610**	11.9571	6.0387	**2.3342**
Mixed lasso	5.0637	**11.8837**	6.0008	2.3348

Table 4: Prediction accuracy comparison using MAPE

MAPE	RMSSD	SDNN	Norm. HF	LF/HF
Fixed effects only (baseline)	25.12	24.30	38.11	168.79
Instantaneous	21.79	23.17	35.83	167.01
30 mins cumulative lag	21.46	23.10	36.18	166.40
60 mins cumulative lag	21.39	22.97	36.12	166.23
All cumulative lag	21.80	23.17	35.88	166.86
MinAIC	21.42	23.17	36.09	166.27
Mixed lasso	**21.37**	**22.92**	**35.79**	**166.15**

Hierarchical linear models (HLM) were fit for all the previously stated input feature-sets, for each of the four outcomes. We included the environmental inputs as fixed as well as random effects across participants [18]. Model fit was checked using pseudo R-Squared [12]. Predictive performance was compared using Root Mean Squared error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Prediction Error (MAPE) on the test dataset.

Using the mixed lasso method, 60 minutes cumulative lagged versions of for temperature, CO2, pressure, relative humidity and instantaneous version of noise level were identified as optimal feature representations for the environment factors across all four HRV outcomes (SDNN, RMSSD, norm HF and LF/HF).

The model fit and prediction accuracy comparisons across models described in previous section are shown in tables: Table 1, Table 2 Table 3 and Table 4. The model fit and error estimates for best performing models are highlighted for reader convenience. Better model fit and prediction accuracy corresponds to a higher value in Table 1 and lower values in Table 2, Table 3 and Table 4 respectively. We see from Table 1 that models fit using the proposed mixed lasso method do not over-fit, but the performance is at least second best in terms of prediction accuracy as seen in Table 2, Table 3 and Table 4. This shows that models with input feature-set derived using the mixed lasso method have best overall predictive performance, indicating that the intervals selected for cumulative lagged effects of each of the inputs are optimal.

5 DISCUSSION

The proposed regularization based ranking method for determining optimal input feature transformation in smart health applications has better prediction accuracy than existing approaches. It is robust in the presence of noise in data and avoids manual errors in inspection or stepwise methods. This study contributes to statistical method literature as well as to environment-wellbeing domain literature, by suggesting a 60-minute cumulative lag effect for four out of the five inputs of interest.

There are some limitations to this study based on assumptions made. We have focused on the problem of determination of optimal lagged effects of inputs over an outcome in a longitudinal setting, but do not delve into the problem of determining significance of the inputs themselves. However, our proposed approach can be used as a pre-processing step based on which subsequent analysis can determine the relative significance of each input in the model. A second limitation is the assumption of availability of continuous streaming data. Streaming data can often have missing values, e.g., removal or disconnection of sensors, values beyond sensor range, etc. [19]. In our work, we manually inspected such intervals and handled the missing values accordingly. However, the presence of discontinuity in streaming data can limit the possibility of assessing validity of larger cumulative lags. Hence, the set of candidate longitudinal transformations, such as cumulative lagged effects have to be determined based on available data. A third constraint on our approach is that it is applicable for multilevel data streams, and not designed to determine cumulative lagged effects in single-level longitudinal data applications (e.g., studying effect of alcohol consumption on brain activity in a single person over a time period).

In addition to the linear multilevel regression models, we also tested the feature extraction method using the RE-EM tree, a tree-based multilevel model fitting data mining algorithm [17]. The models with all positive feature representations of inputs (full model) performed better than any other model, including the model fit using mixed lasso method. Hence, we conclude that our proposed mixed lasso method is more suitable for explanatory statistical modeling than tree-based predictive modeling.

6 CONCLUSIONS

Innovative methods of collecting recording artifacts in real-time using multiple mobile/wearable sensors have opened up unexplored channels of exploratory research in smart health applications. In this study, we show that modeling delay in output response is as important as capturing the magnitude of the effect of the corresponding input. We proposed a novel method for identifying the cumulative lagged effect for inputs in multilevel streaming data. It uses a regularization-based ranking approach to determine best set of features representing the environment factors. It is robust, efficient and provides better prediction accuracy for the given data.

Our proposed method is validated and therefore the identified cumulative lagged effects are more reliable. The environment-wellbeing study presented in this paper is just one of the several smart health applications which need to take into account cumulative lagged effects. Our work facilitates environment-wellbeing research, by serving as a guideline for transforming variables such as, temperature, noise level, CO2, relative humidity and pressure into specific cumulative lagged versions, when analyzing heart rate variability.

We presented the application of our method in identification of optimal cumulative lagged effects in the environment-wellbeing data. However, our approach is capable of addressing a wider set of problems of selecting one out of k transformations of inputs of multilevel models, when there is little or no prior theory for the input to output functional relationship. With an increasing number of smart health applications, this method can prove useful to improve prediction performance of explanatory models as well as contribute to the literature pertaining to the domain of functional relationships.

As part of future work, we plan to apply our method to identify optimal input transformation in other smart health problems. We also plan to build a software package implementing the mixed lasso method described in this study.

7 COMPETING INTERESTS

The authors have declared that no competing interests exist.

REFERENCES

[1] Bergouignan, A., Legget, K.T., De Jong, N., Kealey, E., Nikolovski, J., Groppel, J.L., Jordan, C., Raphaela, O., Hill, J.O. and Bessesen, D.H. 2016. Effect of frequent interruptions of prolonged sitting on self-perceived levels of energy, mood, food cravings and cognitive function. *Int J Behav Nutr Phys Act.* 13, 1 (2016), 113.

[2] Chan, C.C., Chuang, K.J., Su, T.C. and Lin, L.Y. 2005. Association between nitrogen dioxide and heart rate variability in a susceptible population. *European Journal of Cardiovascular Prevention and Rehabilitation.* 12, 6 (2005), 580–586.

[3] Chan, M., Estève, D., Fourniols, J.-Y., Escriba, C. and Campo, E. 2012. Smart wearable systems: Current status and future challenges. *Artificial Intelligence in Medicine.* 56, 3 (2012), 137–156.

[4] Fowler, F.J. 2009. *Survey research methods.*

[5] Gold, D.R., Litonjua, A., Schwartz, J., Lovett, E., Larson, A., Nearing, B., Allen, G., Verrier, M., Cherry, R. and Verrier, R. Ambient Pollution and Heart Rate Variability.

[6] Hastie, T., Tibshirani, R., Friedman, J. and Hastie, Trevor, Tibshirani, Robert, Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* The Mathematical Intelligencer.

[7] Heerwagen, J. and Zagreus, L. 2005. The human factors of sustainable building design: post occupancy evaluation of the Philip Merrill Environmental Center. *Center for the Built Environment UC Berkeley.* (2005), 0–13.

[8] Kuch, B., Hense, H.W., Sinnreich, R., Kark, J.D., von Eckardstein, A., Sapoznikov, D. and Bolte, H.D. 2001. Determinants of short-period heart rate variability in the general population. *Cardiology.* 95, 3 (Jan. 2001), 131–8.

[9] MacNaughton, P., Spengler, J., Vallarino, J., Santanam, S., Satish, U. and Allen, J. 2016. Environmental perceptions and health before and after relocation to a green building. *Building and Environment.* 104, (2016), 138–144.

[10] Mitchell, C.S., Zhang, J.J., Sigsgaard, T., Jantunen, M., Lioy, P.J., Samson, R. and Karol, M.H. 2007. Current state of the science: health effects and indoor environmental quality. *Environmental health perspectives.* 115, 6 (Jun. 2007), 958–64.

[11] Mizukoshi, A., Kumagai, K., Yamamoto, N., Noguchi, M., Yoshiuchi, K., Kumano, H. and Yanagisawa, Y. Real-time measurements of VOC exposure and heart rate variability in indoor and outdoor environments.

[12] Nakagawa, S. and Schielzeth, H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution.* 4, 2 (2013), 133–142.

[13] Pieper, S., Brosschot, J.F., van der Leeden, R. and Thayer, J.F. 2007. Cardiac Effects of Momentary Assessed Worry Episodes and Stressful Events. *Psychosomatic Medicine.* 69, 9 (2007), 901–909.

[14] Ren, C., O'Neill, M.S., Park, S.K., Sparrow, D., Vokonas, P. and Schwartz, J. 2011. Ambient Temperature, Air Pollution, and Heart Rate Variability in an Aging Population. *American Journal of Epidemiology.* 173, 9 (2011), 1013–1021.

[15] Riojas-rodr Guez, H., Antonio Escamilla-cejudo, J., Antonio Gonz Lez-hermosillo, J., Mar Té Llez-rojo, M.A., Vallejo, M., Santos-burgoa, C. and Rojas-bracho, L. 2006. Personal PM 2.5 and CO exposures and heart rate variability in subjects with known ischemic heart disease in Mexico City. *Journal of Exposure Science and Environmental Epidemiology.* 167500453, (2006), 131–137.

[16] Satish, U., Mendell, M.J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S. and Fisk, W.J. 2012. Is CO2 an indoor pollutant? direct effects of low-to-moderate CO2 concentrations on human decision-making performance. *Environmental Health Perspectives.* 120, 12 (2012), 1671–1677.

[17] Sela, R.J. and Simonoff, J.S. 2009. RE-EM Trees: A New Data Mining Approach for Longitudinal Data. *Statistics Working Papers Series.* (2009), 1–27.

[18] Singer, J.D. and Willett, J.B. 2009. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.*

[19] Srinivasan, K. et al. 2016. Feature importance and prediction modeling for multi-source healthcare data with missing values. *Proceedings of the 6th International Conference on Digital Health 2016* (2016), 1–8.

[20] Sun Sim, C., Hyun Sung, J., Hyeon Cheon, S., Myung Lee, J., Won Lee, J. and Lee, J. 2015. The Effects of Different Noise Types on Heart Rate Variability in Men. *Yonsei Med J http.* 56, 1 (2015).

[21] Thayer, J.F., Verkuil, B., Brosschot, J.F., Kampschroer, K., West, A., Sterling, C., Christie, I.C., Abernethy, D.R., Sollers, J.J., Cizza, G., Marques, A.H. and Sternberg, E.M. 2010. Effects of the physical work environment on physiological measures of stress. *European Journal of Cardiovascular Prevention & Rehabilitation.* 17, 4 (2010), 431–439.

[22] Tibshirani, R. 1996. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B.*

[23] Xhyheri, B., Manfrini, O., Mazzolini, M., Pizzi, C. and Bugiardini, R. 2012. Heart Rate Variability Today. *Progress in Cardiovascular Diseases.* 55, 3 (2012), 321–331.