# Unraveling Temporal Dependencies in Stress: Insights from Self-Reported Stress Data

Heejeong Lim
KAIST
Graduate School of Data Science
Daejeon, Korea
hj.lim@kaist.ac.kr

Duri Lee
KAIST
School of Computing
Daejeon, Korea
durilee@kaist.ac.kr

Uichin Lee
KAIST
School of Computing
Daejeon, Korea
uclee@kaist.ac.kr

## ABSTRACT

Understanding how stress evolves over time is crucial for improving stress detection research. This study examines temporal dependencies in self-reported stress data. We analyzed three self-reported stress datasets to explore how past and present stress levels correlate, utilizing the autocorrelation function (ACF). Our finding quantitatively showed that temporal dependencies in stress levels vary across participants, and the degree of these dependencies differs across datasets collected in different contexts. We provide some insights on how to consider temporal dependencies in self-reported stress for stress detection models, taking into account individual and contextual variations.

## CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Applied computing** → *Health informatics.*

## KEYWORDS

Stress, Temporal Dependency, Self-Reported Data, ACF

**ACM Reference Format:**
Heejeong Lim, Duri Lee, and Uichin Lee. 2024. Unraveling Temporal Dependencies in Stress: Insights from Self-Reported Stress Data. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3675094.3678421

## 1 INTRODUCTION

Accumulated stress in everyday life can harm human physiological and psychological health [1, 3]. Therefore, measuring stress in daily life and providing timely stress-relief interventions can improve well-being. Today, mobile and wearable sensors continuously collect physiological, behavioral, and contextual information from users [1], helping us indirectly understand their daily lives. This opportunity has led to many studies that aim to detect stress in human daily life in real-time based on sensor data [6, 8, 9, 13]. Existing research on stress detection mainly uses self-reported stress as ground truth for measuring stress in daily life [8, 9]. The most widely used method for collecting self-reported stress data is the "experience sampling method (ESM)" [9]. ESM is a technique that periodically sends a short questionnaire asking participants to report their inner states of mind, such as mood and stress [10].

Depending on the assumption of how stress occurs over time, self-reported stress data and sensor data can be utilized differently in stress detection research. Several researchers assume that stress occurs independently at any given time (or each two consecutive stress self-reports are independent of one another) [2, 4]. These studies built stress detection models that predict self-reported stress labels based only on sensor data. In contrast, other studies suggest that past and current stress levels may be interdependent, indicating the importance of using past self-reports to predict current stress. For example, Toshnazarov et al. used the self-reported stress data just before the current self-reported label as a predictor [13]. Additionally, Li et al. extracted temporal features from the wearable sensor data collected over the 24 hours preceding the target stress data as predictors [11].

Since the assumptions regarding the temporal dependency of stress differ across studies, it is necessary to investigate how temporal dependency occurs. In this paper, we analyze a self-reported dataset about stress to quantify temporal dependency. Three datasets of self-reported stress collected from daily lives were used for the analysis. We utilized the autocorrelation function (ACF), which measures the degree of correlation between current and past data. Our results showed that various temporal dependencies are exhibited among participants and datasets. Given these observations, we discuss how temporal dependency within self-reported stress can be taken into account in stress detection models.

## 2 DATA

We used self-reported stress labels from the three different datasets collected in daily life. While two datasets were collected throughout the entire day, the other dataset was collected exclusively from the workplace (call center). The details about each dataset (i.e., participants, methods, and results) are described below.

*Dataset 1*: The first dataset is the K-EmoPhone dataset, which collected daily stress from 77 university students (F: 24, M: 53; age: Mean = 21.9, SD = 3.9) for a week [8].[1] Participants received the questions to assess their stress levels between 10 AM and 10 PM, with an average interval of 45 minutes. They recorded their stress level at the moment of response using a 7-point Likert scale. As a result, participants provided an average of 10.4 of self-reported data in a day, and the average interval between answers was 70.5 minutes (SD = 75.7 minutes).

[1] K-EmoPhone is available at https://zenodo.org/records/7606611



*UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia*

ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3678421

*Dataset 2*: The second dataset is the DeepStress dataset, which collected daily stress from 24 university students (F: 9, M: 15; age: Mean = 21.3, SD = 2.1) for six weeks [7].[2] Participants reported their stress levels at any time during the day. They were restricted from reporting their stress if the interval was shorter than 30 minutes, to prevent excessive repetition of self-reporting. Additionally, reminders were sent if the participants did not respond within an hour to collect enough data at appropriate intervals. As a result, participants reported an average of 12.8 of self-reported data in a day (SD = 3.0), and the average interval between answers was 93.5 minutes (SD = 108.5 minutes).

*Dataset 3*: The third is the dataset at work, which collected a self-reported stress level after one call from 18 customer service agents (F: 17, M: 1; age: Mean = 36.8, SD = 5.91) for four weeks. Participants were asked to rate the stress level after each call, which is the basic task unit in the call center, as a 5-point Likert scale. As a result, participants answered an average of 37.2 of self-reported data in a day (SD = 5.5), and the average interval between answers was 13.2 minutes (SD = 15.7 minutes).

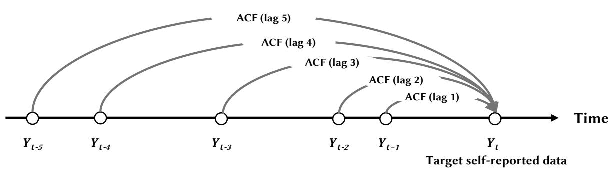


**Figure 1: Flow of using ACF.**

## 3 IRREGULAR TIME SERIES ACF ANALYSIS

We analyze the temporal dependency among the self-reported stress data utilizing autocorrelation function (ACF), which is widely used for time series data analysis [5, 12]. The ACF measures the correlation between present and past data within a single time series sequence. ACF can be useful for time series data because such data vary over time, and as a result, data points collected at adjacent times may be correlated. Since self-reported stress data is collected through ESM to capture people's stress as it changes over time, ACF can uncover hidden correlations within the data.

When calculating the ACF, lag indicates how far in the past it is from the present. Lag values can be integer values, such as 0, which is never lagged from the present (i.e., itself), 1, which is lagged 1 time from the current data (i.e., immediate past point), and 2, which is lagged 2 times from the present. A large lag value represents a more distant past, and the ACF values for it indicate a correlation between current data and data from the distant past.

ACF value is between -1 and 1, with the large absolute values indicating a higher correlation. For example, a positive autocorrelation value for lag 1 suggests that the current self-reported stress level is likely similar to the previous one. This can be interpreted as indicating the possibility that past data may affect the current stress level. In contrast, a negative autocorrelation value indicates that the current self-reported stress level tends to be opposite to the previous one. If the autocorrelation value nears zero, it indicates that the current self-reported stress level does not correlate with past values. The formula of ACF is as shown in Equation 1, where k is the lag and $y_t$ is the data point value at time $t$; the process of using ACF is illustrated in Figure 1.

$$ACF(k) = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \quad (1)$$

The results of ACF can be more accurate when the time intervals are regular rather than irregular. However, it is inevitable to have irregular time intervals due to the missing value inherent in the nature of ESM [14]. In this paper, we used the original data without additional data processing to regularize the intervals. Instead, to minimize the influence of the irregular time interval on the value of ACF, we only get the value of ACF on the same day, excluding relatively long-term intervals such as sleep and getting off work. Dataset 1 and 3 has a fixed time to collect the data in a day. However, in Dataset 2, participants could respond at any time. This led to some data points being collected during early morning hours (e.g., midnight, 1 AM). These times could be associated with the previous day's activities, making it challenging to categorize them as data for the next day. Hence, we only used data collected from 10 AM to 10 PM in Dataset 2. We used the day with more than 10 labeled data [9] and set the maximum lag to 5 which is half of the 10. After calculating the value of ACF, we aggregated the data per participant and obtained the average value.

## 4 RESULTS

Figure 2 describes the average ACF per participant at each time lag. The graphs on the left and center are the results of Datasets 1 and 2, which were collected in a similar data collection context. The boxplots for these datasets show that the average ACF value tends to be positive at lag 1, but becomes negative as the lag increases. In particular, over 64% of participants in Dataset 1 and over 83% in Dataset 2 exhibit positive ACF values at lag 1. In contrast, at subsequent lags, at least 66% of participants in both datasets show negative ACF values. This indicates that for a majority of participants, the immediate past self-reported stress data is relatively similar to the current self-reported stress data, while more older past data tends to show negative autocorrelation. In addition, we observed that, except for lag 1, the distributions of the boxplots at other lags are quite similar. This can show that even at more distant past points, there is a similar level of negative autocorrelation.

In Datasets 1 and 2, we identified that each participant has different average ACF values because the boxplots are spread up and down based on correlation 0. For Dataset 1, the ranges of ACF values (i.e., minimum and maximum values) were calculated as follows: for lag 1, it was (-0.33, 0.38); for lag 2, it was (-0.36, 0.24); for lag 3, it was (-0.46, 0.33); for lag 4, it was (-0.37, 0.23); and for lag 5, it was (-0.33, 0.28). For Dataset 2, the ranges of ACF values were as follows: for lag 1, it was (-0.03, 0.29); for lag 2, it was (-0.26, 0.11); for lag 3, it was (-0.19, 0.07); for lag 4, it was (-0.21, 0.01); and for lag 5, it was (-0.19, -0.10). In Dataset 1 with larger participants, the distribution of ACF values among participants is more widely spread. This shows that the degree of temporal dependency can vary between individuals.

[2]Deepstress dataset is available at https://github.com/Kaist-ICLab/DeepStress_Dataset

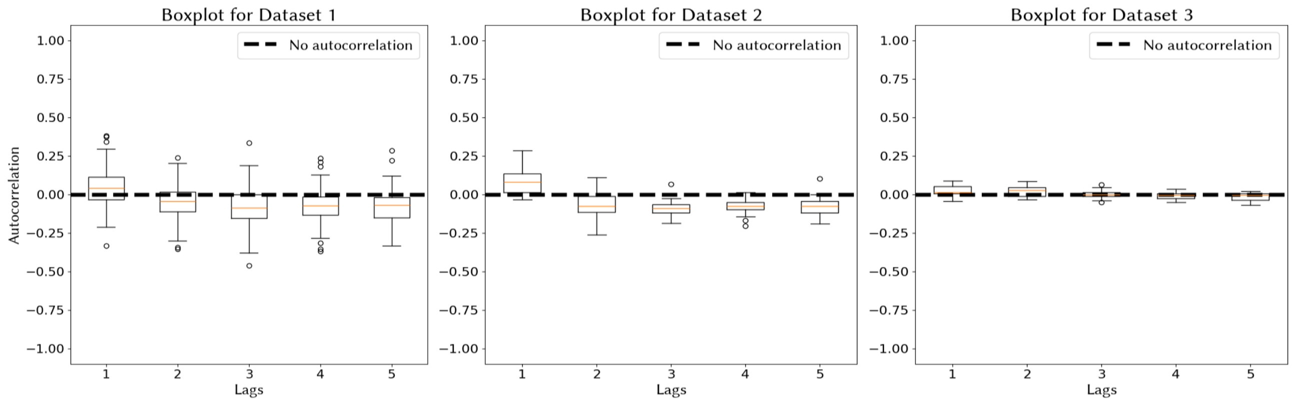


**Figure 2: Results of ACF plots**

The graphs on the right are the results of Dataset 3. All participants had weak correlations close to zero, ranging from a minimum of -0.06 to a maximum of 0.09 across all lags. This indicates that the self-reported stress data collected in Dataset 3 is relatively uncorrelated to each other compared to the other datasets.

## 5 DISCUSSION

We quantitatively assessed how dependent self-reported stress data are on past time points utilizing ACF. Through this analysis, we observed that in Datasets 1 and 2, a significant number of participants exhibited the highest similarity between current stress level and immediate past stress level among the five past time points. In contrast, Dataset 3 showed that there was almost no dependency on past stress data for all participants. This observation highlights that we can determine the optimal duration of past data to include when building stress detection models for a general user population. For example, for Datasets 1 and 2, using immediate past self-reported stress data and sensor data collected approximately 60–90 minutes earlier could be beneficial. In contrast, using older past data in Dataset 3 could add noise to the stress detection model, and thus, using a shorter period of historical data might be beneficial. Additionally, we quantitatively identified the individual variation in temporal dependency of self-reported stress data among participants in Datasets 1 and 2. This can help optimize the use of past data for each individual when we build personalized stress detection models. For example, if a participant has a high ACF value, current stress levels are highly dependent on past stress levels. In such cases, it may be beneficial to include features related to past data in the personalized model. Utilizing ACF can provide an opportunity to effectively incorporate the temporal dependency within self-reported stress data into stress detection models.

Meanwhile, we observed that Dataset 3, which was collected in the workplace, exhibited almost no temporal dependency when compared to Datasets 1 and 2, which were collected over everyday life contexts. This finding showed that temporal dependency can vary depending on the context of data collection. Dataset 3 was collected from call center workers about how stressed they were about a past call event that happened just before. Since Dataset 3 only asked about stress for one target event, participants may not have been influenced by distant past experiences. Therefore, to appropriately consider temporal dependency in stress detection models, it is necessary to assess temporal dependency under diverse contexts, whenever self-reported stress data is collected.

## ACKNOWLEDGMENTS

This research was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2022-0-00064) and by the NIPA (National IT Industry Promotion Agency) grant funded by the Korean government (MSIT) (No. H1001-24-1002).

## REFERENCES

[1] Yekta Said Can, Bert Arnrich, and Cem Ersoy. 2019. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of biomedical informatics* 92 (2019), 103139.
[2] Matteo Ciman and Katarzyna Wac. 2016. Individuals' stress assessment using human-smartphone interaction analysis. *IEEE Transactions on Affective Computing* 9, 1 (2016), 51–65.
[3] Sheldon Cohen, Ronald C Kessler, and Lynn Underwood Gordon. 1997. *Measuring stress: A guide for health and social scientists.* Oxford University Press, USA.
[4] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. 2015. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics* 20, 4 (2015), 1053–1060.
[5] Martin T Hagan and Suzanne M Behr. 1987. The time series approach to short term load forecasting. *IEEE transactions on power systems* 2, 3 (1987), 785–791.
[6] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing.* 493–504.
[7] Gyuwon Jung, Sangjun Park, and Uichin Lee. 2024. DeepStress: Supporting Stressful Context Sensemaking in Personal Informatics Systems Using a Quasi-experimental Approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–18.
[8] Soowon Kang, Woohyeok Choi, Cheul Young Park, Narae Cha, Auk Kim, Ahsan Habib Khandoker, Leontios Hadjileontiadis, Heepyung Kim, Yong Jeong, and Uichin Lee. 2023. K-emophone: A mobile and wearable dataset with in-situ emotion, stress, and attention labels. *Scientific data* 10, 1 (2023), 351.
[9] Zachary D King, Judith Moskowitz, Begum Egilmez, Shibo Zhang, Lida Zhang, Michael Bass, John Rogers, Roozbeh Ghaffari, Laurie Wakschlag, and Nabil Alshurafa. 2019. Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 3 (2019), 1–22.
[10] Reed Larson and Mihaly Csikszentmihalyi. 1983. The experience sampling method. *New directions for methodology of social & behavioral science* (1983).
[11] Boning Li and Akane Sano. 2020. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. *Proceedings of the ACM on Interactive, Mobile,*

*Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–26.

[12] Vijay P Singh, Rajendra Singh, Pranesh Kumar Paul, Deepak Singh Bisht, and Srishti Gaur. 2024. Time series analysis. In *Hydrological Processes Modelling and Data Analysis: A Primer*. Springer, 35–71.

[13] Kobiljon Toshnazarov, Uichin Lee, Byung Hyung Kim, Varun Mishra, Lismer Andres Caceres Najarro, and Youngtae Noh. 2024. SOSW: Stress Sensing with Off-the-shelf Smartwatches in the Wild. *IEEE Internet of Things Journal* (2024).

[14] Niels van Berkel, Jorge Goncalves, Lauri Lovén, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2019. Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* 125 (2019), 118–128.