

# Summary Report and Analysis of the Intel Lab Data

*by linhai\_li@hotmail.com*

## Information from the Visualization of Original Data

Figures 1 and 2 show the time series of temperature, humidity and voltage for each sensors. Figures 3 and 4 show the time series of light and voltage. The x-axis is the time offset in second from the earliest deployment time. The reason to show the time offset is because (1) epoch does not correspond to the increase of time and (2) it is much easier and consistent to look at the temporal behavior of each sensor.

An obvious fact is that sensor 5 and 28 have some problems, meaning they have data missing. For the remaining sensors, I cannot see much from light data when I observe some interesting pattern on the temperature, humidity and voltage data. For example, the temperature and humidity go crazy when voltage goes below 2.2 – 2.4 volts. This suggests that maintenance is required to ensure the sensors to work appropriately by monitoring the remained voltage. The data also suggests that humidity inversely correlated with temperature, and temperature positively correlated with the voltage data after detrending.

## Data Cleaning and More Information

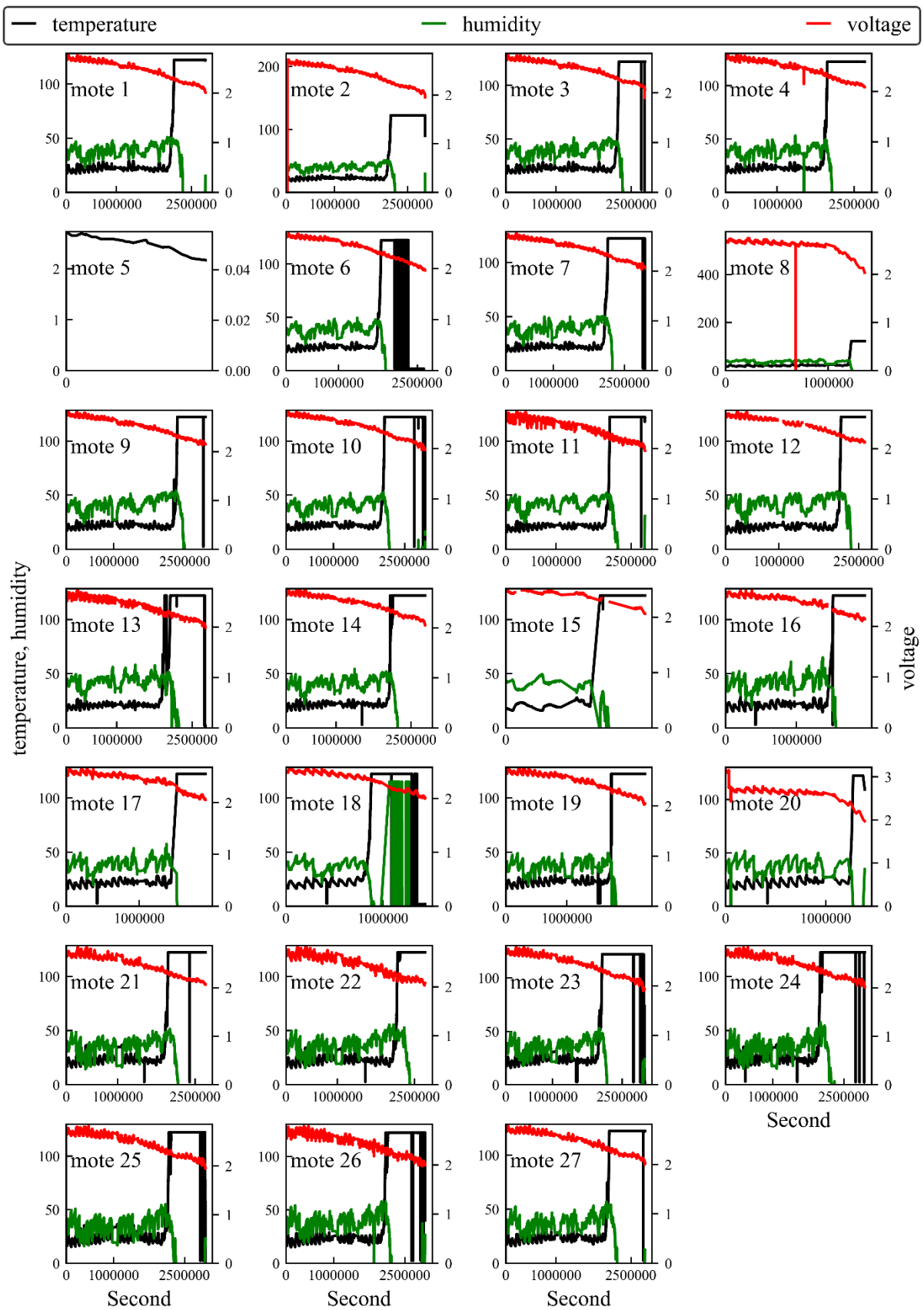
Based on the initial analysis of original data, data cleaning was performed. The strategy to clean the data is tentative and is not necessarily great one. I simply decided that temperature higher than 40 °C is not realistic inside an office. The very high temperature usually corresponds to very low humidity and to voltage below 2.2 – 2.4 volts. It means that I removed the unrealistic data at the later time. In addition, I also despiked each quantity and only kept temperature and humidity within their 1% - 99% percentile. The cleaned data are shown in Figures 5 to 8. They are in the same form as in Figures 1 to 4 but for cleaned data.

From the cleaned data, there are many things to discover. I only show a couple examples that I think are interesting. As shown in Figure 9(a), how long good measurements can be expected from temperature and humidity sensors are predictable based on the initially available voltage. It means we need to schedule maintenance of the sensors, i.e., changing batteries, at the time of deployment based on the initial voltage of a battery. Figure 9(b) also shows that the drainage rate of the battery is correlated with the standard deviation of temperature, and Figure 9(c) shows relationship between the drainage rate and standard deviation of humidity. It means that the schedule to change batteries should be adjusted according to the environment where the sensors are deployed.

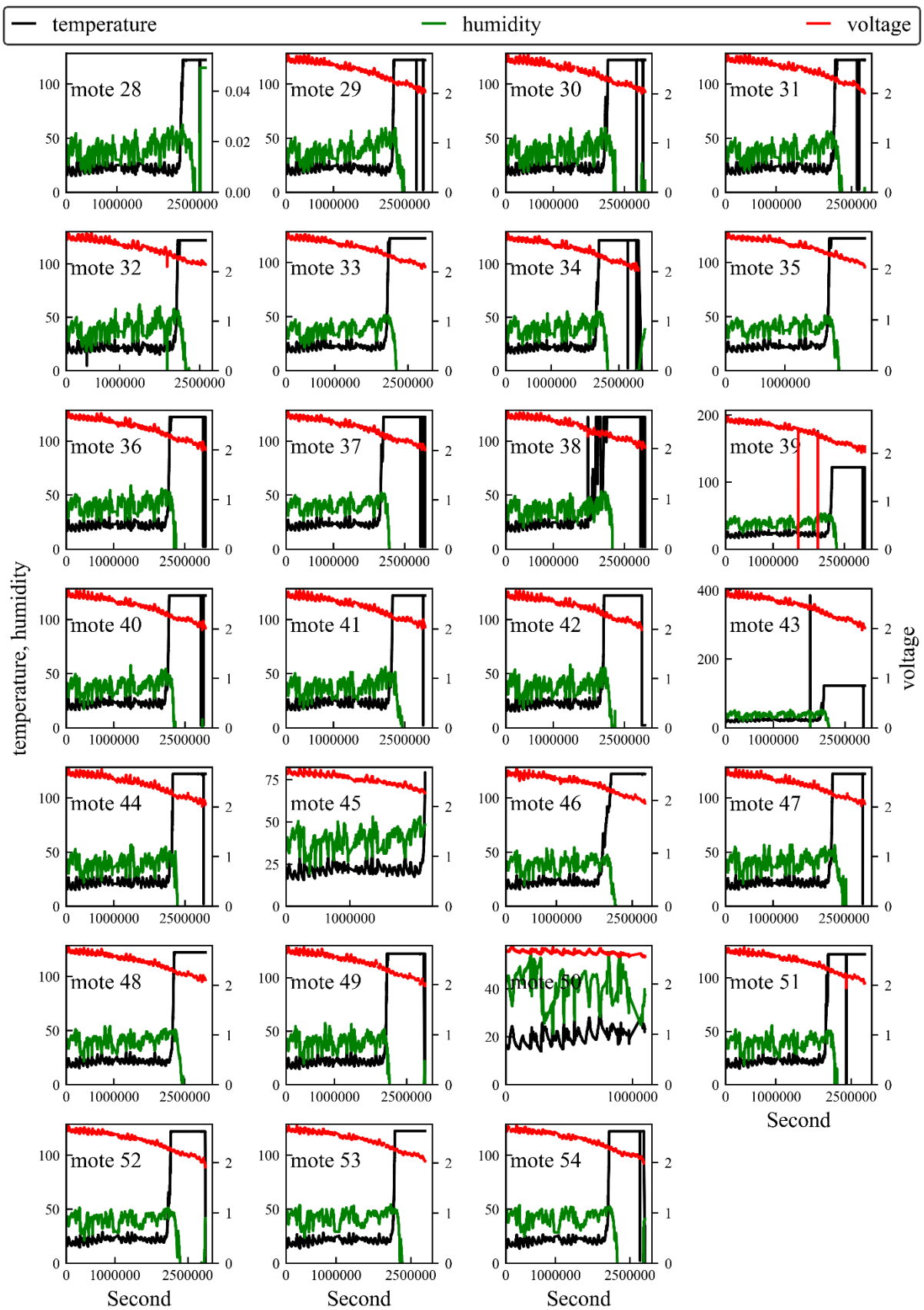
## A Little Information by Clustering the Cleaned Data

After cleaning the data as described above, I extract information such as max, min, mean, and standard deviation (STD) of temperature, humidity, and light. Based on such information, I attempted to cluster the sensors into different groups/environments. As described earlier, different environments may require different attention for maintenance.

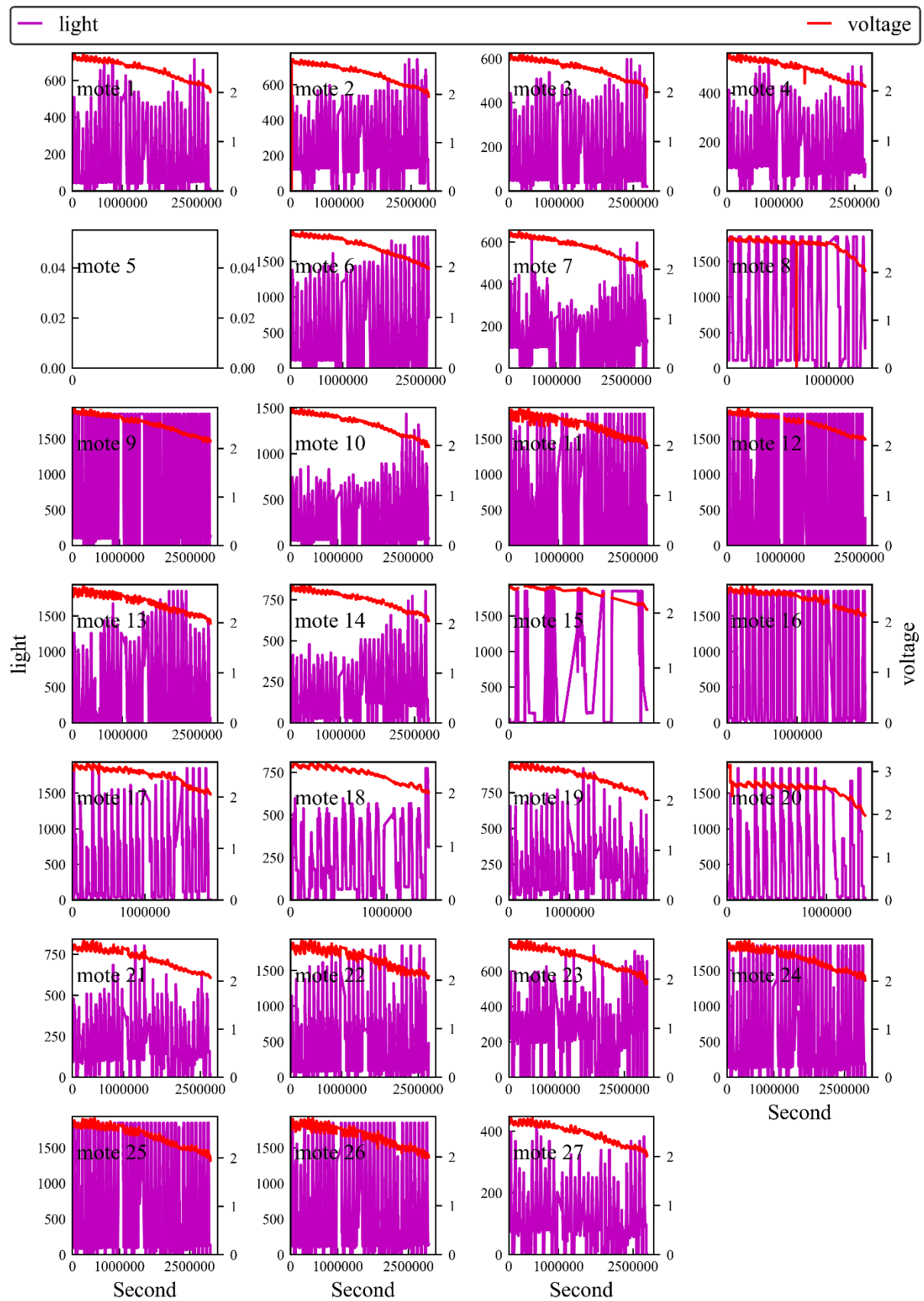
I used the k-means to cluster the data. Based on the "elbow feature" by plotting the cost function and number of clusters, I decided to cluster the sensors into 10 groups. The results is shown in Figure 10. For example, sensors represented by red dots seem to align against the wall. More thorough analysis should be done to understand why the sensors at each group has similar environmental features (such as influence of ventilation, facing direction of windows) and it will help us to better understand the maintenance needs for each group. I do not have complete data to do such analysis; therefore, I show this as an example to extract information using machine-learning techniques.



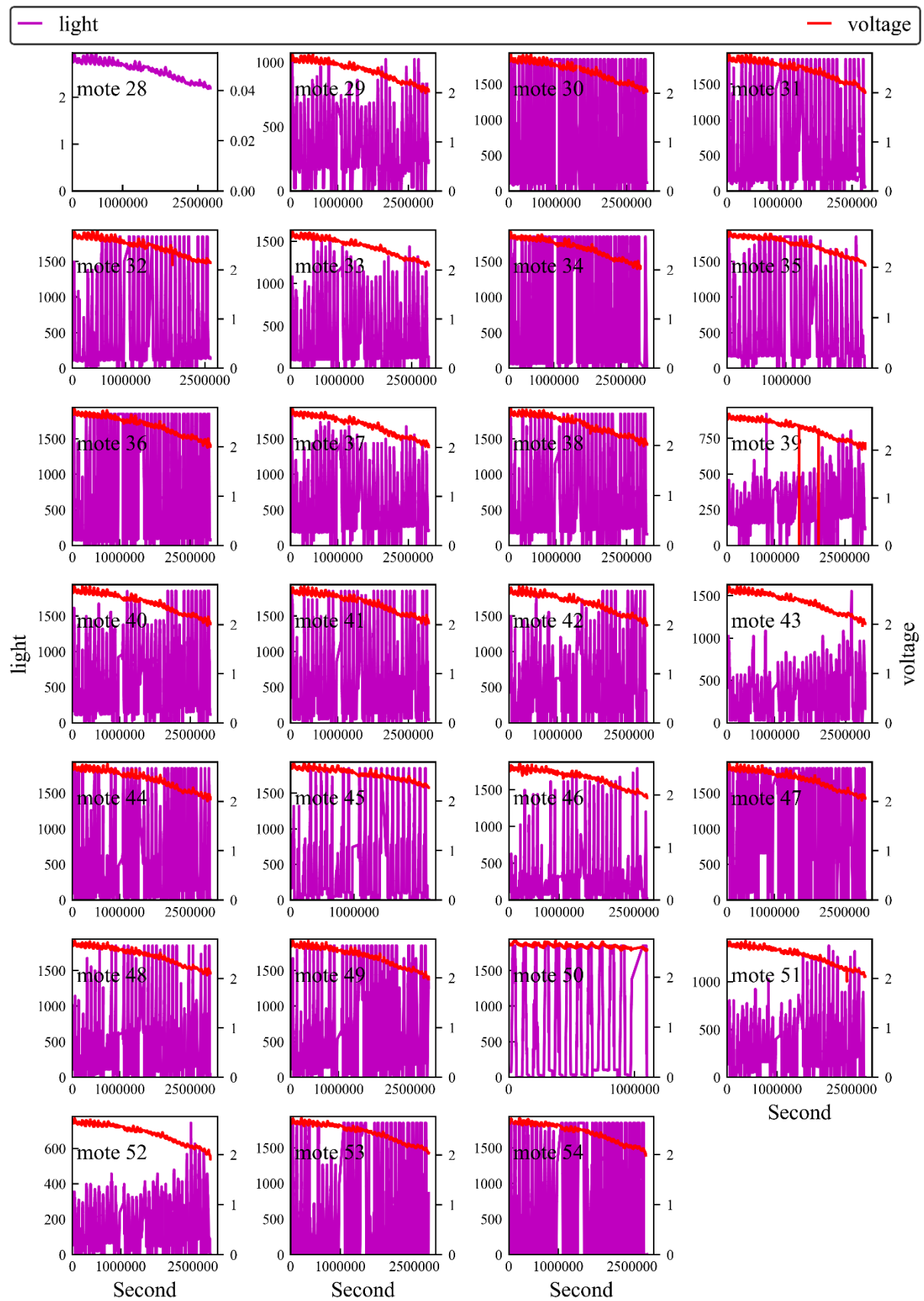
**Figure 1.** Time series of the temperature, humidity, and voltage for each sensor for sensors 1 to 27.



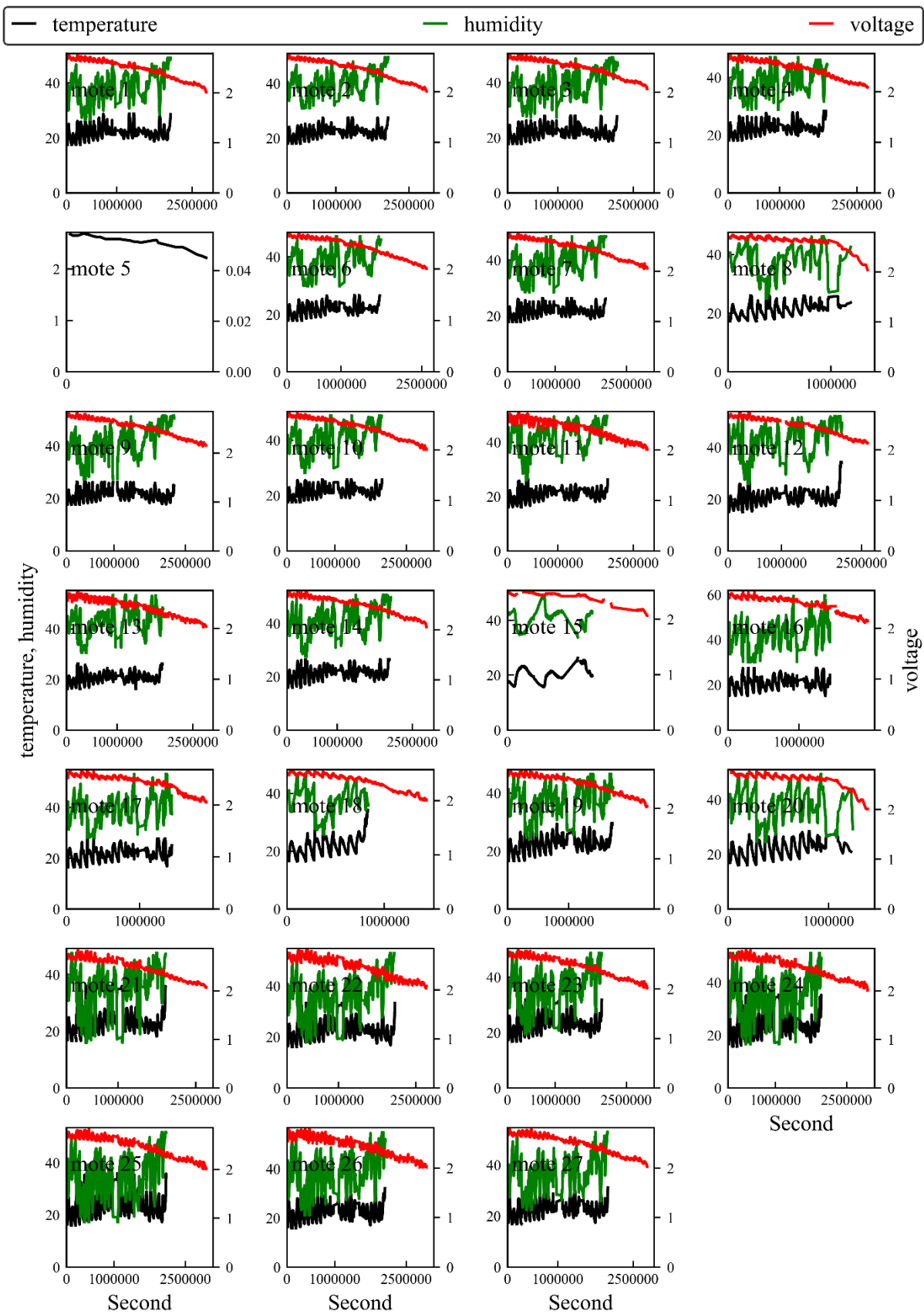
**Figure 2.** Time series of the temperature, humidity, and voltage for each sensor for sensors 28 to 54.



**Figure 3.** Time series of the light and voltage for each sensor for sensors 1 to 27.

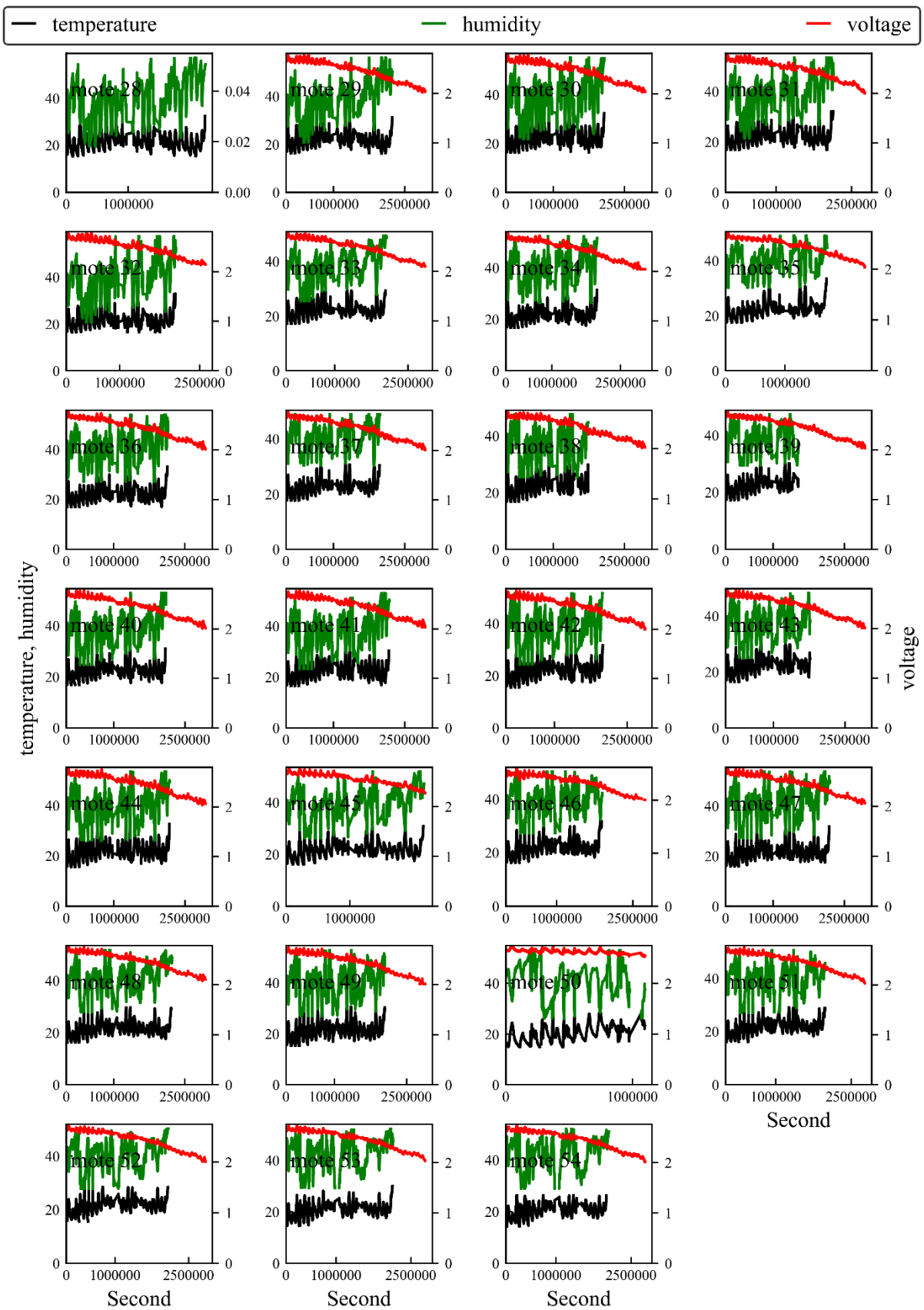


**Figure 4.** Time series of the light and voltage for each sensor for sensors 28 to 54.

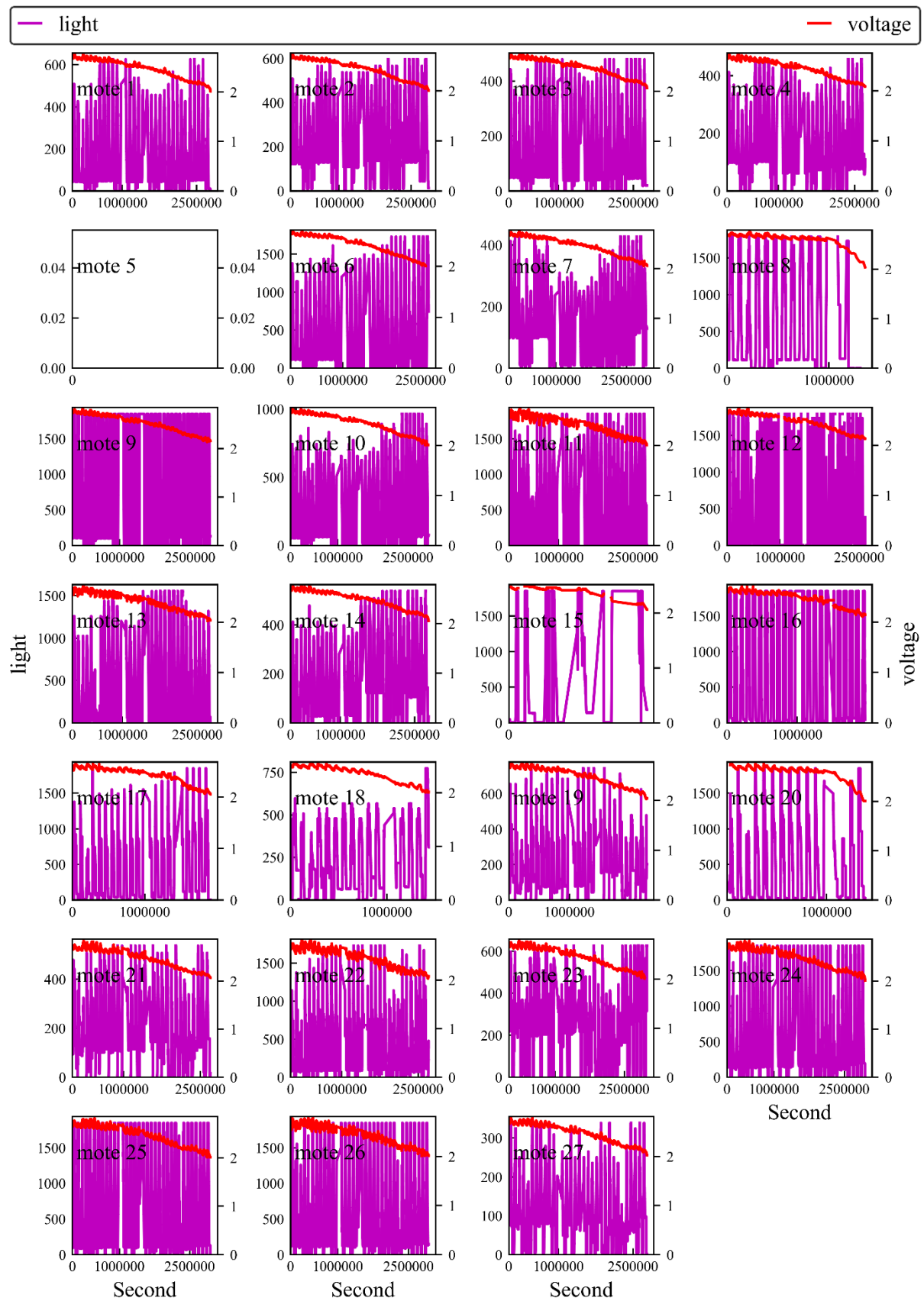


**Figure 5.** Cleaned time series of temperature, humidity, and voltage for each sensor for sensors 1 to 27.



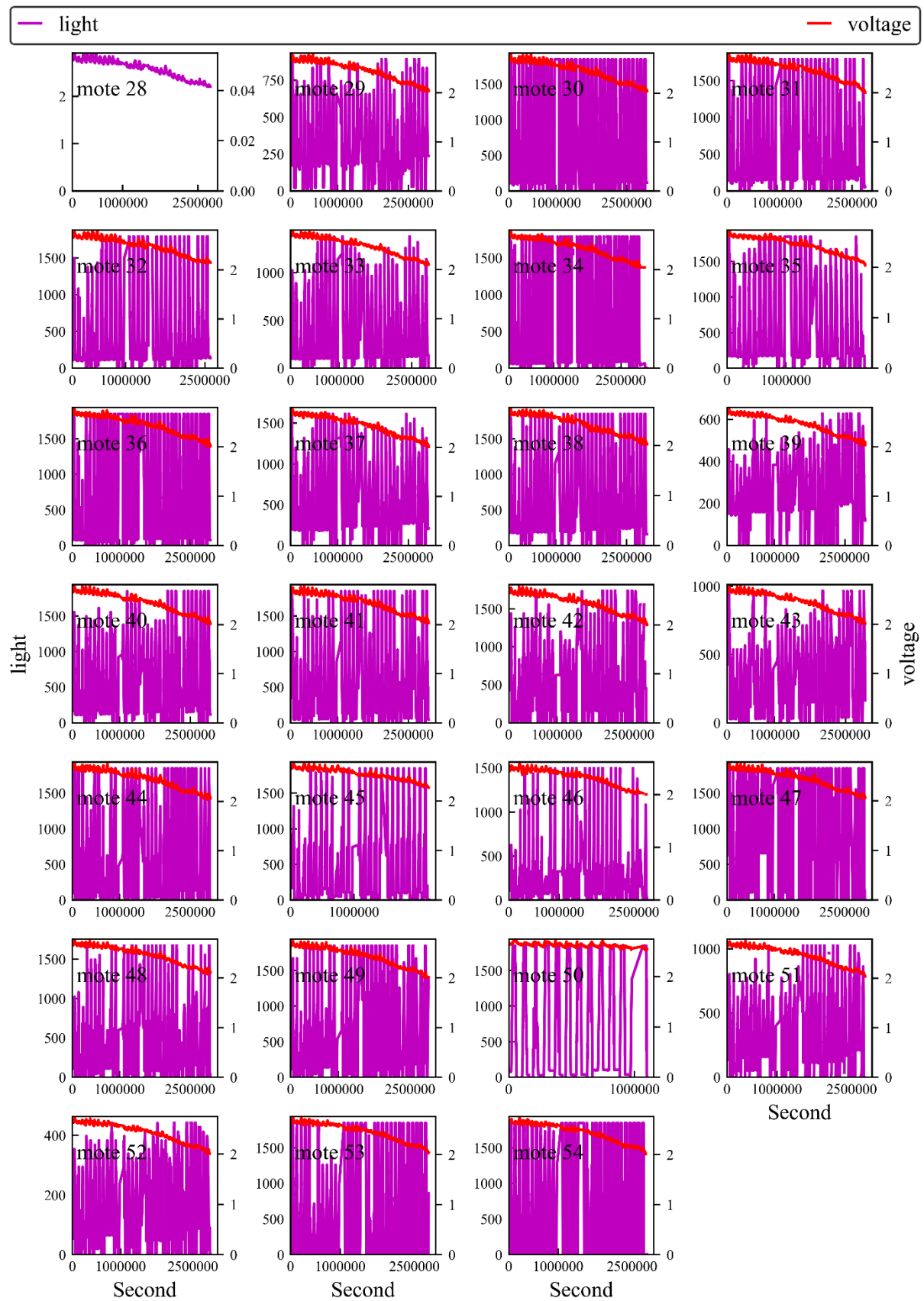


**Figure 6.** Cleaned time series of temperature, humidity, and voltage for each sensor for sensors 28 to 54.

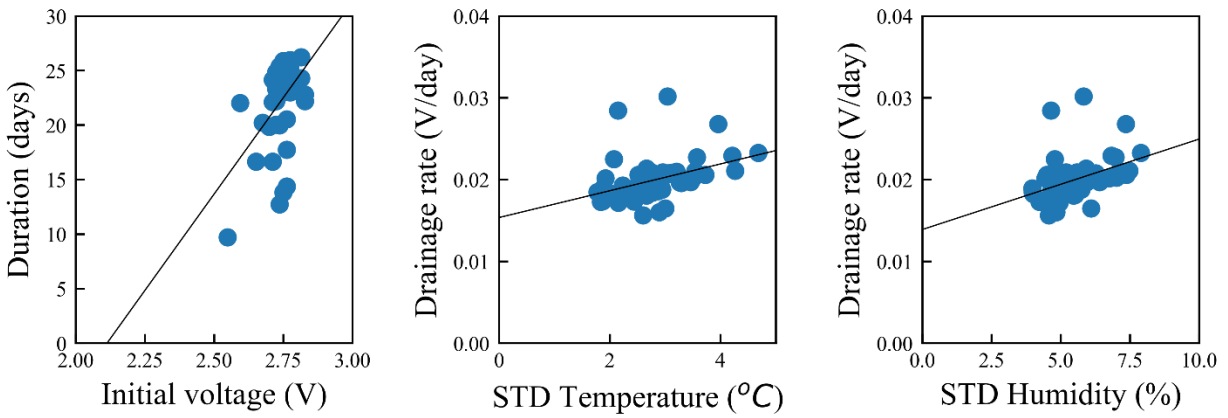


**Figure 7.** Cleaned time series of the light and voltage for each sensor for sensors 1 to 27.

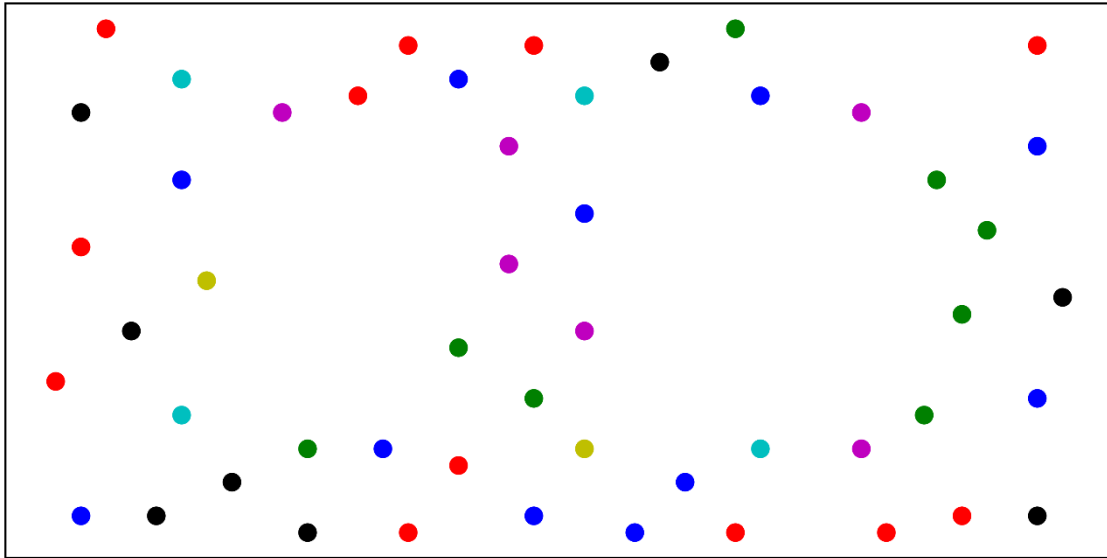




**Figure 8.** Cleaned time series of the light and voltage for each sensor for sensors 28 to 54.



**Figure 9.** Attempt to (Left) estimate the duration of good measurements from initial voltage; (Middle) estimate the drainage rate of battery from standard deviation of temperature; and (Right) same as the (Middle) but from standard deviation of humidity.



**Figure 10.** Clustering results of the sensors. Each color represents a cluster. Each dot represents a sensor in the actual geo-location as shown on the website.