# A Correlation-Driven Approach to IoT Data Filtering for Optimized CEA Monitoring and Control

Tian Yang, Souhaima Stiri, Maria Rita Palattella*
Luxembourg Institute of Science and Technology (LIST), Esch-sur-Alzette, Luxembourg.
Email:name.surname@list.lu

***Abstract*—The control of climate conditions in Controlled Environment Agriculture (CEA) results in increasing use of sensors and Internet of Things (IoT) devices. They generate vast amounts of redundant environmental data, often overwhelming storage, transmission, and decision-making systems. This paper introduces a correlation-driven approach to IoT data filtering which allows detecting meaningful environmental changes and reduces data flow. This translates into saving bandwidth and reducing energy consumption of battery-equipped IoT devices.**

**By applying the statistical process control method CUSUM (Cumulative Sum) to IoT datasets collected in an experimental greenhouse in Luxembourg, this paper focuses on distinguishing drifted data from stable data across specific time periods in different seasons, based on the specific fluctuation patterns exhibited before or after an event occurred in the greenhouse (e.g., opening/closure of vents). With data statistics extracted from the collected IoT dataset - including temperature T, humidity H, among other parameters - our work estimates the change point position and focuses on observing a unique fluctuation pattern within each assigned data observation window, which ensures independent data analysis over the event and avoids inaccurate detections triggered by cross-event observations. Furthermore, with the two most correlated features identified (T, and H), the data filtering decisions are made based on their shared drift positions, thereby enhancing the filtering ratio of redundant data. As a result, the proposed correlation-based CUSUM data filtering scheme increases the filtering ratio from 15.2% for temperature and 52.6% for humidity as individual features to 60.5% for the most correlated feature pair.**

***Index Terms*—Controlled Environment Agriculture, Internet of Things, Edge Computing, Event-Triggered Data Filtering, Drift Detection.**

## I. Introduction

The growing global need for sustainable and resilient food production is driving a shift toward technology-enabled solutions. The population is growing and the global warming is impacting yield productivity and agricultural lands quality. In this context, innovative agricultural systems are gaining attention. Particularly, within Controlled Environment Agriculture (CEA), farmers are able to produce high quality crops all year around, thanks to its resistance to external climate conditions. However, most of CEA systems, lean on a combination of heating, cooling, ventilation and lighting systems to keep the optimal crop growth conditions inside the greenhouse. This makes it a complex system, with multiple inputs, outputs and dependencies that need to be optimally tracked and timely controlled. Thus, the deployment of Internet of Things (IoT) systems in greenhouse environments offers a powerful means to monitor and manage key microclimate parameters; such as temperature, humidity, CO2 concentration, and soil moisture; enabling real-time and data-driven interventions that enhance crop productivity, reduce resource consumption, and improve energy efficiency. Different types of sensors need to be installed in different positions, to cover local changes over the greenhouse surface and subsystems [1].

However, as sensor networks become more pervasive, two major limitations emerge. First, the sheer volume, velocity and variety of data generated by dense sensor deployments may overwhelm communication networks, storage systems, and analytics platforms. Many existing greenhouse IoT systems rely on continuous sensing and periodic reporting, which often leads to redundant or low-value data being transmitted and processed. This results in inefficient use of network bandwidth, increased actuation latency, and higher energy consumption, especially in edge-constrained and battery-powered end devices. Second, current monitoring frameworks frequently lack contextual awareness: they treat all sensor readings equally, with limited capability to distinguish between routine fluctuations and meaningful environmental events that necessitate immediate control action or alert generation. The result is sub-optimal control, unnecessary actuator cycles, and degraded system responsiveness.

To address these challenges, this paper aims to develop an intelligent, event-aware IoT and correlation-driven monitoring framework tailored for greenhouse environments. Our contributions are three-fold: (i) we propose an event correlation methodology to identify relationships among heterogeneous sensor streams and detect meaningful environmental changes; (ii) we design a correlation analysis and drift detection mechanism that filters out low-value or redundant data, both spatial and temporal, thereby optimizing the data flow and preserving only readings likely to impact system operations; and (iii) we evaluate the framework in a Pilot greenhouse, in real-world conditions to quantify the gains in data optimization, communications efficiency, and control responsiveness. By integrating these capabilities, the proposed system supports high-fidelity environmental monitoring while reducing network and

This work is part of the Green ERA-Hub project LIFE, funded by the Fonds National de la Recherche (FNR), Luxembourg.


ISBN: 978-3-903176-79-9

computation burden, thereby advancing the state of the art in smart greenhouse automation.

The work presented in this paper is conducted in the context of the LIFE project [2], funded under the Green Era Hub program. The project leverages on the Digital Twin (DT) technology to closely monitor and optimize different processes in a CEA hydroponic farm, in particular use of water resources, and manure-based nutrient solutions, while aiming to increase crop productivity [1]. In this paper, we focus specifically on data filtering from the environmental monitoring subsystem, using data collected in the greenhouse, from a single IoT device, equipped with multiple sensors. The remainder of this paper is organized as follows: Section II reviews related work on intelligent IoT monitoring and data optimization in greenhouse systems; Section III introduces the overall system architecture and data flow; Section IV details the proposed framework for event correlation and relevance detection; Section V presents the experimental setup and performance evaluation; and finally, Section VI concludes the paper and outlines future research directions.

## II. Related Work

IoT architectures for monitoring and control systems in precision agriculture [3] have been proposed, notably toward the indoor greenhouse scenario [4], [5]. Among them, Ullah et al. [4] studied **learning-based prediction and control** of smart greenhouse, including CO2 generation, cooling/heating, fogging/dehumidification and ventilation operations. As an optimization scheme, ANN-based learning module has been applied based on user defined ranges, to make real-time adjustment to the weights and automatize actuators' operations. IoT-based precision farming in soilless cultivation systems, typically including aeroponic, hydroponic and aquaponic farming has been examined in [5], where pH, electrical conductivity (EC) and water levels were jointly analyzed with the IoT data collected from sensors.

Following the modeling of environmental control systems in time series, **event-driven data filtering and aggregation approaches** have been proposed [7]–[9]. In particular, Zhang et al. [7] investigate event-triggered control and filtering mechanisms for maintaining stability and performance in networked systems; Lu et al. [8] focuses on event-driven data collection and aggregation along a fixed topology from sensors to the sink, optimizing energy consumption and transmission efficiency through spatio-temporal correlation analysis. Gkoulis et al. proposed a symmetrical architecture for event stream processing [9] in an IoT-enabled application context. Based on an IoT-enabled greenhouse with a roof vent, data is collected from six sensors under three levels of variation. A Complex Event Processing (CEP) model is then used to analyze how the sensor data changes over time within defined time windows. However, although the hop size and window size are predetermined according to the type of IoT event, the proposed scheme [9] still fully relies on event timestamps to identify triggering behaviors, without defining specific thresholds or decision-making criteria for each feature as done in [4]. Since no automated approach is introduced in [9] for events such as opening or closing a roof vent, the role of CEP remains limited to recognizing events and extracting business insights, making it subject to significant delays caused by the greenhouse's physical response time to the event.

In particular, S. Xia et al. [11] proposed a hierarchical data fusion scheme for intelligent greenhouses that integrates underlying sensor data with varying processing and transmission time. The study compared batch, sequential, and parallel fusion frameworks, showing that the proposed parallel fusion approach improved computational efficiency while maintaining high fusion accuracy.

In this paper, cross-feature redundancy is analyzed to further improve the data reduction ratio using the statistical control algorithm CUSUM (Cumulative Sum) [12], [13]. Unlike window-based change detectors such as ADWIN (Adaptive Windowing) that reports discrete change timestamps, CUSUM evaluates deviations at each timestamp and is therefore suitable for sample-based dynamic drift detection in time series. Hence, within the scope of this work, CUSUM is adopted as the main drift detection method. Furthermore, sample-to-sample drift detection is extended to focused, event-triggered change point analysis of time series to further improve the performance. With appropriate parameters (e.g., drift parameter, window size, warm-up onset and duration, etc.) being defined, these methods allow to identify the occurrence of highly-fluctuated or redundant data in IoT dataset.

Under the framework of event-driven detection, the subsequent step refers to accurately **identifying data redundancy** over time through distinct fluctuation patterns to further reduce the data at the feature level without losing data relevance. Typically, redundant IoT data arises from high correlations and accumulates either across multiple features or over time within a single feature.

Thus, it is necessary to define tailored detecting thresholds for each feature to characterize their varying fluctuation scales (i.e., to capture and reduce correct amount of redundant data) and to determine an appropriate window size that enables an in-depth statistical analysis of the time-series data.

For window-based detection methods, shorter window size and longer training / warm-up period will help in improving the detection accuracy [13]. Greenhouse events, such as the operation of vents or fans, can affect data statistics and fluctuation patterns, leading to a migration of the window's starting time as determined by the control strategy. In this paper, we investigate event-driven changes in data statistics, taking into account both fluctuation scales and observation durations, to detect drifts and optimize data-filtering decisions.

## III. System Architecture and Data Flow

In this work, we consider a closed-loop hydroponic farming system as CEA. The system is installed in an 8mx26m greenhouse (Fig. 1), located in Greiveldange, in Luxembourg, and managed by Fesch Haff S.A.R.L. Different sensors are installed in different positions to monitor environmental conditions, irrigation systems and crop growth. The focus of

Fig. 1: Milesight EM500 IoT device (left). Pilot greenhouse from Fesch Haff S.A.R.L.: Greiveldange, Luxembourg (right)

this paper is on environment monitoring using a single IoT device, measuring different features in a single fixed position. Multi-sensor IoT devices are installed within a LoRaWAN communication infrastructure and edge computing techniques to collect data and trigger actions, whenever needed.

### A. IoT Monitoring System

For the environmental monitoring system, we use as IoT devices eight Milesight EM500, measuring Temperature T, Humidity H, CO2 and atmospheric Pressure P. Four devices are positioned near the greenhouse entrance, (presented in Fig. 1 with yellow squares) where environmental conditions are expected to be more affected by door opening. In addition, four other devices are deployed toward the rear of the greenhouse. All devices are installed in a fixed position: two near the plant, two in the roof of the greenhouse (both on the front and on the back) to capture microclimate variations inside the greenhouse. In addition, further sensors are expected to be installed in other plant rows, depending on the crop type, to monitor microclimate and transpiration of specific plants. This progressive densification of the sensing infrastructure will significantly increase the volume of data generated and transmitted through the network, thereby impacting overall bandwidth usage and system scalability.

For this reason, the selection of an appropriate sampling frequency becomes a critical design parameter. After sensor calibration, and data analysis in different seasons and operational conditions, a sampling frequency of 10 minutes is selected to reflect near-real time fluctuation of data and ensure a balance between measurement accuracy and system scalability. Data from the different sensors is sampled at an interval of 10 minutes and sent to a LoRaWAN gateway installed inside the greenhouse, and connected to a Chirpstack server via cellular backhauling network. HEX formatted frames are received by Chirpstack server and transformed into numerical data using the decoder implemented in the server. The decoded data is then stored in a local database for analysis and visualization.

To illustrate the amount of data sent by a single IoT device, we refer to the payload decoder of the considered device [14]. The payload generated by the 4 sensors (T, H, P CO2) readings is encoded in 15 Bytes sent 6 times per hour. This results in approximately 2.1 kB of payload data per device per day, in addition to protocol overhead and retransmissions. Although this volume looks modest for a single device, it scales linearly with the number of devices deployed and the accumulation of data over the long operational periods of greenhouses.

### B. Data Flow Management and Data Redundancy

The management of data flow refers to optimizing the IoT data of high relevance to be sent in each transmission based on predefined detecting thresholds, as well as the onset of reporting interval. Important changes should not be missed when collecting data, but also redundant transmissions should be avoided. To define a dynamic time interval and frame size, a large data set was collected from the operating greenhouse over the year, covering different seasons and activities. In the collected dataset, varying fluctuation patterns such as frequency and change scale are observed on each feature.

Even at a relatively low sampling frequency of one sample every 10 minutes, redundancy still accumulates because the observation window spans weeks or months, while greenhouse IoT data typically evolves through small-scale changes. To reduce the transmission of data flow, we identify two types of data redundancy. The first is *spatial redundancy*, which refers to high positive or negative correlation among features at a given timestamp, allowing to indicate feature-level similarity or duplication, and consequently to reduce the frame size. The second is *temporal redundancy*, which captures data values remaining unchanged over long time for a single feature or sensor, allowing to adjust transmission intervals. Both types of redundancy can lead to unnecessary consumption of bandwidth and end device's battery.

### C. Greenhouse Event Analytics

Before starting the analysis of collected data, it is important to have a comprehensive awareness and understanding of the events happening in the physical world, i.e. the greenhouse. This helps interpreting data variations and identifying relevant change points that may influence data flow and system operations.

The considered greenhouse incorporates some automated processes to guarantee some core functionalities of the CEA and avoid critical conditions that may impact the crops' health. For instance, the control of the environmental conditions inside the greenhouse requires maintaining indoor temperature within a defined threshold range. For this, an automatic opening of the vents is triggered once the greenhouse temperature is higher than 21°C, and closed again once the temperature drops below 18°C, to preserve the desired microclimate. This control action, opening of the greenhouse vents, introduces an abnormal change of data trend, mainly a noticeable shift of temperature and humidity patterns, that needs to be tracked more precisely.

Similarly, fluctuations in CO2 concentration can serve as indicators of external events. Fig.2 shows an example of the impact of human presence in the greenhouse, on the fluctuations of measured CO2 levels. In normal days without human presence, CO2 concentration follows a consistent daily pattern:

the CO2 concentration increases by night and decreases during the day (yellow curve). While, with the presence of humans in the greenhouse, a distinct anomalous peak of CO2 is detected, impacting the CO2 variation trend (blue curve). These peaks are directly attributable to human respiration and clearly demonstrate how CO2 measurements can capture short-term events in addition to longer-term environmental dynamics.

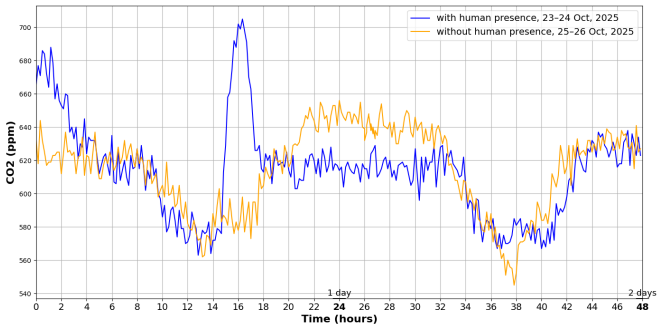



Fig. 2: Evolution of CO2 concentration in the greenhouse over 48 hours with/without the presence of humans.

Beyond optimizing data transmission and storage, the detection of data drifts and correlations in edge devices contributes to system robustness and security. Timely identification of such events enables the generation of alerts and adaptive feedback mechanisms, fostering a more resilient and intelligent greenhouse monitoring and control framework.

## IV. Event Correlation and Relevance Detection Framework

### A. Correlation Analysis among Features in IoT Dataset

To achieve common drift detection by jointly analyzing the fluctuation patterns of individual features, correlation analysis is applied across features. This analysis enables accurate identification of shared environmental changes over time and improves overall data-filtering efficiency. Highly-correlated features can effectively represent main shared variations in the greenhouse environment, allowing the system to disregard and eventually discard occasional fluctuations in individual features. For this purpose, a two-step analysis of data correlation is presented in this section to identify the potentials of lowering spatial redundancy.

The first step of characterizing data redundancy is **to calculate the correlation coefficient of two specific features** (i.e., a feature pair) out of N from the dataset. The higher the correlation, the higher the spatial redundancy it reflects. Then, through the observation of shared temporal patterns identified by overlapped drifts of correlated features, we identify the opportunities to reduce redundant data on the time series.

The second step is to **identify the highest correlation value among all paired combinations of features**: T–H, T–P, H–P, CO2–H, CO2–T, and CO2–P. As shown in Fig. 3, we plotted all data points collected during a selected period. In the example of the T–H feature pair, temperature tends to be low when humidity is high, this indicates that humidity

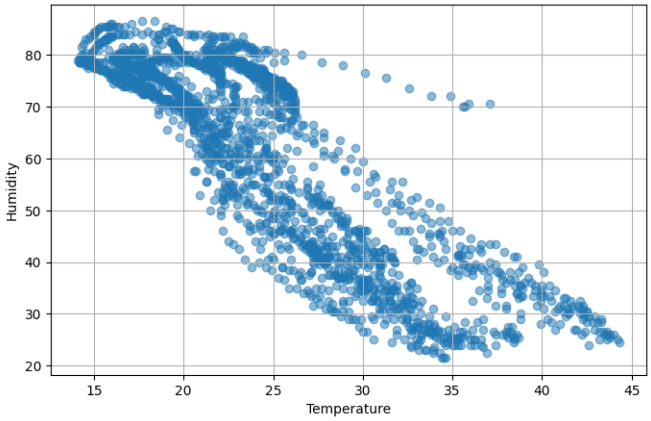



Fig. 3: High negative correlation score = -0.8472 for T–H as most-correlated feature pair, indicating intrinsic, strong correlation among selected data features in the greenhouse.

may increase when temperature drops. Beside a few discrete data points, most of collected IoT samples follow this strong negative correlation.

The absolute value of the correlation coefficient indicates the strength of the linear relationship between two variables, reflecting how consistently the data trend follows a linear pattern. A higher coefficient makes it more rational to identify key, representative environmental changes through common drift detections of highly correlated feature pairs. After comparison, the correlation coefficient for the temperature–humidity pair (Fig. 3) is identified as the highest among six combinations.

### B. Relevance Assessment: Filtering and Feature Extraction

The correlation coefficient is computed per pair of two features, reflecting common data trends over time across the $\binom{N}{2}$ possible feature combinations (**T-H: -0.8472**; CO2-H: 0.7779; CO2-T: -0.6643; P-T: -0.0784; P-H:-0.1198; CO2-P: 0.1095). The coefficient further compares the extent of feature relevance using its absolute value, quantifying both positive and negative correlations. Based on the data relevance among features and the quality of each individual feature, decisions are made to precisely capture data changes over time and identify the most relevant and representative features that require close monitoring.

Once the most correlated pair is identified, it becomes easier to pinpoint the data points where notable changes occur (i.e., the concept drifts), given the pre-defined fluctuation patterns (e.g., the scale of drifts, positive or negative drifts, and frequency of occurrence). As the most-correlated pair among all the features, the common drifts highlight the most significant changes despite high data redundancy. **Therefore, detecting common drifts in most-correlated feature pair helps in solving the trade-off between optimal data reduction and accurate representation of relevant environmental changes.**

### C. Integration of Event-based Control Logic

As shown in Fig. 4, the traditional, non-event-based statistical control process involves data reading from the IoT dataset,

extracting fluctuation patterns, determining an appropriate drift parameter $\delta$, and outputting the detected drifts accordingly. In particular, a warm-up process is performed at the beginning of time window to extract the main data statistics according to historic data, and make drift decisions for the rest of dataset. As compared to traditional statistical control logic, an event-driven statistical control scheme is illustrated on the right side of Fig. 4. In this scheme, the change point is estimated to separate data trends when varied fluctuation patterns are detected within a single window. Starting from the identified change point, the warm-up period is reset to locally capture more realistic patterns with an updated drift parameter. This approach enables the detection of newly emerging local fluctuation patterns following the occurrence of an event, ensuring high accuracy of drift detection based on repositioning the warm-up phase.

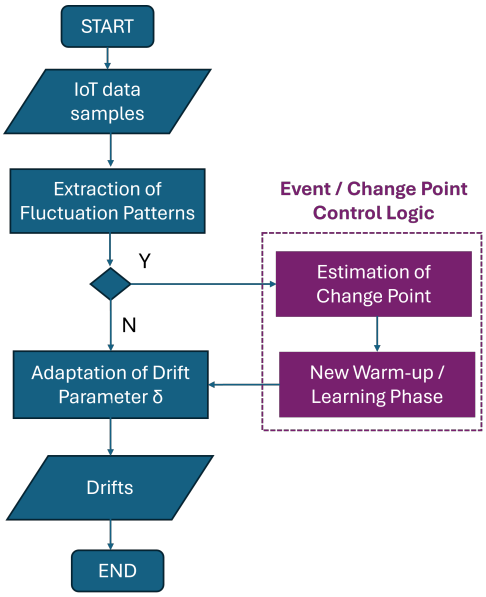



Fig. 4: Event-based control logic with extensive warm-up or training phase from the estimated position of change point.

## V. Experimental Setup and Evaluation

### A. Description of simulation environment

The IoT data in our test was collected from the pilot greenhouse with an average sampling frequency of one measurement every 10 minutes. For the statistical analysis, we selected data from two periods: 17–30 March 2025 and 30 June–13 July 2025; representing winter conditions where the greenhouse is most of the time closed to keep indoor warmth, and summer period where the vents might be opened for longer periods if the indoor temperature goes above the thresholds, resp. The objective is to compare the impact of distinct fluctuation patterns on drift detection decisions in time series, in different weather conditions. The performance test was conducted under Python version 3.9.13.

### B. Data collection, drift detection and training process

The raw IoT data is collected across different seasons of the year, with varying fluctuation patterns on each feature and depicted in different units (e.g., the unit of data fluctuation is quantified as 0.1 degree for temperature and 0.5 per cent for humidity). Based on identified fluctuation patterns, we detect drifts for each individual feature. To determine drifted data and the redundant normal data to eliminate, the common drifted data captured in the most correlated feature pairs are considered as the indicators of key environmental changes in the greenhouse. Finally, the positions of shared drifted data samples are recorded as a main outcome after data collection and drift detection process.

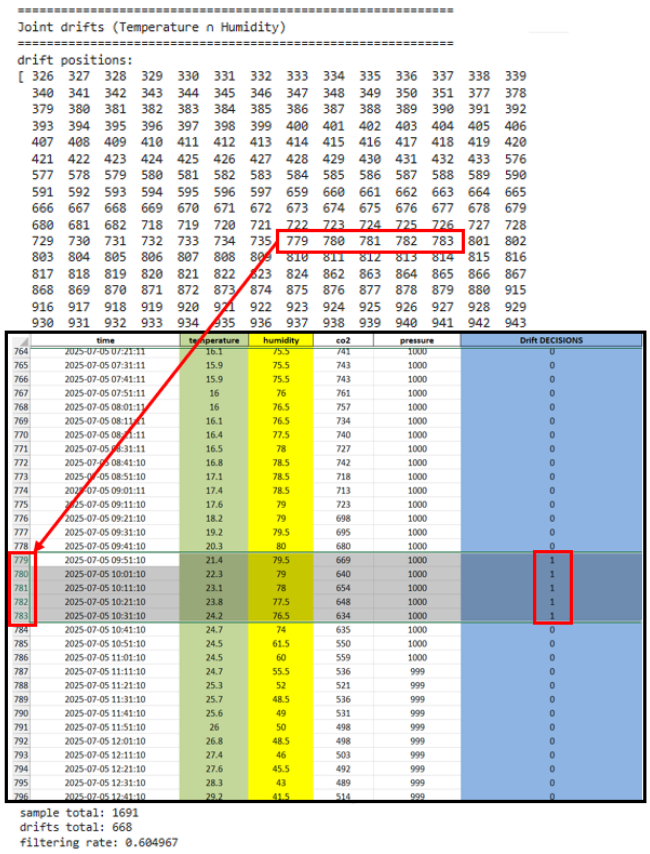



Fig. 5: Integrating detected positions of common drifts into the corresponding rows of raw IoT dataset for training.

Following the acquisition of positions of common drift decisions, we mark these positions back into the corresponding data records in raw dataset, as shown in Fig. 5. When reading data from one record to the next, this helps further analyze how changes in the data values of different features can lead to the emergence or disappearance of common drifts (i.e., drift decisions in the last column). Each change of four individual features in Fig. 5 can affect the decision-making on data filtering.

### C. Performance Metrics in Event-Triggered Detections

It is worth noting that the drift decisions in Fig. 5 are obtained under thresholding criteria for fluctuation patterns, drift parameter, window size, starting time and duration of warm-up period. Depending on the data quality denoted by data statistics extracted during the warm up, drift thresholds are defined to capture fluctuated data of corresponding scales.

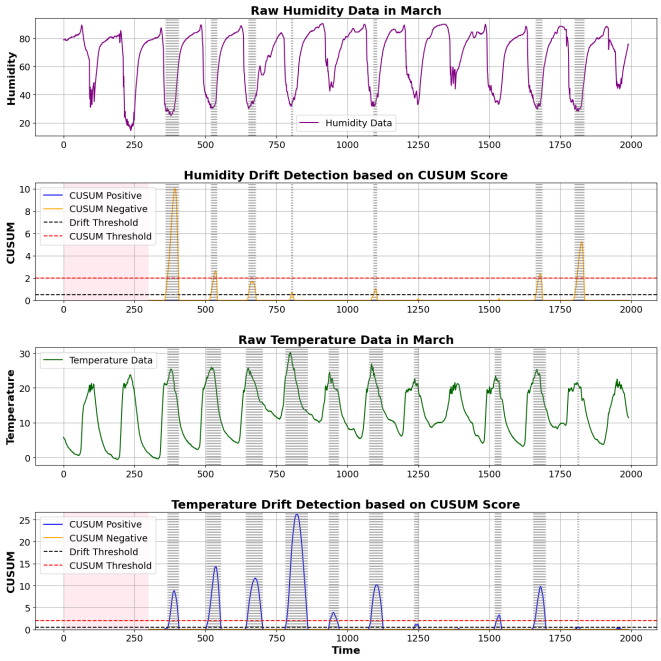



Fig. 6: $\delta$=1.5: capture only medium/large-scale drifts in March (moderate fluctuations)

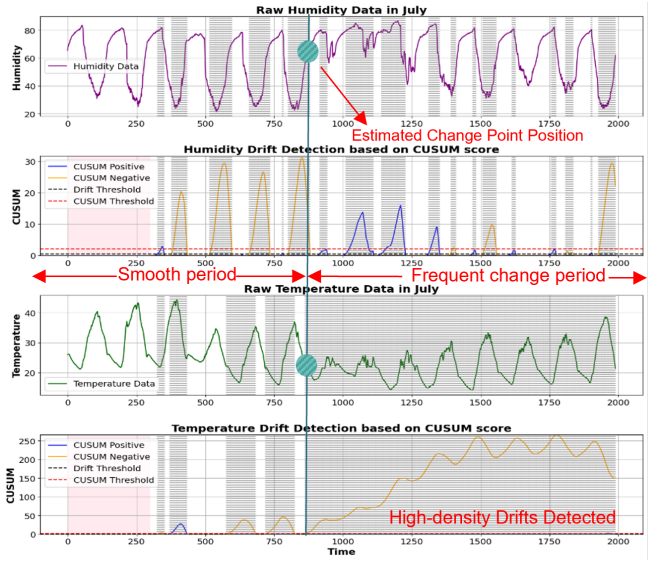



Fig. 7: $\delta$ =1.0 to capture small-scale drifts in July, where a change point is added to distinguish smooth/fluctuated periods.

As an important objective of our test, we analyze data changes over different periods/seasons, and define appropriate drift parameter $\delta$ accordingly. As can be observed in Fig. 6, both temperature and humidity data exhibit relatively moderate data fluctuations in March: temperature is higher than the threshold (T>21°C) for shorter periods, thus vents are opened for a short time. Thereby, to allow higher tolerance on fluctuated data while only capturing large-scale changes, we enlarge $\delta$ to 1.5. This setting helps in avoiding miss-detection during moderate period and ensures a high filtering ratio.

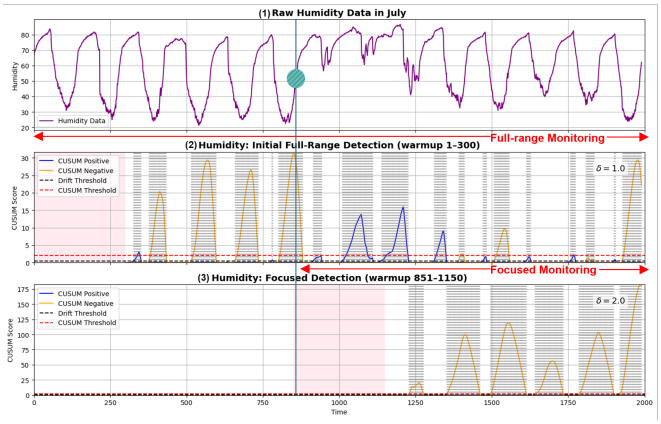
In Fig. 6 and Fig. 7, the gray vertical slicing reflects the detected drifts positions. Before using $\delta$ to control the sensitivity to fluctuations, a warm-up window of 300 samples is defined to extract baseline data statistics. Along the vertical axis of Fig. 6, the CUSUM threshold is set to 2 (red dashed line) to capture the peak changes, while a relaxed drift capture threshold is set to 0.5 (black dashed line) to capture additional data changes around those peaks. Consequently, the algorithm captures the blue and yellow lines, indicating positive and negative drifts, respectively.

Similarly, it can be observed in Fig. 7 that more frequent drifts have occurred, notably during the second half of the monitoring period. Even if $\delta$ has been lowered to 1.0 to be more sensitive for capturing small-scale drifts, the temperature data still exhibits strong fluctuations during the second half which has not emerged during the first half. When analyzing the temperature changes, it can be seen that the first half of the monitoring period is very hot, the greenhouse temperature was higher than the threshold for closing the vents, thus they were opened almost all the time. While in the second half, temperature dropped and started fluctuating around higher and lower thresholds, resulting in more frequent opening and closing of the vents. Therefore, a single observation window with a single warm up period will hardly be sufficient to detect distinct fluctuation patterns before or after the change point. Also, using a single drift parameter $\delta$ risks causing missed detections during sparse periods and false alarms during periods of intense change. Thus, identifying the change point corresponding to a greenhouse event becomes necessary.

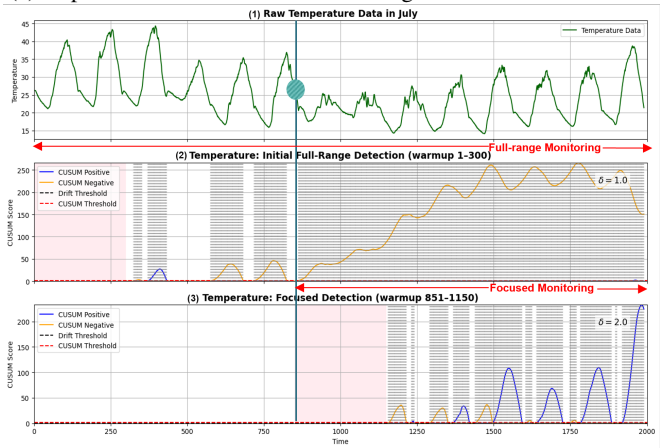
To perform correlation analysis and drift detection in time series, it is crucial to distinguish the distinct fluctuation patterns of the selected feature with the estimated change point in the dataset. However, multiple features may exhibit different change points even when triggered by a same event. Thus, between the two most-correlated features, it is necessary to estimate a common change point to determine when to start the joint drift detection window and when to re-initialize the warm-up process to relearn the newly-emerged data patterns.

### D. Final Results

In Fig. 8, following the distinct fluctuation patterns in July's dataset triggered by an environmental event (temperature drop), the CUSUM function was re-initialized from the estimated common change point, thus enabling focused monitoring toward the frequent change period after the 851st sample. Two main changes have been made compared to Fig. 7: (1) the starting point of warm up period is moved from 1 to 851 to allow normalized, adaptive drift detections




(a) Repair underestimated drifts resulting from missed detections.

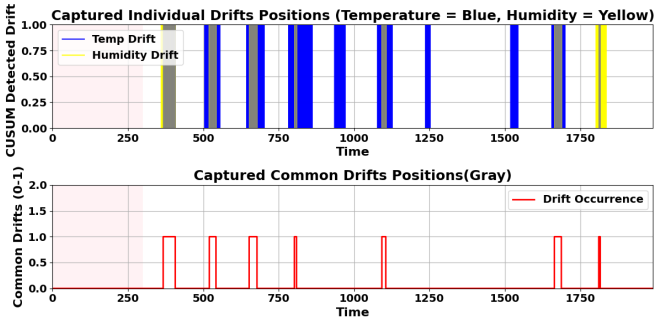



(b) Normalize & reduce redundant drifts resulting from false alarms.

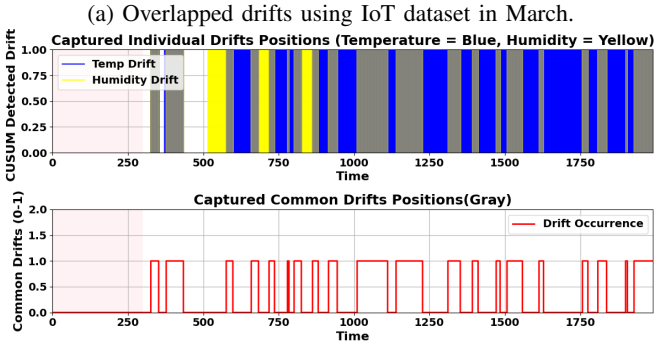
Fig. 8: The migration of starting point of warm up period (1 to 851) to normalize drift detections during focused period.

after the event; (2) as recent data has been used in warm up to relearn from window 851-1150, the drift parameter $\delta$ is relaxed from 1.0 to 2.0 to allow a higher tolerance during the focused monitoring period from the change point. Consequently, we can observe in Fig. 8 (a-3) and Fig. 8 (b-3) that the event-triggered detection enables normalization of captured changes on both features during the focused time period, providing moderate drift decisions that avoid miss detection of humidity in Fig. 8 (a-2) and false alarm of temperature in Fig. 8 (b-2) resulting from full-range monitoring. It can also be observed that with drift parameter $\delta$ being relaxed to lower the sensitivity of drift detection, the data filtering ratio marked by the black spaces in Fig. 8 (b-3) is much higher than in Fig. 8 (b-2), which is very close to 0.

In the next step, following the training process shown in Fig. 5, Fig. 9 shows the identification of common drifts from the overlapped positions of most correlated feature pair: temperature and humidity, in the selected periods of winter and summer. The occurrence of common drifts relies on predetermined drift detection criteria of individual features. In the upper part of the figures, the presence of yellow blocks represents humidity drifts, and blue blocks represent




(a) Overlapped drifts using IoT dataset in March.




(b) Overlapped drifts using IoT dataset in July.

Fig. 9: Detection of joint drifts from H-T feature pair

temperature drifts, respectively. In the lower part of figures, the occurrence of common drifts are marked as 1 based on the overlapped positions of both features, corresponding to detected common parts in gray. Based on the data statistics in July using this detect-and-filter approach, among 1691 testing samples, 1434 temperature drifts are detected, with a filtering rate at only 15.2%; 802 humidity drifts are detected, with a filtering rate at 52.6%. This reflects that temperature fluctuates more frequently than humidity, but these fluctuations may be ordinary and contain redundancy since the detected drifts are exclusively temperature-based, and the data is barely filtered without using correlation-based common drift detection. In contrast, 668 common drifts are detected, leading to a higher filtering rate at 60.5%. This indicates that common drift detection can filter out normal data or individual changes to the highest extent, while only keeping the shared drifts, allowing deeper data reduction while maintaining the relevant data. Moreover, in winter period, when the greenhouse is kept closed most of the time to keep the warmth inside, less common drifts are noticed, only 136 drifts are detected out of 1691 sample. Resulting in a higher filtering ratio of 91.9%.

## VI. Discussion and Future Work

In the context of CEA, our paper discussed the challenges in identifying appropriate change point to allow seamless, highly-accurate event-triggered drift detections even under varying fluctuation patterns. This approach contributes to a higher filtering ratio, notably among the highly-correlated features.

In this work, we mainly focus on drift detections on the data collected from a single IoT device, during stable periods. However, distinct fluctuations may occur, in different locations, due

to uneven sunlight exposure, human activities, or the operation of vents. These events may require parameter adaptations on certain devices to correctly capture drifts and accurately reflect data change patterns. Therefore, it is crucial to monitor drifts in coordination with the other IoT devices deployed in the greenhouse, which helps identify uneven or localized fluctuations within the entire greenhouse environment. In the time series, by fine-tuning the parameters, identifying and repositioning change points, our method successfully avoids false alarms that typically occur in full-range, non-focused monitoring and detection, thereby further enhancing the data filtering efficiency.

From a network performance perspective, the proposed approach focuses on filtering out feature specific drifts while keeping only common drifts that correspond to meaningful environmental changes. This strategy significantly reduces the size of transmitted payloads. In particular, when no common drifts are detected, the payload is reduced by approximately 50%, which directly translates into shorter packet time-on-air and, consequently, a lower probability of collisions when multiple devices transmit simultaneously using the same spreading factor. This reduction in transmitted data also leads to more efficient bandwidth utilization and contributes to extending the battery lifetime of the devices. Over the evaluated periods, the proposed method resulted in the transmission of only 965 bytes of data during the two weeks of March, compared to 11.8 kB with the baseline approach. Similarly, during the two weeks of July, 4.7 kB were transmitted instead of 11.8 kB. These results highlight the effectiveness of the proposed approach in reducing communication overhead while preserving relevant environmental information from a single device, and motivate future implementations of these on-line filtering methodologies in edge IoT devices.

In future work, it is also essential to automate this process over a wider range of time. To achieve this, not only should the thresholds be dynamically adjusted according to the incoming data, the estimation of change point positions also needs to be refined and validated through a feedback loop, rather than relying on experimental selections.

## VII. Conclusion

In greenhouse-based CEA systems, there is a growing need for deploying multiple types of IoT devices to enable full-range monitoring and control, ranging from surrounding climate conditions to crop-level sensing. However, the continuous data transmission translates in battery depletion for the IoT devices, in addition to waste of network bandwidth, when compiting with parallel data streams (e.g., from cameras, etc.). To address this challenge, this paper proposes an efficient, event-triggered data-filtering scheme that enables prompt recognition of actionable insights while minimizing redundant data among correlated features. Through correlation-driven statistical process control, the detected common drifts allow to reflect key representative environmental changes in the greenhouse. By identifying distinct fluctuation patterns before and after the event, our work contributes to the extraction of highly-relevant data features over time, while enabling further data reduction. Simulation results show that the total data filtering ratio approaches 60.5%, while the missed detection and false alarm ratios are considerably decreased.

## Acknowledgment

This work was conducted as part of the LIFE project, carried out under the Green ERA-Hub, a Coordination and Support Action (CSA), funded through the European Union's Horizon Europe research and innovation (R&I) programme under Grant Agreement No. 101056828. The work was funded by FNR under INTER/GREEN23/18401855/LIFE. The authors would like to thank Manuel Arrillaga from Fësch Haff SARL, for supporting our measurements campaign and testing in the hydroponic farm.

## References

[1] S. Stiri and M. R. Palattella, "Digital Twins for Hydroponic Farms Management: Communication Needs and Challenges," 2025 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 2025, pp. 781-787.

[2] GEH LIFE Project website: https://www.geh-life.com/, visited in November 2025

[3] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758–3773, 2018.

[4] I. Ullah, M. Fayaz, M. Aman, and D. Kim, "Toward autonomous farming—A novel scheme based on learning to prediction and optimization for smart greenhouse environment control," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25300–25323, 2022.

[5] M. Dutta, D. Gupta, S. Tharewal, D. Goyal, J. K. Sandhu, M. Kaur, and J. M. Alanazi, "Internet of Things-based smart precision farming in soilless agriculture: Opportunities and challenges for global food security," *IEEE Access*, 2025.

[6] M. F. Almufareh, M. Humayun, Z. Ahmad, and A. Khan, "An intelligent LoRaWAN-based IoT device for monitoring and control solutions in smart farming through anomaly detection integrated with unsupervised machine learning," *IEEE Access*, vol. 12, pp. 119072–119086, 2024.

[7] X. M. Zhang, Q. L. Han, and B. L. Zhang, "An overview and deep investigation on sampled-data-based event-triggered control and filtering for networked systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 4–16, 2016.

[8] Y. Lu, H. Jiang, Z. Pang, Z. Wang, J. Xu, Y. Liu, and H. Sun, "Data collection study based on spatio-temporal correlation in event-driven sensor networks," *IEEE Access*, vol. 7, pp. 175857–175864, 2019.

[9] D. Gkoulis, C. Bardaki, G. Kousiouris, and M. Nikolaidou, "Transforming IoT events to meaningful business events on the edge: Implementation for smart farming application," *Future Internet*, vol. 15, no. 4, p. 135, 2023.

[10] F. Mahmood, R. Govindan, A. Bermak, D. Yang, and T. Al-Ansari, "Data-driven robust model predictive control for greenhouse temperature control and energy utilisation assessment," *Applied Energy*, vol. 343, p. 121190, 2023.

[11] S. Xia, X. Nan, X. Cai, and X. Lu, "Data fusion based wireless temperature monitoring system applied to intelligent greenhouse," *Computers and Electronics in Agriculture*, vol. 192, p. 106576, 2022.

[12] T. Yang, M. F. Awan, A. Jakovljevic, D. Drajic, M. Minardi, C. Politis, M. Afhamisis, and M. R. Palattella, "Edge Computing to Enable Front-End Data Processing Capabilities in Smart Agriculture and Forestry," in Proc. 8th Conf. Cloud and Internet of Things (CIoT'25), London, UK, 2025.

[13] P. Q. F. Yi, "An adaptive CUSUM chart for drift detection," Quality and Reliability Engineering International, vol. 38, no. 2, pp. 887-894, 2022.

[14] Milesight-IoT EM500 Sensor Decoder: https://github.com/Milesight-IoT/SensorDecoders/tree/main/em-series/em500-co2

[15] L. Yang, D. M. Manias and A. Shami, "PWPAE: An Ensemble Framework for Concept Drift Adaptation in IoT Data Streams," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, pp. 01-06, 2021.