New Method of Analysis of Plant Data

February 5, 2014

PreFEED Corporation

Yoshio Kumagae

Introduction

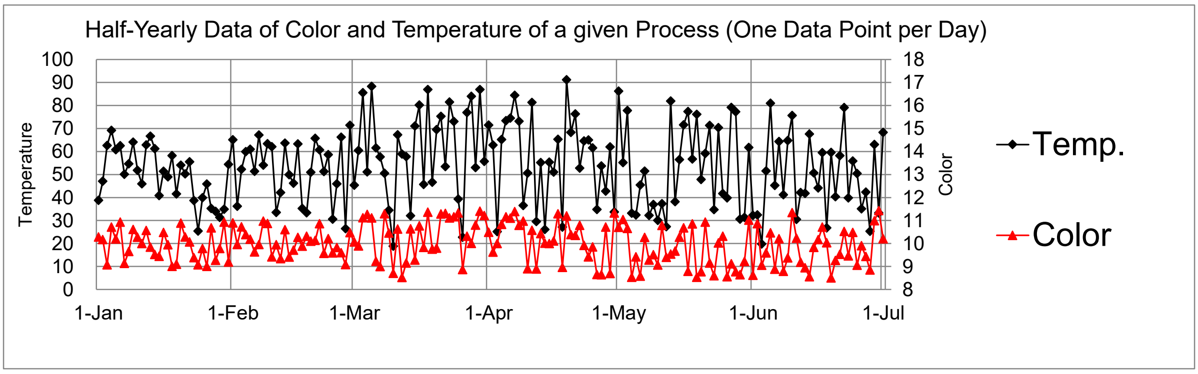
Methods such as modeling and multivariate analysis are widely used to analyze experimental data and plant data. However, within the large amounts of data collected by DCS and the like, temporal variation factors also increase, and it is often impossible to clearly identify correlations due to the influence of many factors.
In such cases, it is often useful to index the data to several levels in terms of "abnormal or normal" and the like, and to examine the relationships between the indexes. This method (discretized data analysis) is introduced here.

This method is protected by the following patent.
System for identifying cause of abnormality, method of identifying cause of abnormality, and program for identifying cause of abnormality
United States Patent 8972222
Inventors: Kumagae, Yoshio (Tokyo, JP)
Assignee: Prefeed Corporation (Tokyo, JP)

Application Example

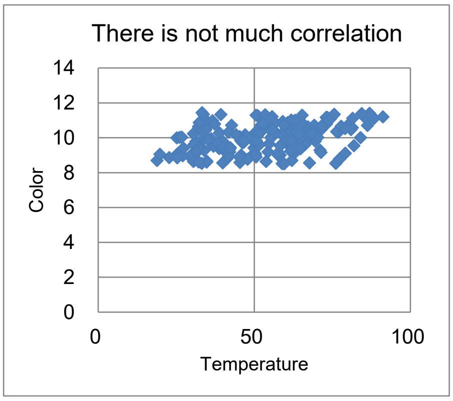
We monitored the color of a product, and its temperature as well, since at the time of research and development it was known that its color is related to the temperature.
Our customer informed us that the product quality had recently deteriorated, so we decided to investigate when and why the quality had deteriorated.

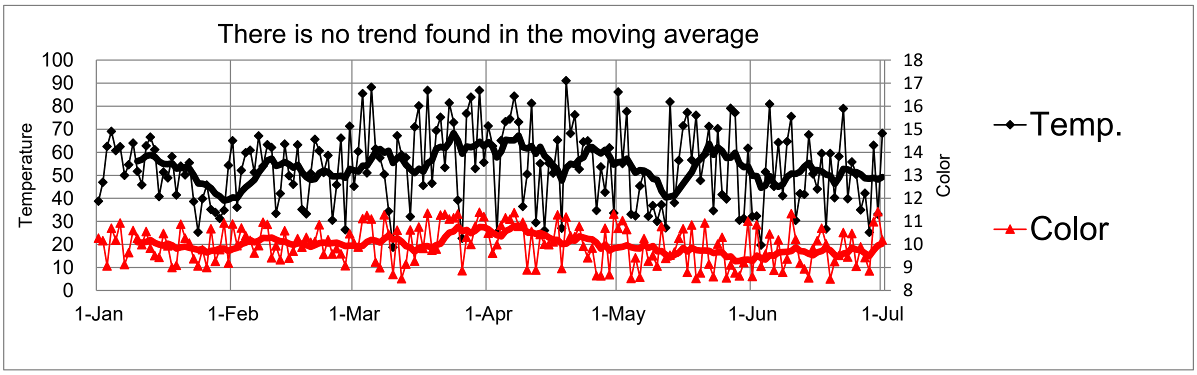
Date	Temperature	Color
1-Jan	38.80	10.28
2-Jan	47.05	10.17
3-Jan	62.55	9.07
4-Jan	69.15	10.71
5-Jan	60.75	10.19
6-Jan	62.45	10.93
7-Jan	50.05	9.13
8-Jan	54.65	9.65
9-Jan	64.05	10.61
10-Jan	51.75	10.27
25-Jun	50.38	9.05
26-Jun	35.05	9.90
27-Jun	42.27	9.43
28-Jun	25.28	8.85
29-Jun	63.05	10.99
30-Jun	33.27	11.43
1-Jul	68.30	10.20



Examination of Correlation Analysis

From the findings at the time of research and development, temperature and color were supposed to be correlated, but there was no obvious relationship.
Although moving average processing was performed to check the overall trend rather than individual variation, it was not clear at what point anomalies started to occur.
First, let us try to index the color as "1 (abnormal) or 0 (normal)?"



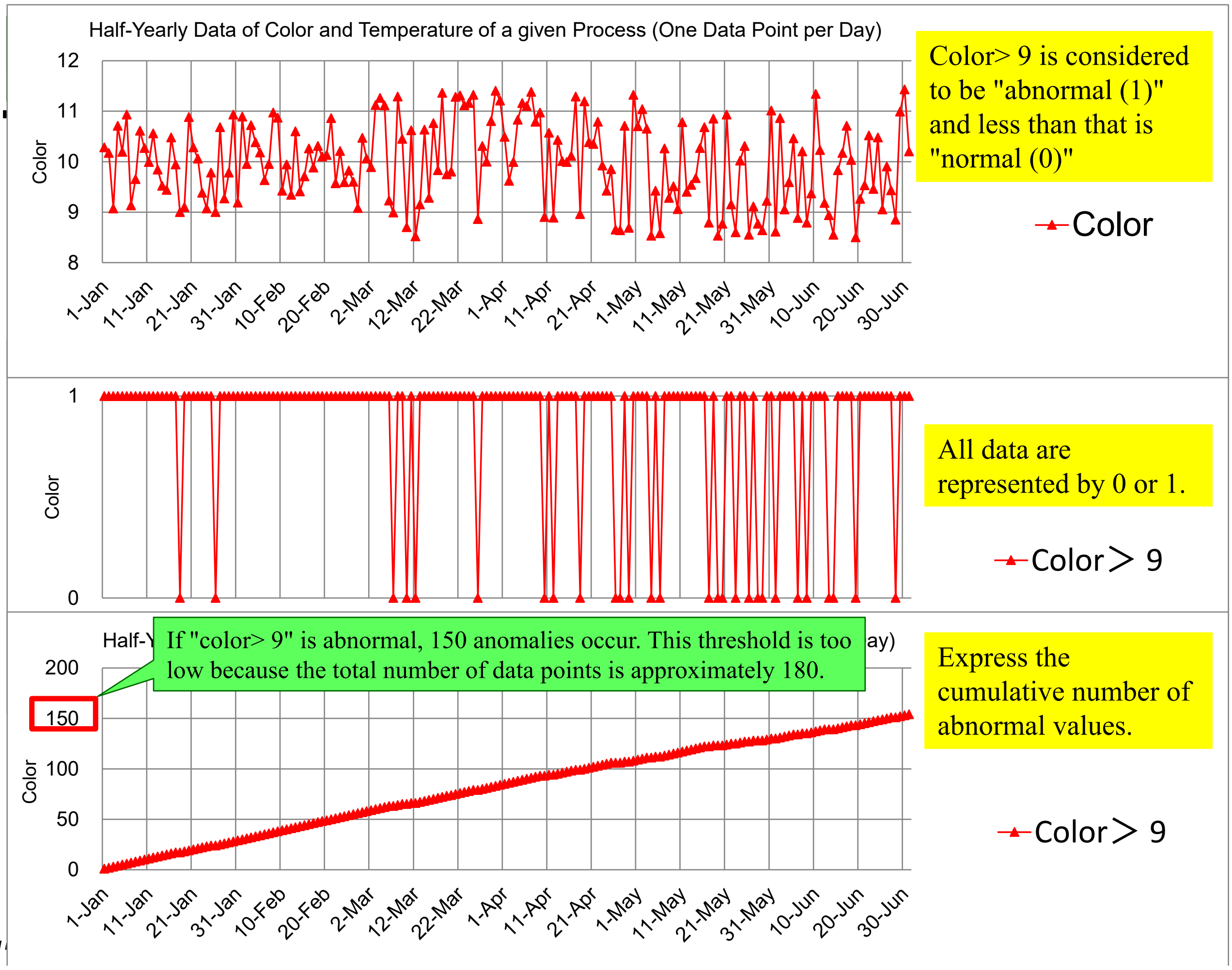


Indexing Data

Date	Temperature	Color	Color > 11	Total Number of Anomalies
1-Jan	38.80	10.28	0	0
2-Jan	47.05	10.17	0	0
3-Jan	62.55	9.07	0	0
4-Jan	69.15	10.71	=IF(C6>D2,1,0)	0
5-Jan	60.75	10.19	0	0
6-Jan	62.45	10.93	0	0
7-Jan	50.05	9.13	0	0
8-Jan	54.65	9.65	0	0
9-Jan	64.05	10.61	0	0
10-Jan	51.75	10.27	0	0
1-Mar	45.33	10.06	0	0
2-Mar	60.40	9.89	0	0
3-Mar	85.56	11.12	1	1
4-Mar	51.13	11.26	1	2
5-Mar	88.31	11.11	1	3
6-Mar	61.57	9.23	0	3

Indexing is done using Excel's if statement. For example, with "color> 11", a logical expression is incorporated so as to be "1". For instance, the color of January 4 is 10.71 and the index is 0, and the color of March 3 is 11.12 and the index is 1.

Once all data have been indexed, their accumulations are calculated. This makes it possible to know the total number of "index 1 events (anomalies)" in the target period, and to clarify the pattern of increase (such as the time when they have started to increase rapidly).

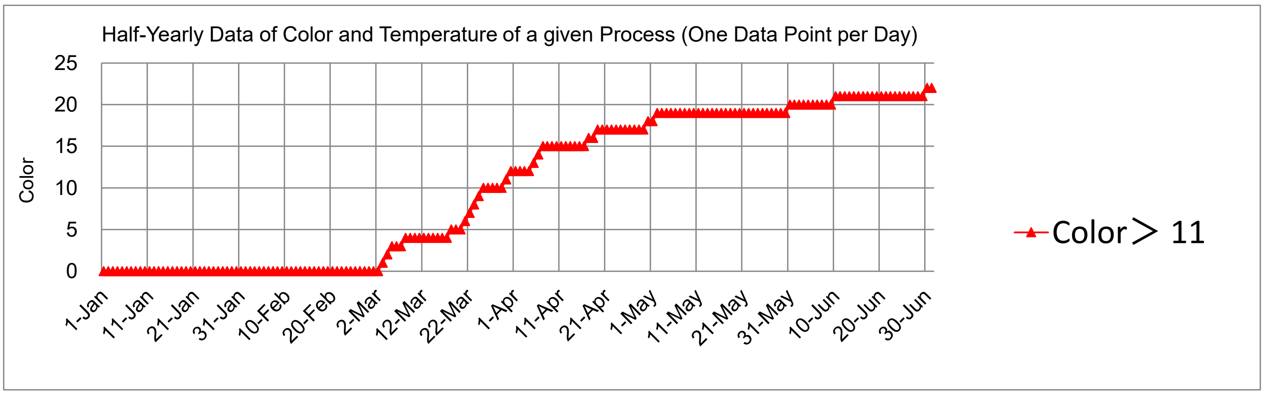


On Setting the Threshold

One of the purposes of analyzing plant data is often to find a cause when there is an event that exceeds a certain threshold. When there is a standard such as a product standard as a threshold, it can be conceivably applied. When there is no such objective standard, it can be considered as follows: out of the total number of data points, when the number of events of which the behavior is different from the others is around 10% to 20% of the total, it can be considered that an attempt should be made to find out some cause for "abnormal" events. In other words, the setting of the threshold is a standard that can be determined subjectively depending on what percentage of all the data you want to use to explain an anomaly, and it seems reasonable that it should be set so that the cumulative number of anomalies is around 10% to 20%.

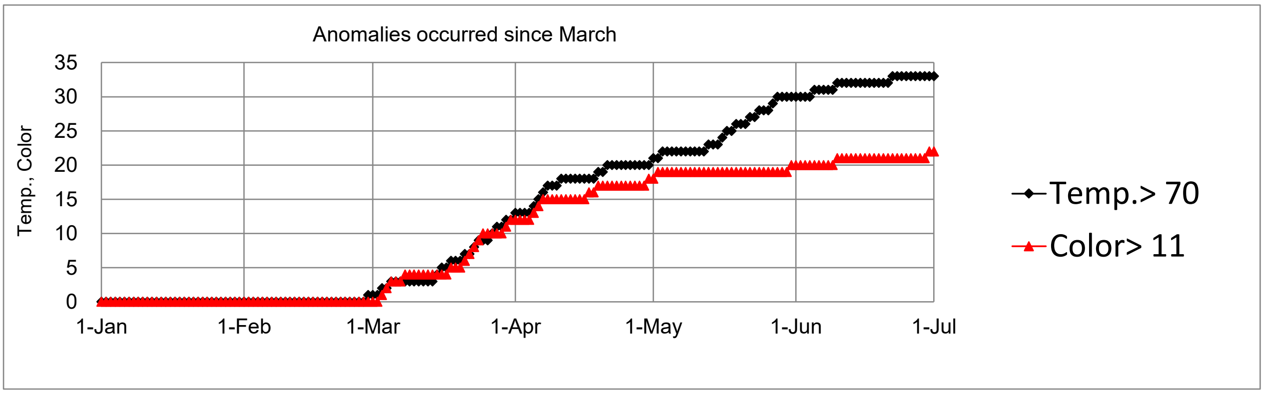
On the other hand, if a variable that is the cause (in this example, the cause of the color anomaly) is measured, and if the threshold for that variable is adjusted, the cumulative trend of anomalies should be consistent with the trend of color anomalies.

Appropriate Threshold Setting



If the threshold value is determined to be 11, the number of abnormal cases is about 10% of the total measured data.
If the cause that can explain this is measured, the trend should be consistent with the trend of the color anomaly.
In this case, only the temperature is measured, so the threshold value is adjusted for the temperature data, and we consider making a cumulative number of anomalies of the same level as the number of color anomalies.

Comparison of Cumulative Abnormal Trend of All Data



- The threshold value was examined for the variable considered to be the cause (in this case, temperature), and the number of anomalies within the examination period was made to be similar to the number of color anomalies. As a result, in the above example, temperatures greater than 70 were assumed to be abnormal.

Interpretation of results

- It turns out that the anomalies had started to occur in March.
- On the other hand, color anomalies decreased from May, but temperature anomalies continued to increase. From this, it is understood that temperature is not the only cause, but that there are also other causes.

Conclusion

In research and development stages or in laboratory experiments, it is common practice to thoroughly consider operation conditions in order to examine the relationship between variables of interest. For example, in order to see the influence of the raw material molar ratios on reaction selectivity, experiments are typically conducted by keeping other factors constant (temperature, pressure, amounts of catalyst and the like).
In a plant, however, it is often the case that the relationships cannot be understood clearly by taking out only the variables that were originally related. Indeed, variables cannot be kept constant due to the effects of recycling, conditions are not the same every time for raw materials and catalyst lots, conditions can change due to the accumulation of contaminants in equipment, and there can be various disturbances due to fluctuations in utilities and the like.
However, since such data show the actual behavior of the plant under a wide range of conditions, stable operation of the plant and quality and yield improvements can be achieved if they can be analyzed effectively.
Cumulative anomaly analysis can be effectively used for such purposes.