Monitoring Photochemical Pollutants for Anomaly Detection based on Symbolic Interval-Valued Data Analysis

Liang-Ching Lin
(Joint work with Meihui Guo and Sangyeol Lee)

Department of Statistics
National Cheng Kung University, Tainan, 701, ROC

January 18, 2022

2021 TMS Annual Meeting

Outline

Photochemical Data

Table: Photochemical Data
https://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx

Date	Time	Ethane	Ethylene	⋯	Propylene	Isobutane
2016/12/28	4	–	0.36	⋯	0.04	0.11
2016/12/28	5	0.02	0.43	⋯	–	0.1
2016/12/28	6	–	0.27	⋯	–	0.1
2016/12/28	7	0.02	0.43	⋯	0.06	0.12
2016/12/28	8	0.03	0.48	⋯	0.08	0.1

Introduction

- Ozone: photochemical secondary pollutant
  photochemical pollutant$\uparrow \longmapsto$ exposure to sunlight $\longmapsto$ ozone
- Ozone recently surpassed particulate matter (PM) as a main source of air pollution
- Photochemical Assessment Monitoring Stations: to identify the source of air pollutants
- dataset: 56 variables recorded hourly from January 1, 2016 to December 31, 2017.

Photochemical Data

Table: Photochemical Data
https://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx

Date	Time	Ethane	Ethylene	⋯	Propylene	Isobutane
2016/12/28	4	–	0.36	⋯	0.04	0.11
2016/12/28	5	0.02	0.43	⋯	–	0.1
2016/12/28	6	–	0.27	⋯	–	0.1
2016/12/28	7	0.02	0.43	⋯	0.06	0.12
2016/12/28	8	0.03	0.48	⋯	0.08	0.1

- detect the abnormal days with higher values & the main pollutants;
- missing/repeated data $\longrightarrow$ daily mean
- preserve more information $\longrightarrow$ maximum & minimum values.

Data Cleaning

- no records are found for all hours and all variables → remove that day
- missing rates of some variables are higher than 70% → remove that variable
- missing values on a day are more than 12 hours → interpolate by $[(t-1)+(t+1)]/2$
- The remaining 48 variables, 341 days for 2016, and 343 days for 2017

Principal Component Analysis

- monitoring individually $\rightarrow$ include many false alarms
- First, we perform PCA to detect the main pollution components.
- Let $\mathbf{Y}^{(m)} = ((Y_{i,j}^{(m)}))_{\{1 \le i \le m,\ 1 \le j \le p\}}$ be the data matrix.
- The covariance matrix

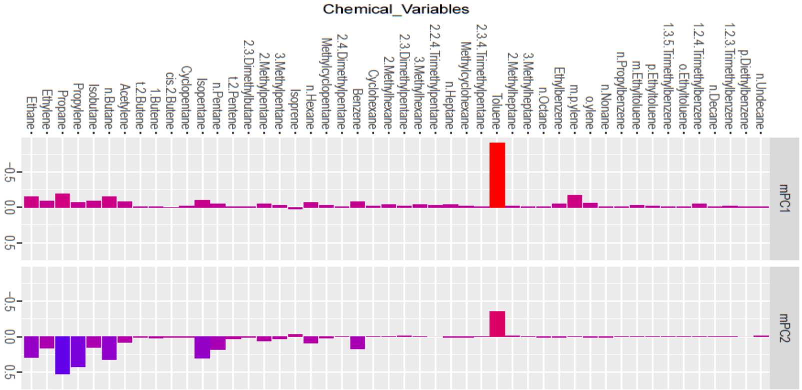
$$\Sigma^{(m)} = (\mathbf{Y}^{(m)} - \mathbf{1}_m(\bar{Y}^{(m)})')'(\mathbf{Y}^{(m)} - \mathbf{1}_m(\bar{Y}^{(m)})')/m$$

- Using the spectral decomposition to $\Sigma^{(m)}$,

$$\Sigma^{(m)} = \lambda_1^{(m)} \nu_1^{(m)} (\nu_1^{(m)})' + \cdots + \lambda_p^{(m)} \nu_p^{(m)} (\nu_p^{(m)})',$$

Results

- The first and first two principal components explain 66.45% and 82.07% of the variability.



Statistical Quality Control – Phase I (2016)

- Second, we monitor the air pollution based on the obtained principal components.
- The projections based on the first two components

$$\mathbf{S}_k^{(m)} = \left(S_{1,k}^{(m)}, \ldots, S_{m,k}^{(m)}\right)' = \mathbf{Y}^{(m)} \cdot \nu_k^{(m)}, \quad k = 1, 2.$$

- Then, we construct the Shewhart control chart (with 6-sigma).

First Principal Scores (2016)

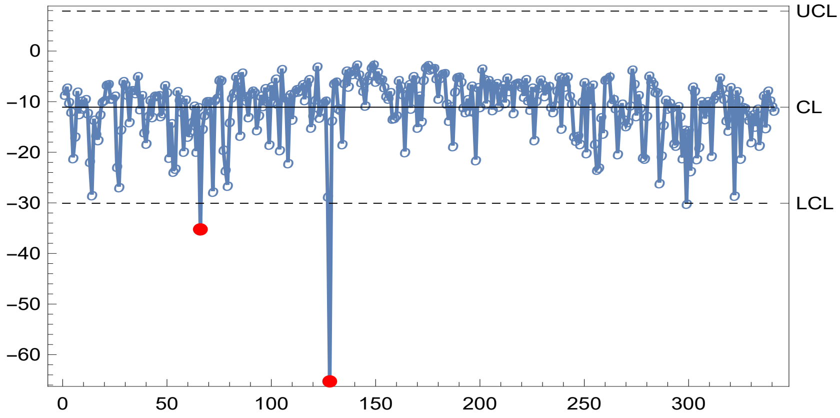


Figure: Shewhart control chart for the first principal scores on 2016.

Second Principal Scores (2016)

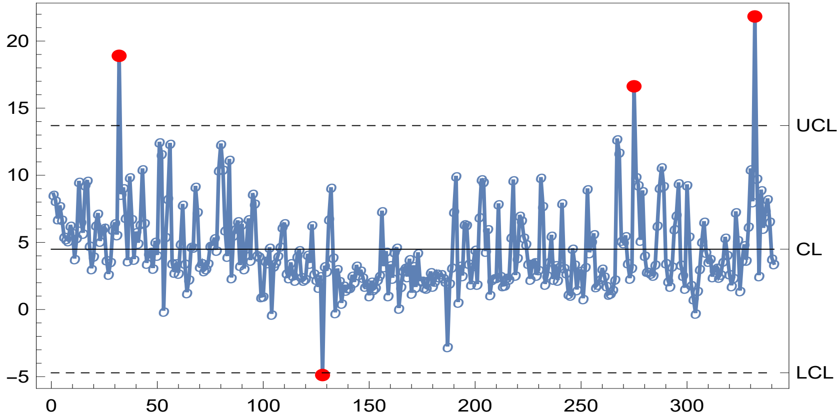


Figure: Shewhart control chart for the second principal scores on 2016.

Out-of control days

Table: Assignable causes based on the first two principal components of daily mean data.

	PCA of daily mean data				
	PC1		PC2		
date	5/21	3/17	12/20	2/9	10/24
causes	31^{st}: 60.62	31^{st}: 30.105	3^{rd}: 17.538	3^{rd}: 15.903	3^{rd}: 7.223

Out-of control days

Table: Assignable causes based on the first two principal components of daily mean data.

	PCA of daily mean data				
	PC1		PC2		
date	5/21	3/17	12/20	2/9	10/24
causes	31^{st}: 60.62	31^{st}: 30.105	3^{rd}: 17.538	3^{rd}: 15.903	3^{rd}: 7.223

largest value can't be detected
(289.08 on 7/26 of 4th compound, propylene)

Statistical Quality Control – Phase II (2017)

- Remove out-of-control points until all points are in-control.
- monitor the first two principal scores for the next month and update monthly (including principals) by using a rolling window

First Principal Scores (2017)

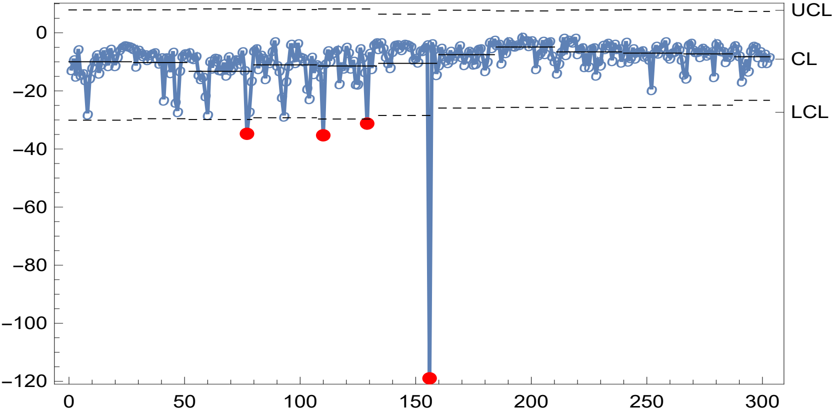


Figure: Shewhart control chart for the first principal scores on 2017.

Second Principal Scores (2017)

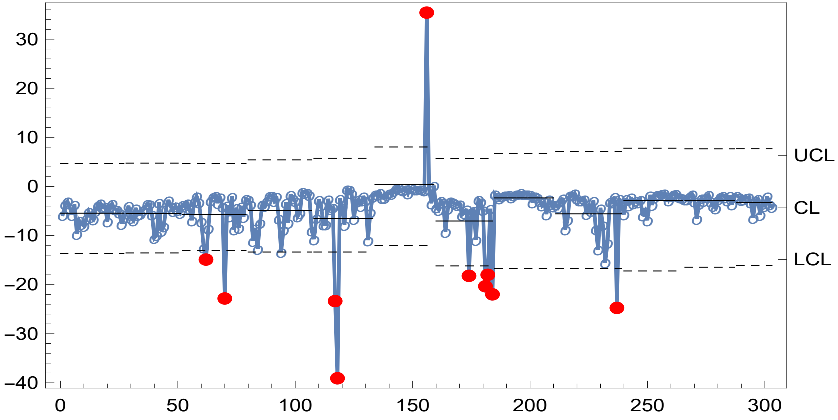


Figure: Shewhart control chart for the second principal scores on 2017.

Out-of control days

Table: Assignable causes based on the first two principal components of the daily mean data of 2017.

	PCA of daily mean data						
	PC1					PC2	
date	6/27	3/29	5/3	5/27		5/11	9/28
causes	31^{st}: 129.48	31^{st}: 30.32	31^{st}: 31.1	31^{st}: 26.95		4^{th}: 46.68	4^{th}: 25.41

	PCA of daily mean data						
	PC2						
date	3/20	5/10	7/30	7/27	7/18	7/28	3/12
causes	4^{th}: 21.9	4^{th}: 25.83 31^{st}: 10.38	4^{th}: 23.14	4^{th}: 17.6 31^{st}: 7.78	4^{th}: 16.83	4^{th}: 17.28	4^{th}: 12.75 31^{st}: 6.43

Out-of control days

Table: Assignable causes based on the first two principal components of the daily mean data of 2017.

	PCA of daily mean data						
	PC1					PC2	
date	6/27	3/29	5/3	5/27		5/11	9/28
causes	31^{st}: 129.48	31^{st}: 30.32	31^{st}: 31.1	31^{st}: 26.95		4^{th}: 46.68	4^{th}: 25.41

	PCA of daily mean data						
	PC2						
date	3/20	5/10	7/30	7/27	7/18	7/28	3/12
causes	4^{th}: 21.9	4^{th}: 25.83	4^{th}: 23.14	4^{th}: 17.6	4^{th}: 16.83	4^{th}: 17.28	4^{th}: 12.75
		31^{st}: 10.38		31^{st}: 7.78			31^{st}: 6.43

2nd largest value can't be detected
(205.69 on 9/19 of 31th compound, toluene)

Symbolic Interval-Valued Variables

- daily mean: lack of comprehensive information and makes the data too concentrated
- to preserve more information: reorganize the data in the form of daily maximum and minimum values
- symbolic (interval-valued) data analysis: Billard and Diday (2003, 2006); Zhang et al. (2019); Su et al. (2015); Brito (2014); Lauro and Plumbo (2005).
- Most literatures have analyzed symbolic data based on uniform distributions.
- In this study, the interval-valued variables are viewed as the largest-order and smallest-order statistics from a normal distribution, as shown in Lin et al. (2021).

Notations

- dataset: $\Omega = \{X_1, \ldots, X_N\}$ where $N = n \times m$
- split Ω into m groups of n elements
- interval-valued data: $\mathbf{X}_i = [X_{l,i}, X_{u,i}]$, $i = 1, \ldots, m$, where $X_{l,i} = \min\{X_{(i-1)n+1}, \ldots, X_{in}\}, X_{u,i} = \max\{X_{(i-1)n+1}, \ldots, X_{in}\}$.
- Assumption: $X_1, \ldots, X_N \sim N(\mu, \sigma^2)$
- This assumption can be easily released to other distributions.

Univariate Descriptive Statistics

- Referring to Blom (1958), the kth order statistics of a standard normal distribution with a sample of size n is

$$E(Z_{(k)}) \approx \Phi^{-1}\left(\frac{k-\alpha}{n-2\alpha+1}\right).$$

- We have

$$\begin{aligned}\hat{\mu} &= \frac{1}{2m}\sum_{i=1}^{m}(X_{l,i}+X_{u,i}),\\ \hat{\sigma}^2 &= \left(\frac{m^{-1}\sum_{i=1}^{m}(X_{u,i}-X_{l,i})}{\Phi^{-1}(\frac{n-\alpha}{n-2\alpha+1})-\Phi^{-1}(\frac{1-\alpha}{n-2\alpha+1})}\right)^2.\end{aligned}$$

Bivariate Interval-Valued Variables

Theorem 1

The likelihood function of $\boldsymbol{\theta}$ based on $(\mathbf{X}_1, \mathbf{Y}_1) \ldots, (\mathbf{X}_m, \mathbf{Y}_m)$ is given by

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{m} [n(n-1)\mathbb{I}^{n-2} A_1 + n(n-1)(n-2)\mathbb{I}^{n-3} A_2 \\
&\quad + n(n-1)(n-2)(n-3)\mathbb{I}^{n-4} A_3],
\end{aligned}
$$

$$
\begin{aligned}
A_1 &= f_{X,Y}(x_u, y_u) f_{X,Y}(x_l, y_l) + f_{X,Y}(x_u, y_l) f_{X,Y}(x_l, y_u), \\
A_2 &= f_{X,Y}(x_u, y_u)\mathbb{I}_x(y_l)\mathbb{I}_y(x_l) + f_{X,Y}(x_u, y_l)\mathbb{I}_x(y_u)\mathbb{I}_y(x_l) \\
&\quad + f_{X,Y}(x_l, y_u)\mathbb{I}_x(y_l)\mathbb{I}_y(x_u) + f_{X,Y}(x_l, y_l)\mathbb{I}_x(y_u)\mathbb{I}_y(x_u), \\
A_3 &= \mathbb{I}_x(y_u)\mathbb{I}_x(y_l)\mathbb{I}_y(x_u)\mathbb{I}_y(x_l), \quad \mathbb{I} = \int_{x_l}^{x_u}\int_{y_l}^{y_u} f_{X,Y}(x,y)\,dxdy, \\
\mathbb{I}_x(b) &= \int_{x_l}^{x_u} f_{X,Y}(x,b)\,dx, \quad \mathbb{I}_y(a) = \int_{y_l}^{y_u} f_{X,Y}(a,y)\,dy.
\end{aligned}
$$

Briefly Proof

(pf): Assume that the joint probability density function of $\{X_l, X_u, Y_l, Y_u\}$ is $g(x_l, x_u, y_l, y_u)$. Since

$$\begin{aligned}
& \int_{x_l}^{\infty}\int_{-\infty}^{x_u}\int_{y_l}^{\infty}\int_{-\infty}^{y_u} g(x,y,z,w)dwdzdydx \\
= \quad & P(X_{(1)} \geq x_l, X_{(n)} \leq x_u, Y_{(1)} \geq y_l, Y_{(n)} \leq y_u) \\
= \quad & P(x_l \leq X_1 \leq x_u, \ldots, x_l \leq X_n \leq x_u, y_l \leq Y_1 \leq y_u, \ldots, y_l \leq Y_n \leq y_u) \\
= \quad & \left[\int_{x_l}^{x_u}\int_{y_l}^{y_u} f_{X,Y}(x,y)dxdy\right]^n .
\end{aligned}$$

Then, differentiating the above equation with respect to all variables on both sides, we obtain the results.

$$\frac{\partial}{\partial x_l} \rightarrow \frac{\partial}{\partial x_u} \rightarrow \frac{\partial}{\partial y_l} \rightarrow \frac{\partial}{\partial y_u}$$

$f(x, y) = ?$

- To ensure wide applicability, we consider the copula-linked function to the joint distribution function.
- Let $u = \Phi[(x - \mu_x)/\sigma_x]$ and $v = \Phi[(y - \mu_y)/\sigma_y]$, Gaussian copula

$$f_{X,Y}(x, y) = \phi(\Phi^{-1}(u), \Phi^{-1}(v)).$$

- Clayton copula

$$c^{Cl}(u, v) = (1/\rho + 1)(uv)^{-(1/\rho+1)} \left(u^{-1/\rho} + v^{-1/\rho} - 1\right)^{-(\rho+2)},$$

- Gumbel copula

$$c^{Gu}(u, v) = \exp\left\{-\left[(-\log u)^{\rho} + (-\log v)^{\rho}\right]^{1/\rho}\right\} \frac{(\log u \log n)^{\rho-1}}{uv}$$
$$\left[((-\log u)^{\rho} + (-\log v)^{\rho})^{2/\rho-2} + (\rho - 1)\left((-\log u)^{\rho} + (-\log v)^{\rho}\right)^{1/\rho-2}\right].$$

- Then $f_{X,Y}(x, y) = c^{\text{Cl or Gu}}(u, v)\phi((x - \mu_x)/\sigma_x)\phi((y - \mu_y)/\sigma_y)$.

SQC for Univariate Interval-Valued Variables

- Since

$$\begin{aligned}\alpha/2 &= P(X_u > x) = 1 - P(X_u \le x) = 1 - P(X_1 \le x, \ldots, X_n \le x) \\ &= 1 - [P(X_1 \le x)]^n = 1 - \left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^n.\end{aligned}$$

- Then, we have

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = (1-\alpha/2)^{1/n}, \text{ and thus, UCL} = \mu + \sigma\Phi^{-1}[(1-\alpha/2)^{1/n}].$$

- Similarly,

$$\text{LCL} = \mu + \sigma\Phi^{-1}[1-(1-\alpha/2)^{1/n}].$$

SQC for Univariate Interval-Valued Variables

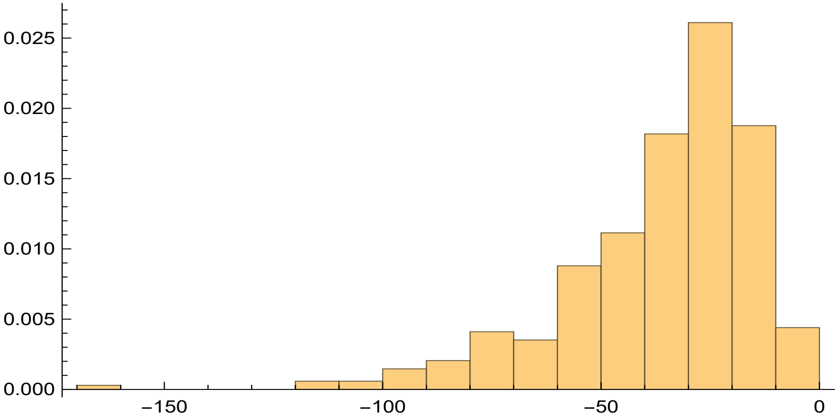
- Let $[\mathbf{X}_i, \mathbf{Y}_i] = [X_{l,i}, X_{u,i}, Y_{l,i}, Y_{u,i}]$, $i = 1, \ldots, m$, be the observed bivariate interval-valued variables.
- We first estimate the covariance matrix $\hat{\Sigma}$.
- Then, perform the spectrum decomposition to $\hat{\Sigma}$

$$\hat{\Sigma} = \lambda_1 \boldsymbol{\nu}_1 (\boldsymbol{\nu}_1)' + \lambda_2 \boldsymbol{\nu}_2 (\boldsymbol{\nu}_2)'.$$

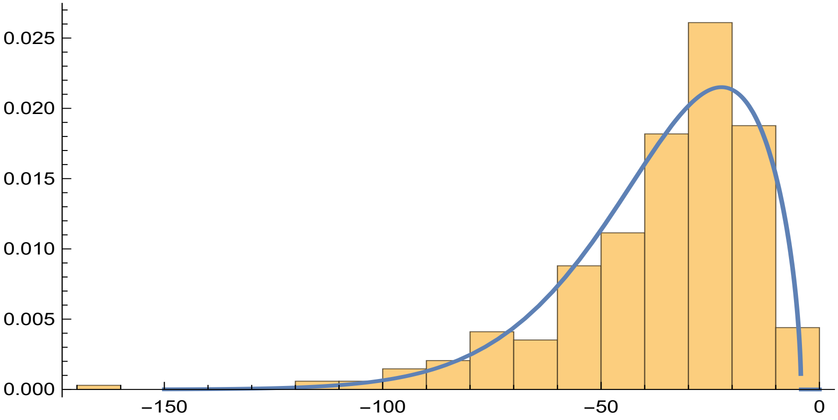
- Referring to Billard and Diday (2006), the principal scores are

$$\begin{aligned} S_{i,k}^{(U)} &= \left(\nu_{k,1}(X_{u,i} - \hat{\mu}_x)\mathbf{1}_{\{\nu_{k,1} \geq 0\}} + \nu_{k,1}(X_{l,i} - \hat{\mu}_x)\mathbf{1}_{\{\nu_{k,1} < 0\}}\right) + \\ &\quad \left(\nu_{k,2}(Y_{u,i} - \hat{\mu}_y)\mathbf{1}_{\{\nu_{k,2} \geq 0\}} + \nu_{k,2}(Y_{l,i} - \hat{\mu}_y)\mathbf{1}_{\{\nu_{k,2} < 0\}}\right), \\ S_{i,k}^{(L)} &= \left(\nu_{k,1}(X_{l,i} - \hat{\mu}_x)\mathbf{1}_{\{\nu_{k,1} \geq 0\}} + \nu_{k,1}(X_{u,i} - \hat{\mu}_x)\mathbf{1}_{\{\nu_{k,1} < 0\}}\right) + \\ &\quad \left(\nu_{k,2}(Y_{l,i} - \hat{\mu}_y)\mathbf{1}_{\{\nu_{k,2} \geq 0\}} + \nu_{k,2}(Y_{u,i} - \hat{\mu}_y)\mathbf{1}_{\{\nu_{k,2} < 0\}}\right). \end{aligned}$$

UCL? LCL?



Extreme Value Distribution



UCL and LCL

- According to the shapes of the histograms, we fit a generalized extreme value distribution to these scores.
- generalized extreme value distribution:

$$
\begin{aligned}
f(x) &= \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}, \qquad F(x) = e^{-t(x)}, \\
t(x) &= \begin{cases} \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ e^{-(x-\mu)/\sigma} & \text{if } \xi = 0. \end{cases}
\end{aligned}
$$

- Finally, for a given α, UCL(LCL) $= 1 - \alpha/2(\alpha/2)$ quantiles of the corresponding fitted extreme value distribution;

Simulation Results on Parameter Estimations – I

Table: Relative errors of the estimators of μ and σ.

	$N(1,1)$	$N(-10,1)$	$N(10,25)$	$N(-10,25)$	$N(20,25)$
$\hat{\mu}$	0.017157	0.001839	0.008008	0.008668	0.004029
$\hat{\sigma}^2$	0.013406	0.011950	0.013342	0.013397	0.012932

Simulation Results on Parameter Estimations – II

Table: Relative errors of the MLE of ρ for Gaussian.

True value	0.85	0.65	0.45	0.25
RE of $\hat{\rho}$	0.018	0.201	1.172	1.791
True value	-0.25	-0.45	-0.65	-0.85
RE of $\hat{\rho}$	1.539	1.010	0.404	0.019

Simulation Results on Parameter Estimations – III

Table: Relative errors of the MLE of ρ for Clayton and Gumbel copulae.

	Clayton copula			
True value of ρ	1.167	0.5	0.214	0.056
Kendall's τ	0.3	0.5	0.7	0.9
RE of $\hat{\rho}$	0.071	0.061	0.056	0.044

	Gumbel copula			
True value of ρ	1.43	2	3.33	10
Kendall's τ	0.3	0.5	0.7	0.9
RE of $\hat{\rho}$	0.036	0.034	0.033	0.021

Simulation Results on Univariate Interval SQC

Table: ARL_0 (out-of-control numbers for each 370 runs)

N(0,1)	N(2,9)	N(10,25)
1.0226	0.9852	0.9118
(0.1067)	(0.1016)	(0.0911)

Table: ARL_1

N(0.5,1)	N(1,1)	N(1.5,1)	$N(0,1.25^2)$	$N(0, 1.5^2)$
68.226	9.557	2.01	6.119	1.22
(68.941)	(8.827)	(1.419)	(5.754)	(0.518)

Simulation Results on Bivariate Interval SQC – I

Table: ARL_0 (out-of-control numbers for each 370 runs)

copula	Gaussian ($\rho = 0.5$)	Gaussian ($\rho = -0.5$)	Gumbel ($\rho = 2$)	Clayton ($\rho = 0.5$)
first principal	1.376 (0.0825)	1.384 (0.0823)	1.329 (0.0762)	1.354 (0.0842)
second principal	1.4933 (0.0797)	1.4813 (0.0985)	1.465 (0.0909)	1.496 (0.0977)

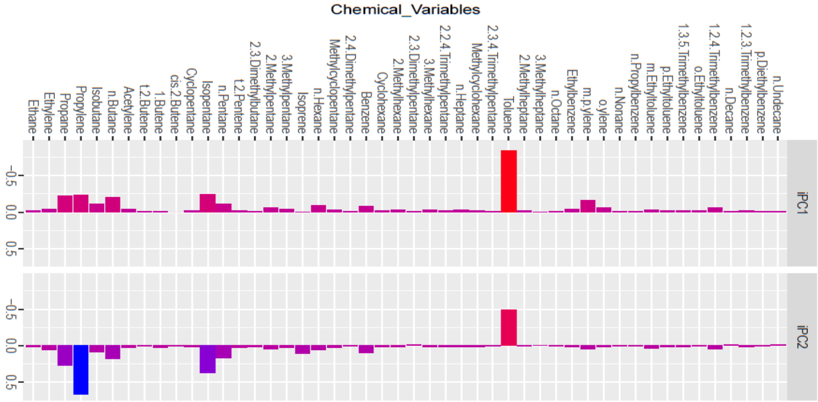
Simulation Results on Bivariate Interval SQC – II

Table: ARL_1

	Gaussian copula ($\rho = 0.5$)				
	$X \sim N(0.5, 1)$	$X \sim N(1, 1)$	$X \sim N(1.5, 1)$	$X \sim N(0, 1.25^2)$	$X \sim N(0, 1.5^2)$
second principal	56.55 (65.98)	8.77 (9.38)	2.16 (1.79)	5.85 (6.11)	1.29 (0.64)
	$Y \sim N(-4, 4)$	$Y \sim N(-3, 4)$	$Y \sim N(-2, 4)$	$Y \sim N(-5, 2.5^2)$	$Y \sim N(-5, 3^2)$
first principal	65.09 (73.58)	10.91 (15.93)	2.05 (1.84)	5.18 (4.81)	1.18 (0.45)
	Gumbel copula ($\rho = 2$)				
	$X \sim N(0.5, 1)$	$X \sim N(1, 1)$	$X \sim N(1.5, 1)$	$X \sim N(0, 1.25^2)$	$X \sim N(0, 1.5^2)$
second principal	52.75 (61.39)	8.46 (9.74)	2.09 (1.84)	6.37 (6.32)	1.34 (0.68)
	$Y \sim N(-4, 4)$	$Y \sim N(-3, 4)$	$Y \sim N(-2, 4)$	$Y \sim N(-5, 2.5^2)$	$Y \sim N(-5, 3^2)$
first principal	82.01 (77.53)	16.67 (20.16)	3.39 (4.12)	6.06 (6.43)	1.23 (0.55)
	Clayton copula ($\rho = 0.5$)				
	$X \sim N(0.5, 1)$	$X \sim N(1, 1)$	$X \sim N(1.5, 1)$	$X \sim N(0, 1.25^2)$	$X \sim N(0, 1.5^2)$
second principal	59.76 (65.42)	8.38 (9.16)	2.21 (1.82)	6.68 (6.95)	1.33 (0.67)
	$Y \sim N(-4, 4)$	$Y \sim N(-3, 4)$	$Y \sim N(-2, 4)$	$Y \sim N(-5, 2.5^2)$	$Y \sim N(-5, 3^2)$
first principal	63.68 (66.70)	10.04 (12.24)	2.03 (1.82)	6.4 (6.16)	1.28 (0.62)

Empirical Results on PCA

- For photochemical data, we consider the Clayton copula.
- The 1^{st} and $1^{st}+2^{nd}$ explain 65.86% and 79.01%



Phase I (2016) SQC Results – 1^{st}

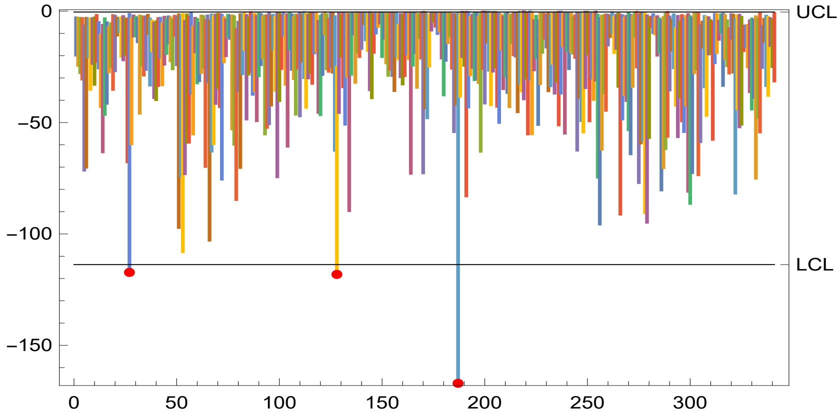


Figure: Control chart of the first interval-valued projections on 2016.

Phase I (2016) SQC Results – 2^{nd}

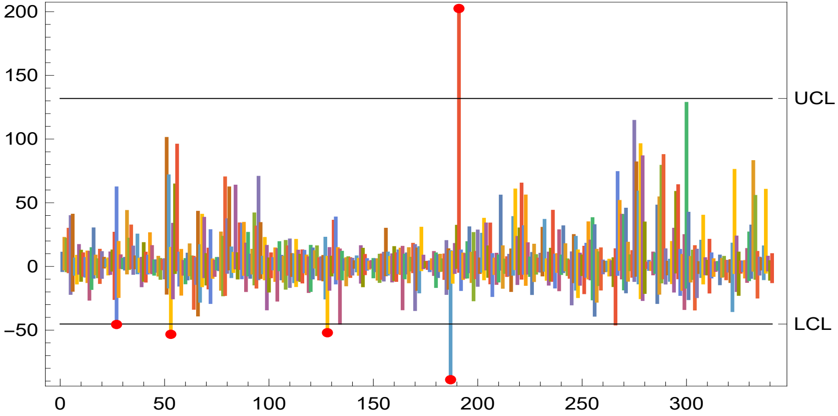


Figure: Control chart of the second interval-valued projections on 2016.

Out-of control days

Table: Assignable causes based on the first two principal components of daily interval-valued data.

	PCA of daily interval-valued data				
	PC1			PC2	
date	7/22	5/21	1/29	7/26	3/4
causes	31^{st}: 184.5	31^{st}: 120.38	31^{st}: 96.58	4^{th}: 289.08	31^{st}: 111.46
			4^{th}: 58.34		

largest value can be detected
(289.08 on 7/26 of 4th compound, propylene)

Phase II (2017) SQC Results – 1^{st}

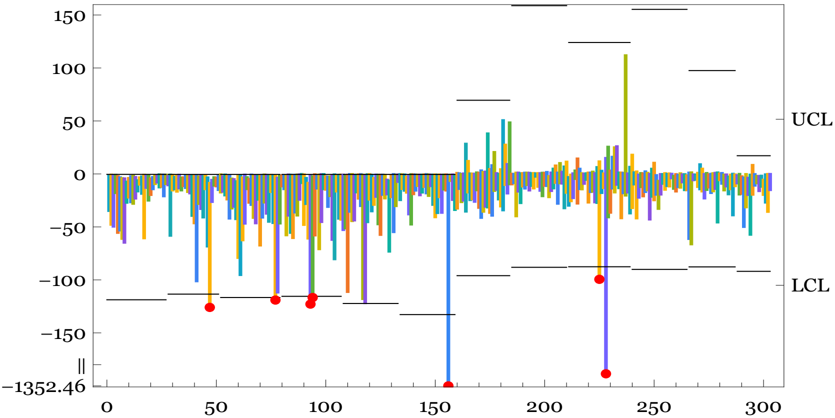


Figure: Control chart of the first interval-valued projections on 2017.

Phase II (2017) SQC Results – 2^{nd}

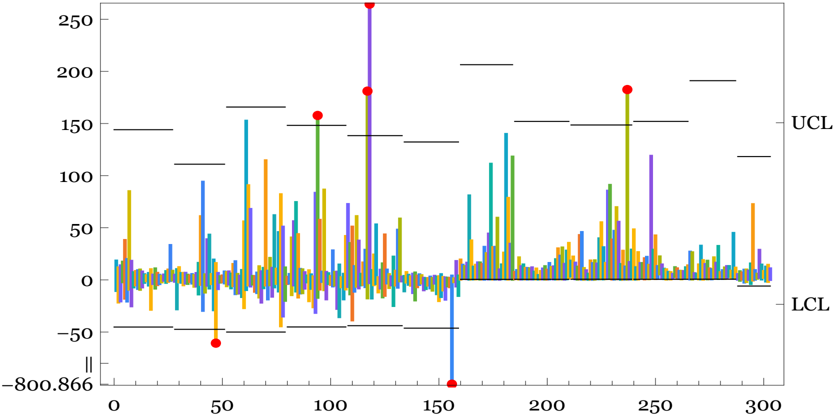


Figure: Control chart of the second interval-valued projections on 2017.

Out-of control days

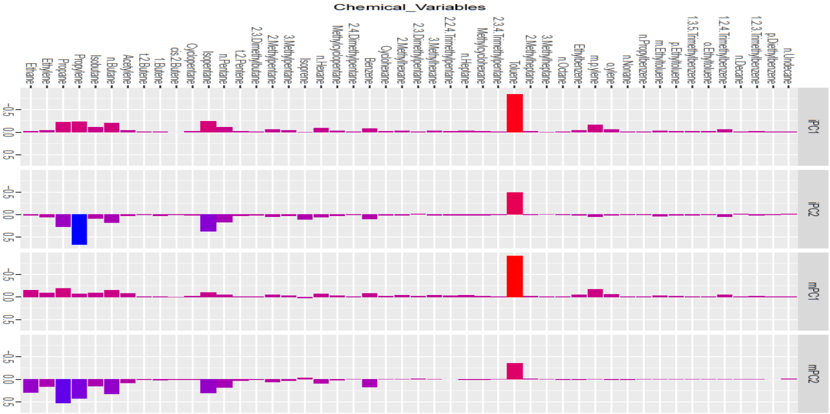
Table: Assignable causes based on the first two principal components of the daily interval-valued data of 2017.

	PCA of daily interval-valued data				
	PC1				
date	6/27	9/19	2/23	4/14	3/29
causes	31^{st}: 1677.25	31^{st}: 205.69	31^{st}: 126.28	31^{st}: 69.72	31^{st}: 91.42
				4^{th}: 67.14	4^{th}: 87.33

	PCA of daily interval-valued data				
	PC1		PC2		
date	4/15	9/16	5/11	5/10	9/28
causes	31^{st}: 38.86	31^{st}: 107.11	4^{th}: 299.25	4^{th}: 201.336	4^{th}: 250.61
	4^{th}: 169.07			31^{st}: 40.77	

2nd largest value can be detected
(205.69 on 9/19 of 31th compound, toluene)

Comparison of Principals



Comparison of Out-of-Control Days on 2016

Table: Assignable causes based on the first two principal component scores on 2016.

	PCA of daily mean data				
	PC1		PC2		
date	5/21	3/17	12/20	2/9	10/24
causes	31^{st}: 60.62	31^{st}: 30.105	3^{rd}: 17.538	3^{rd}: 15.903	3^{rd}: 7.223

	PCA of daily interval-valued data				
	PC1			PC2	
date	7/22	5/21	1/29	7/26	3/4
causes	31^{st}: 184.5	31^{st}: 120.38	31^{st}: 96.58 4^{th}: 58.34	4^{th}: 289.08	31^{st}: 111.46

Comparison of Out-of-Control Days on 2017

Table: Assignable causes based on the first two principal component scores on 2017.

both	6/27(31^{st}: M=129/I=1677) 3/29(31^{st}: M=30/I=91, 4^{th}: I=87) 5/10(4^{th}: M=26/I=201) 5/11(4^{th}: M=47/I=299) 9/28(4^{th}: M=25/I=250)
mean	5/3, 5/27 (31^{st}: 27∼31), 3/12, 3/20, 7/18, 7/27, 7/28, 7/30(4^{th}: $13 \sim 23$)
interval	9/19(31^{st}: 205.69), 2/23(31^{st}: 126.28) 9/16(31^{st}: 107.11), 4/15(4^{th}: 169.07)

Concluding Remarks

- We conducted univariate and bivariate symbolic interval-valued data analysis based on normal distribution.
- Moreover, a copula-linked function provides wide elasticity for the bivariate interval-valued variables.
- In our empirical study, the innovative interval-valued control chart can capture the date on which the abnormal maximum occurred, much better than the method of averaging out with other small values, confirming the validity of the proposed methods.
- The normal distribution can be extended to other distributions, possibly allowing n to be a random variable.

Thanks for Your Attention