C. Model Construction and Evaluation

- stepwise regression algorithm with Akaika information criterion (AIC)
- **P** be the set of indices of the factors selected from factor screening
- |**P'** | to denote the number of elements contained in **P**
- *P* is an arbitrary subset of **P**
- F^{p} denotes the set of the factors whose indices are contained in *P*.
- x^{p} that is a matrix of *N*×*(*|*P*|+1*)*

$$\hat{l}^{\mathcal{P}} = l(\widehat{\boldsymbol{\beta}}(\mathcal{P}),\ \widehat{\sigma}^{2}(\mathcal{P}),\ \widehat{\rho}^{2}(\mathcal{P})|\ \mathbf{y}, \mathbf{x}(\mathcal{P}))$$

$$l^{\mathcal{P}}(\boldsymbol{\beta}(\mathcal{P}),\ \sigma_{\zeta}^{2}(\mathcal{P}),\ \rho^{2}(\mathcal{P})|\ \mathbf{x}^{\mathcal{P}}, \mathbf{y})$$

C. Model Construction and Evaluation (cont'd)

$$\mathrm{AIC}_{\mathrm{SPBM}}^{\mathcal{P}} = -2\hat{l}^{\mathcal{P}} + 2\left(1 + \sum_{p \in \mathcal{P}} d_p\right), \qquad (27)$$

- d_p is the increment of degree of freedom
- $d_p = 1$, if F^p is a numerical factor.
- $d_p = -1$, if F^p is a nominal factor.

C. Model Construction and Evaluation (cont'd)

The proposed algorithm for SBPM with variable selection based on **P'** consists of the following steps:

Step 1. Set $t = 0$ and $\mathcal{P}^0 = \phi$, where ϕ denotes the empty set.

Step 2. Calculate $M = \mathrm{AIC}_{\mathrm{SPBM}}^{\mathcal{P}^0}$.

Step 3. $t = t + 1$.

Step 4. Find $p^* = \arg\min_{p \in \mathbf{P}' \setminus \mathcal{P}^{t-1}} \mathrm{AIC}_{\mathrm{SPBM}}^{\{\mathcal{P}^{t-1}, p\}}$. Set $\mathcal{P}^t = \{\mathcal{P}^{t-1}, p^*\}$.

Step 5. If $M > \mathrm{AIC}_{\mathrm{SPBM}}^{\mathcal{P}^t}$ or $t = |\mathbf{P}'|$, set $M = \mathrm{AIC}_{\mathrm{SPBM}}^{\mathcal{P}^t}$ and go to Step 3.

Step 6. Set $\mathcal{P}^* = \mathcal{P}^{t-1}$, $\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}(\mathcal{P}^{t-1})$, $\widehat{\sigma}^{2*} = \widehat{\sigma}^2(\mathcal{P}^{t-1})$, $\widehat{\rho}^{2*} = \widehat{\rho}^2(\mathcal{P}^{t-1})$ and the final model follows.

Finally, the fitted model is as follows:

$$\hat{Y}_{kn} = \mathbf{X}_{kn}^{\mathcal{P}^*} \widehat{\boldsymbol{\beta}}^*. \tag{28}$$

C. Model Construction and Evaluation (cont'd)

(a) between-batch residuals $e_k^b = \sum_{k=1}^{L} e_{kn}^{\zeta}/N_k$, $k = 1, \ldots, K$ and (b) within-batch residuals $e_{kn}^{w} = e_{kn}^{\zeta} - e_k^b$, $k = 1, \ldots, K, n = 1, \ldots, N_k$, where $e_{kn}^{\zeta} = Y_{kn} - \hat{Y}_{kn}$.

D. Time Complexity

- (a) key factor extraction by statistical analysis
- (b) model construction and evaluation.

- N observations
- P observed factors
- M_T number of trees to be grown in RF

D. Time Complexity (cont'd)

- RF

- CART with the number of factor *P* is $O(\sqrt{P} N \log N)$
- RF is $O(\boldsymbol{M_T} \sqrt{P} N \log N)$
- fit a onedimensional linear regression model by least square approach is $O(N)$

D. Time Complexity (cont'd)

- *SBPM*

- Number of iterations be κ, the time complexity of SBPM part in (a) will be $O(\kappa PN)$.

- Calculating

- λp, $p = 1, \ldots, P$ and sorting them is at most $O(P \log P)$ via a quick sorting algorithm.

D. Time Complexity (cont'd)

- (a) key factor extraction by statistical analysis
- Thus, the time complexity of (a)

  be expressed by $O(\boldsymbol{M_T}\sqrt{P}N \log N) + \kappa PN + P \log P)$
- If $\boldsymbol{M_T}$ and κ are both fixed, it can be replaced by $O(NP \log N \log P)$.
- *(b)* model construction and evaluation.
- In (b), suppose that the number of candidate factors selected in (a) is *P'*
- For the stepwise forward selection algorithm, at most $P'(P' + 1)/2$
- the time complexity of a SBPM is limited by $O(\kappa P'^2 N)$
- Thus, the time complexity of (b) can be expressed by $O(P'^4 N)$ if κ is fixed.

✓ **$O(P^4 N\log N)$, that can be reduced to $O(P^2 N \log N)$ if $P' \approx O(\sqrt{P})$.**

V. VALIDATION

- *1) Simulation Setting*
- Let *Y* denote the response variable
- Effective factor: the factor that effectively affects *Y*.
- Ineffective factor: the factor that is not an effective factor.
- Between-batch noise: denoted by ε_b^k
- Within-batch noise: denoted by ε_{kn}^w
- SBPM-based analysis: The analysis based on SBPM
- Product-based analysis: ignores the dependency of the products within batch
- Batch-based analysis: batch level data
- Factor-assumed analysis : with a dummy factor

V. VALIDATION (cont'd)

- Three types of simulations to facilitate the comparison
- a) Numerical factor case
- b) Nominal factor case
- c) Mixed factor case (half-and-half

- 1000 factors containing 100 interested factors are considered
- In the 100 interested factors, 10 factors are randomly selected and set as the effect factors of effect 1.
- In the remaining 900 factors, 100 factors are randomly selected and set as the effective factors of effect τ , where τ of range 0.1 to 0.3 is a parameter to control the simulation scenario.

V. VALIDATION (cont'd)

- Pure noises are *i.i.d.* generated from a normal distribution with mean 0 and variance σ_w^2 , where σ_w^2 of range 0.2 to 5 is a parameter to ontrol the simulation scenario.

$$Y_{kn} = \beta_0 + \sum_{p=1}^{100} \beta_p X_{pkn} + \sum_{p=101}^{1000} \beta_p X_{pkn} + \varepsilon_{kn}^{w},$$

V. VALIDATION (cont'd)

- ***Validation of Single Factor Analysis***
- The factors with p-values less than the significant level α, set as 0.05 in this study, will be regarded as effective factors.

- (a) Type I error R_α
- (b) screening accuracy R_a

$$R_\alpha = \frac{p_s - p_c}{p_t - p_e},$$

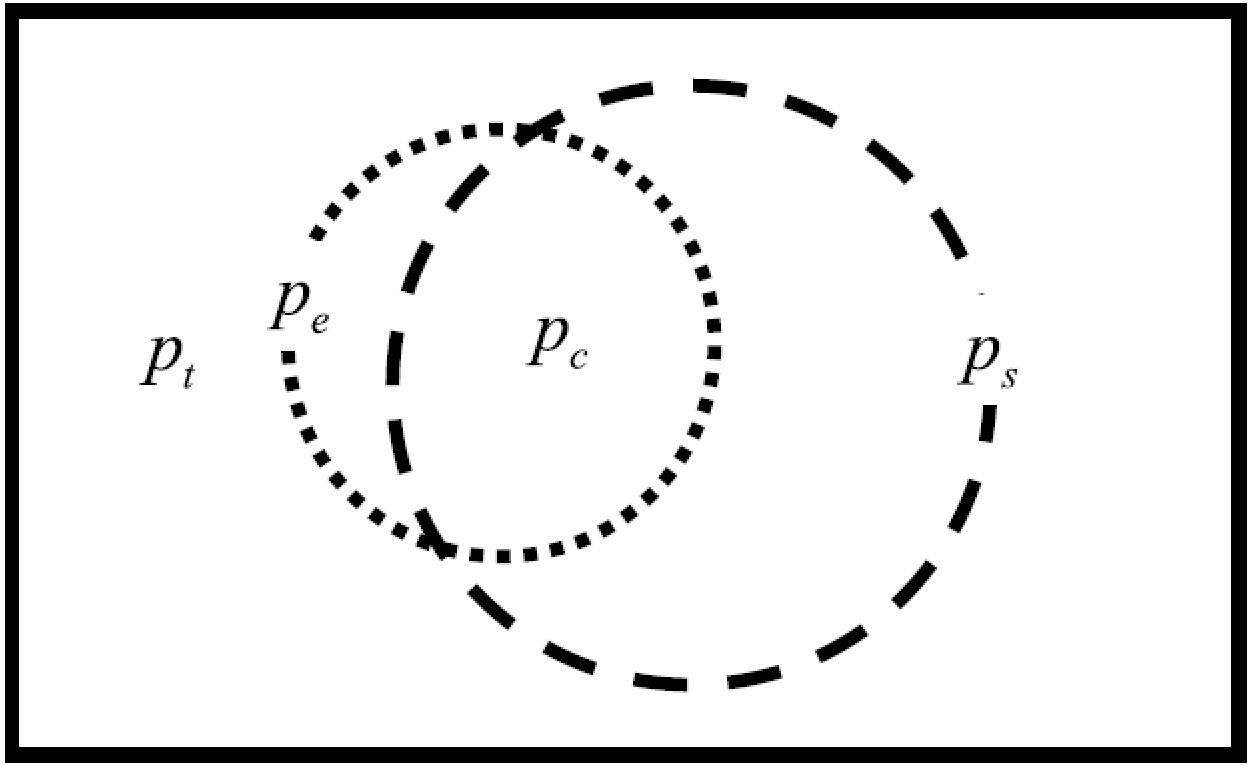
$$\mathrm{R_a} = \frac{2\mathrm{p_c} + \mathrm{p_t} - \mathrm{p_e} - \mathrm{p_s}}{\mathrm{p_t}},$$

V. VALIDATION (cont'd)

- total factors with size p_t
- effective factors with size p_e
- selected factors with size p_s
- catching factors with size p_c

$$R_\alpha = \frac{p_s - p_c}{p_t - p_e},$$

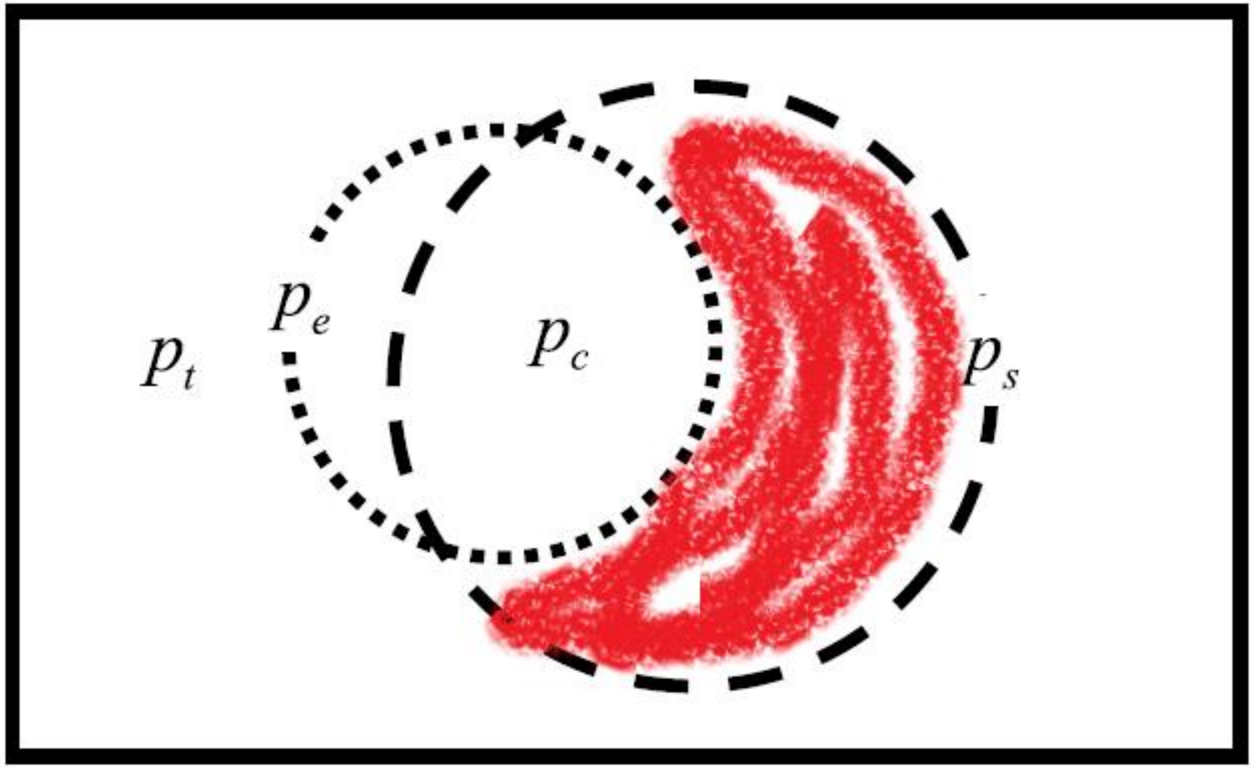
$$\mathrm{R_a} = \frac{2\mathrm{p_c} + \mathrm{p_t} - \mathrm{p_e} - \mathrm{p_s}}{\mathrm{p_t}},$$



V. VALIDATION (cont'd)

$$R_\alpha = \frac{p_s - p_c}{p_t - p_e},$$

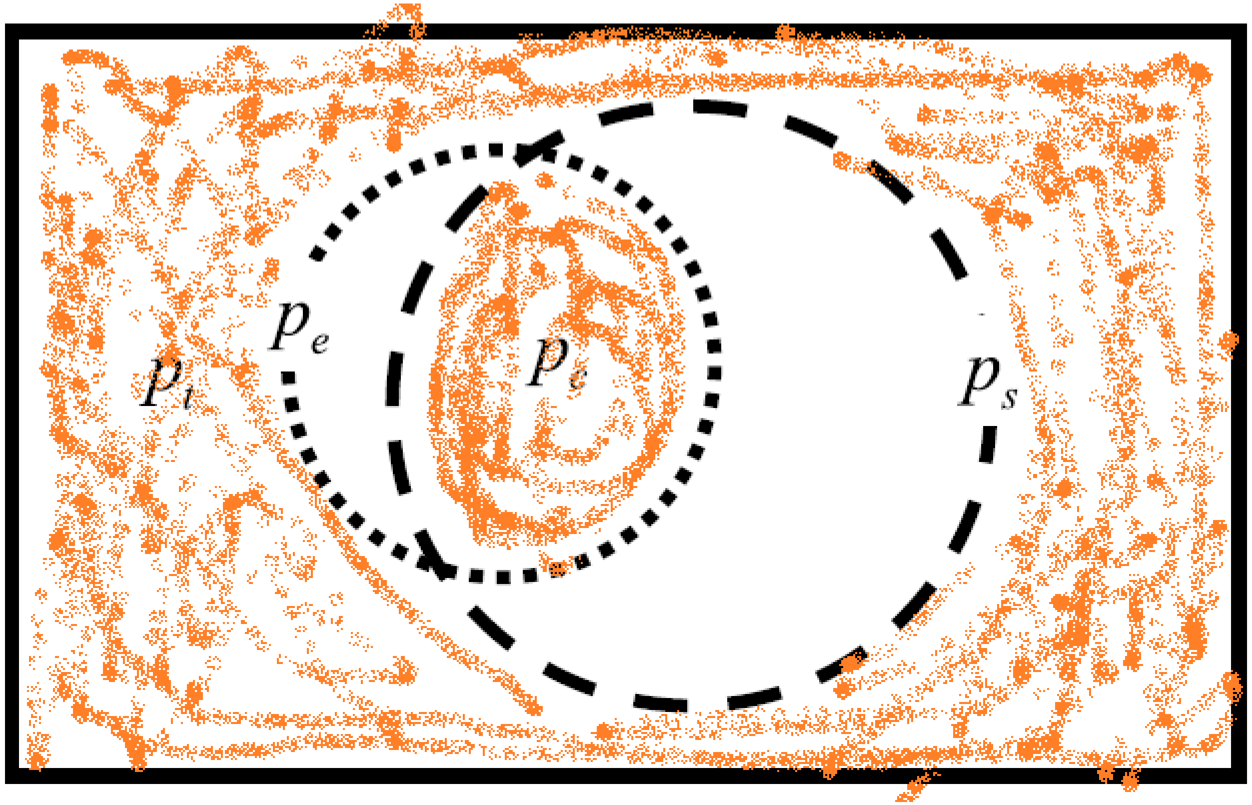
$$R_a = \frac{2p_c + p_t - p_e - p_s}{p_t},$$



V. VALIDATION (cont'd)

$$R_\alpha = \frac{p_s - p_c}{p_t - p_e},$$

$$R_a = \frac{2p_c + p_t - p_e - p_s}{p_t},$$



V. VALIDATION (cont'd)

- Type I error

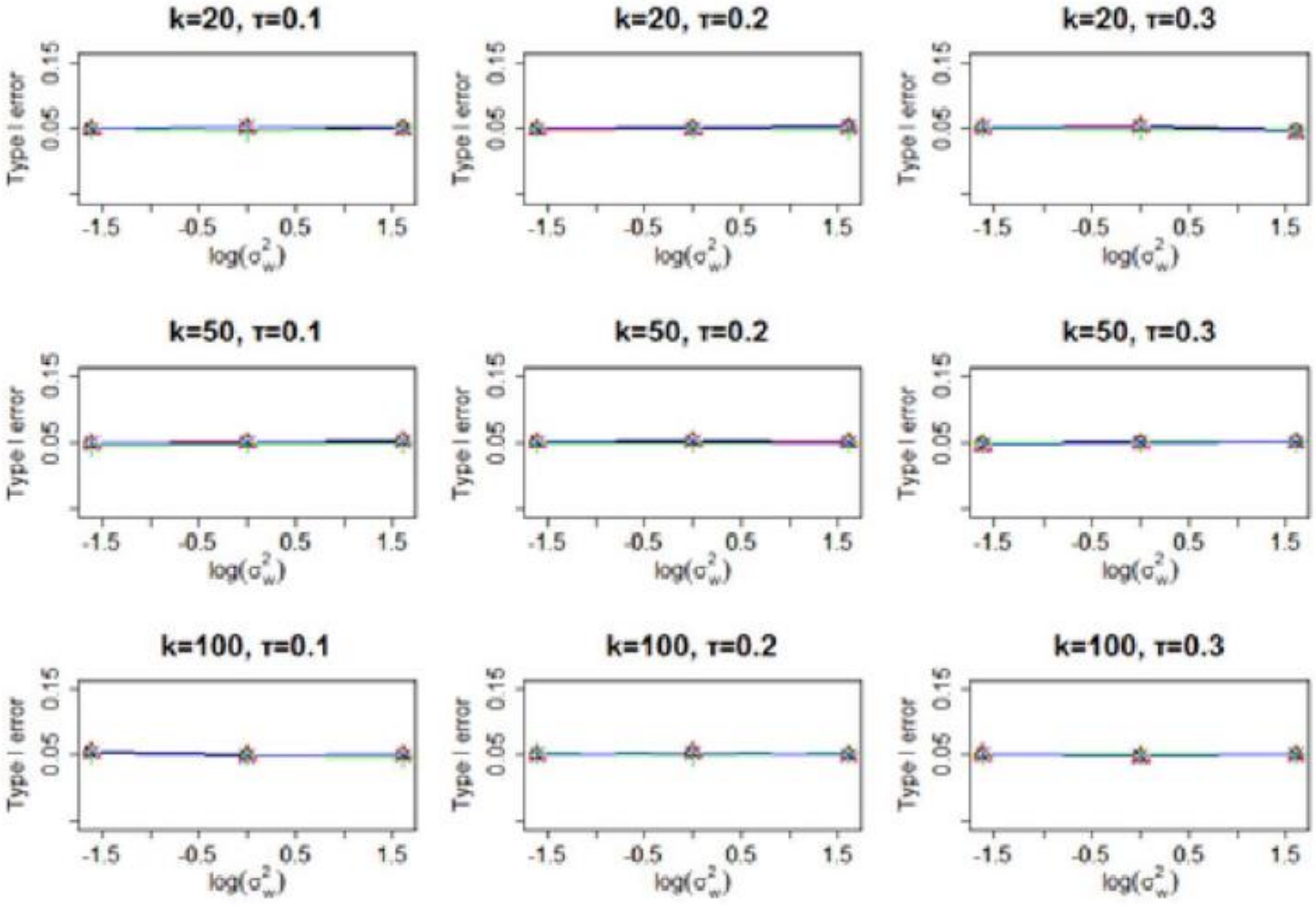
' □' SBPM

'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- numerical input case



V. VALIDATION (cont'd)

- Type I error

'□' SBPM

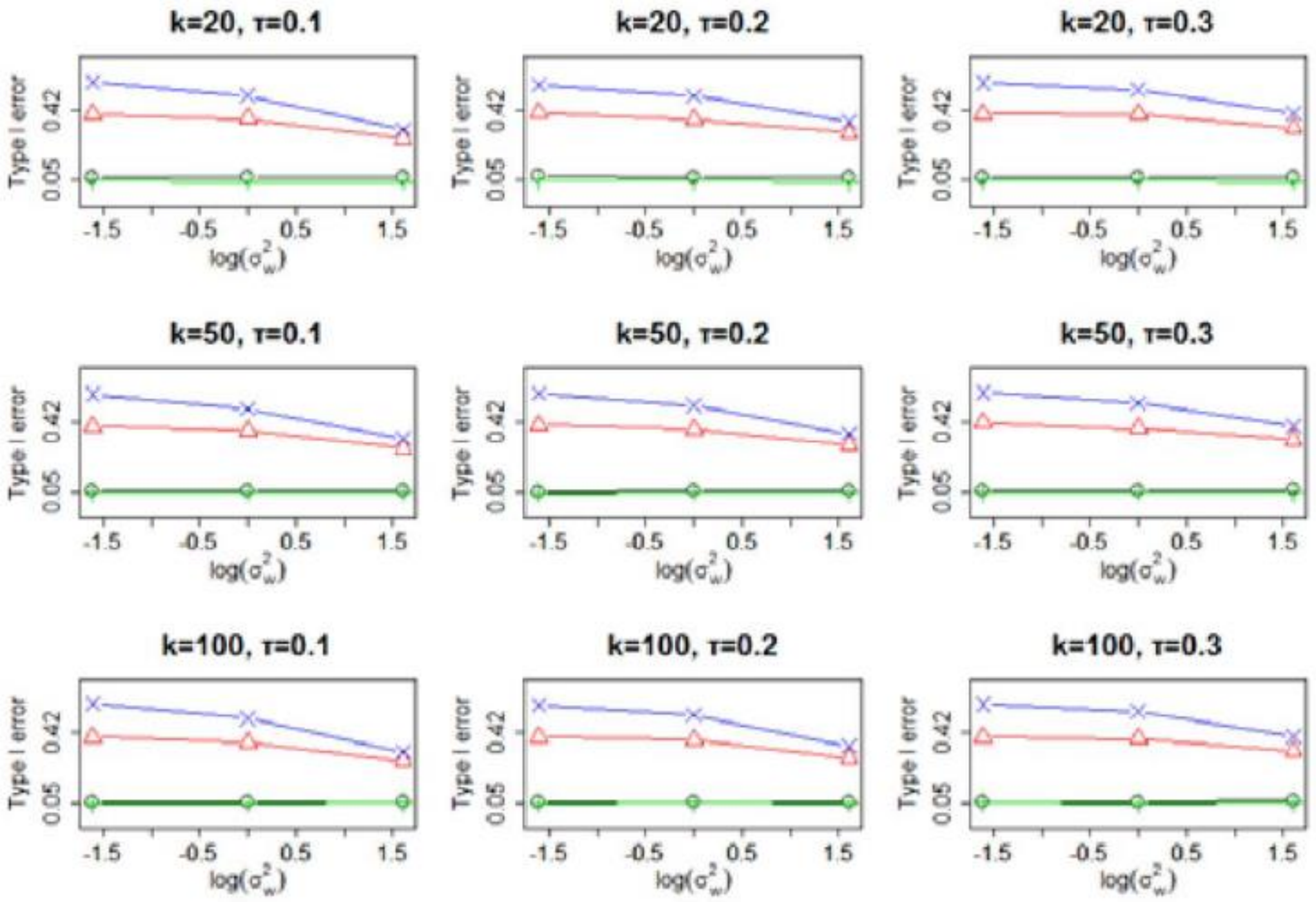
'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- nominal input case.

- $\zeta kn = \xi kn + \varepsilon kn$



V. VALIDATION (cont'd)

- Type I error

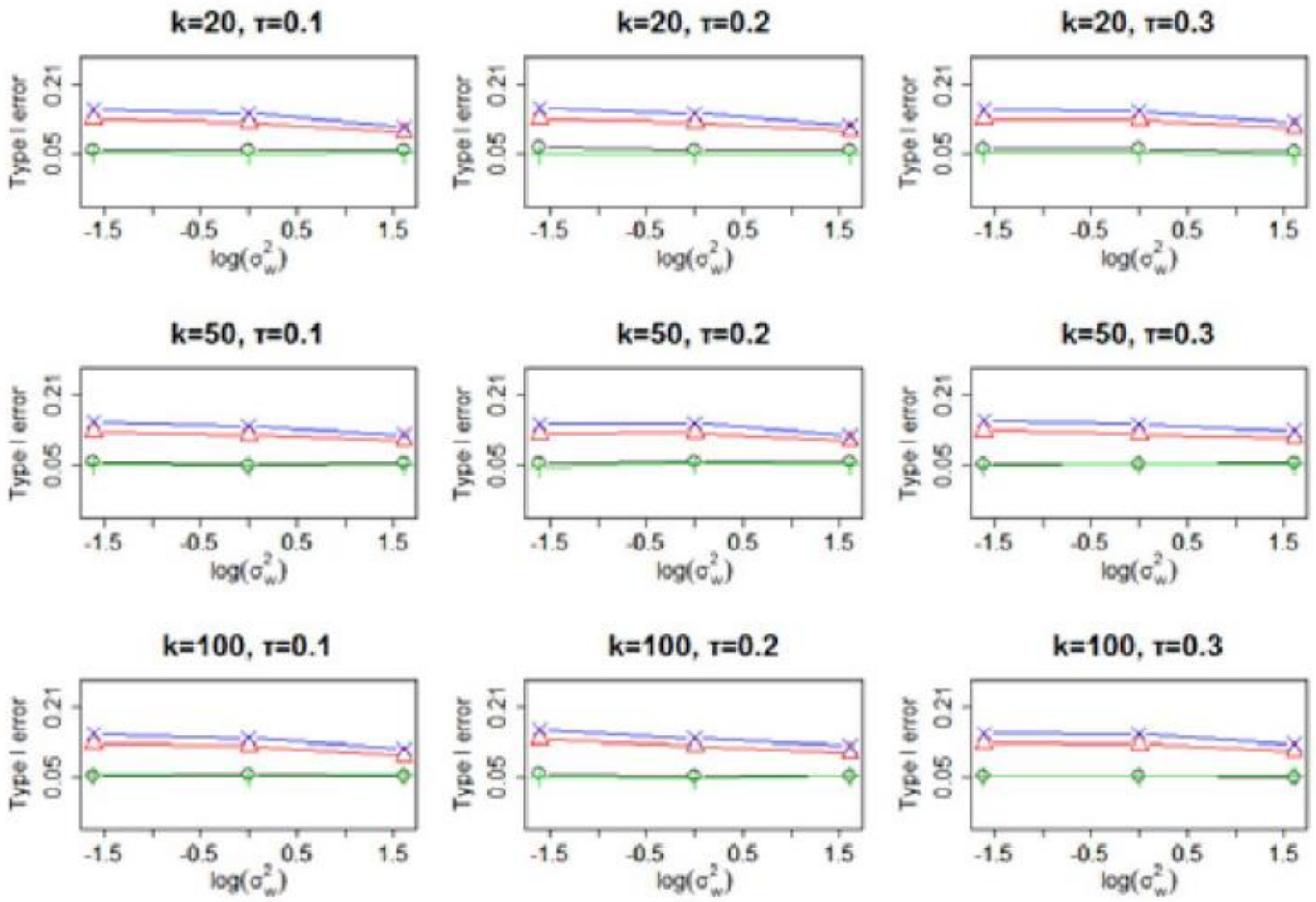
'□' SBPM

'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- mixed input case.



V. VALIDATION (cont'd)

- Accuracy

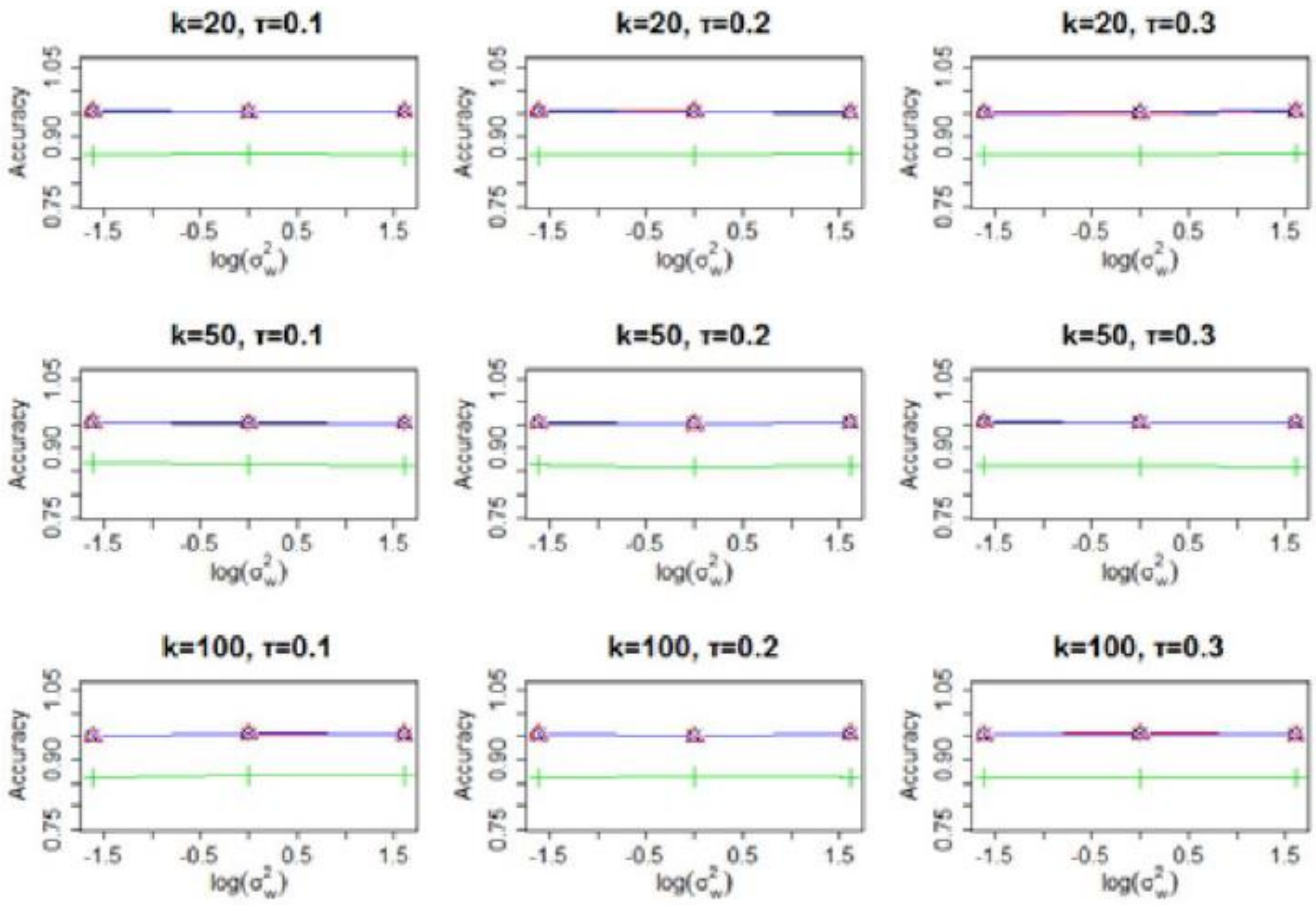
'□' SBPM

'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- numerical input case



V. VALIDATION (cont'd)

- Accuracy

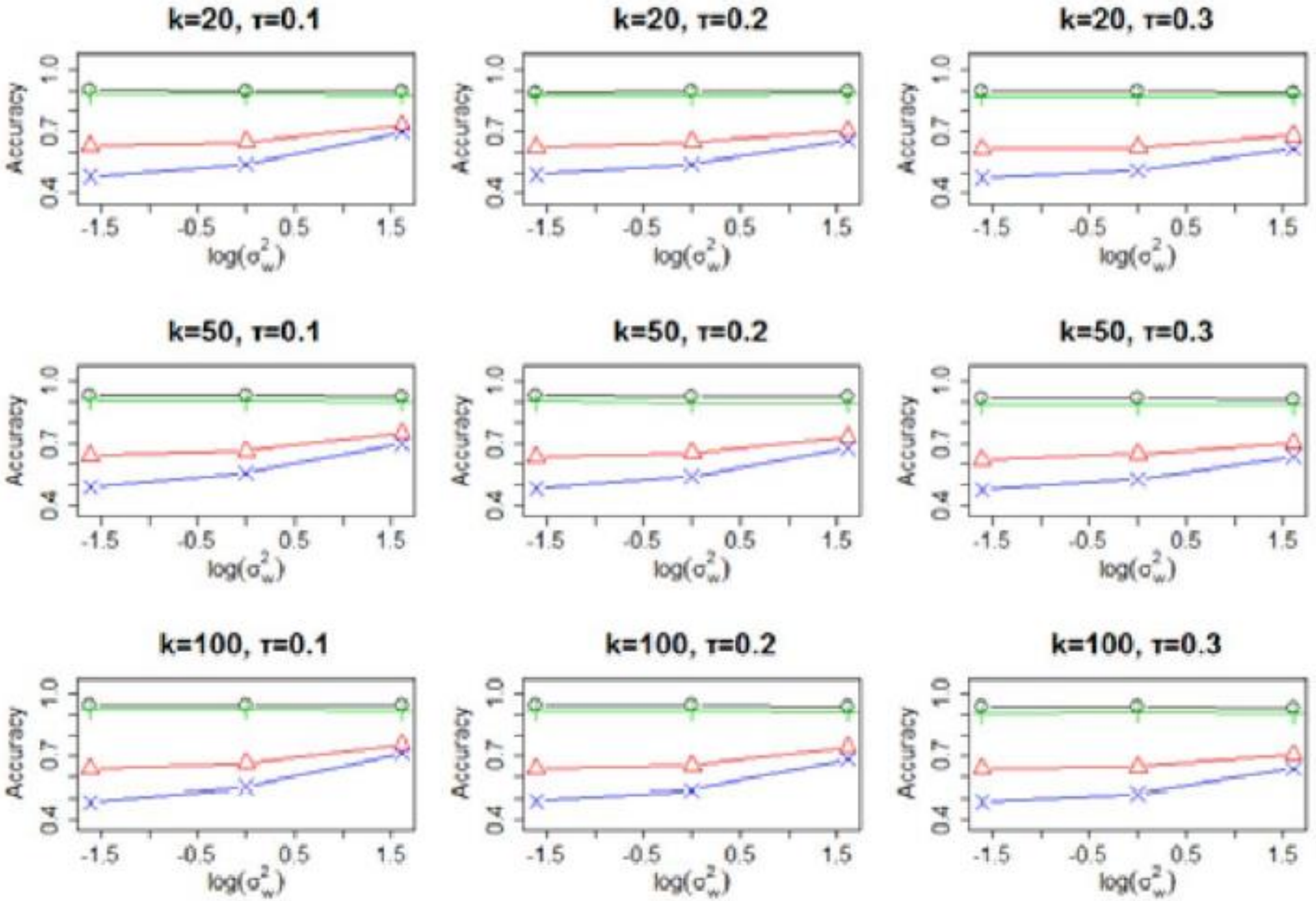
' □' SBPM

'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- nominal input case



V. VALIDATION (cont'd)

- Accuracy

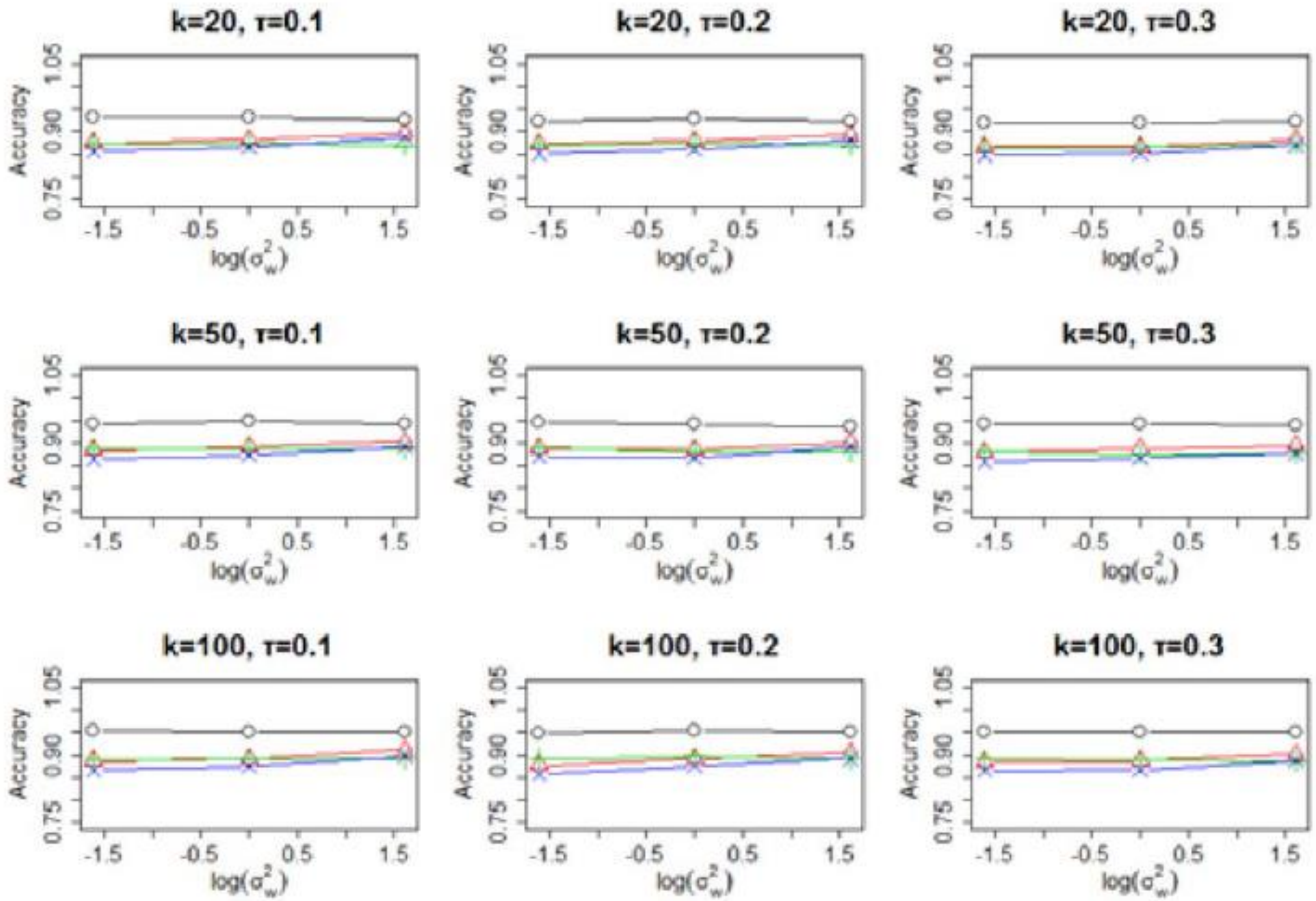
' □' SBPM

'Δ' product-based,

'+' batch-based,

'X' factor-assumed analyses

- mixed input case



V. VALIDATION (cont'd)

- *Validation of SBPM-Based Root Cause Detection Framework:*
- *RSSE* is a smaller-the-better measure.
- *βp* = 1, if *Fp* is an effective factor.
- *βp* = 0, if *Fp* is an ineffective factor.

$$R_{SSE} = \sum_{p=1}^{100} \widehat{\beta}_p - \beta_p,$$

V. VALIDATION (cont'd)

- *Validation*

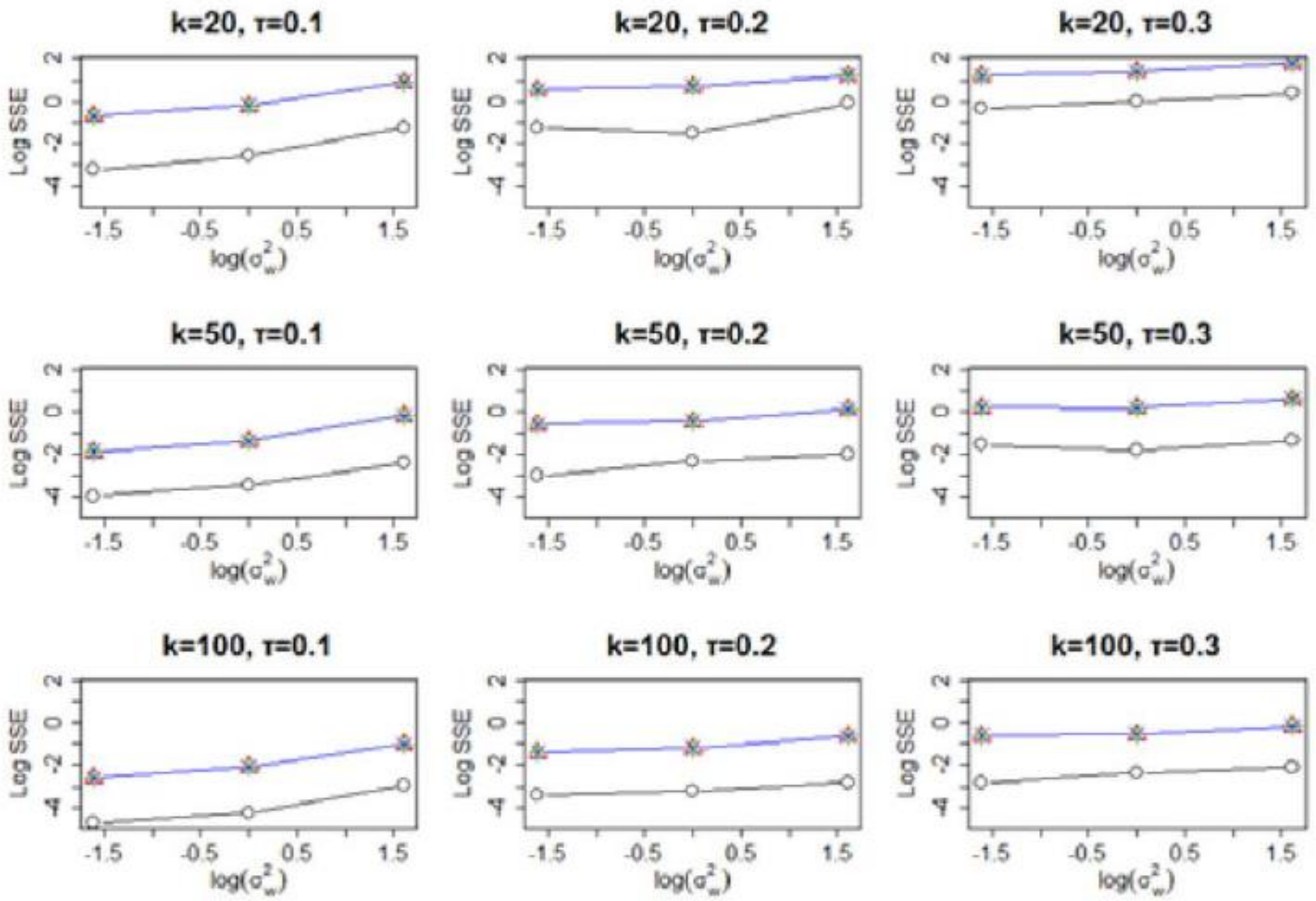
'□' SBPM-based root cause detection approach

'Δ' stepwise regression

'+' Lasso

'X' LARS

- numerical input case



V. VALIDATION (cont'd)

- *Validation*

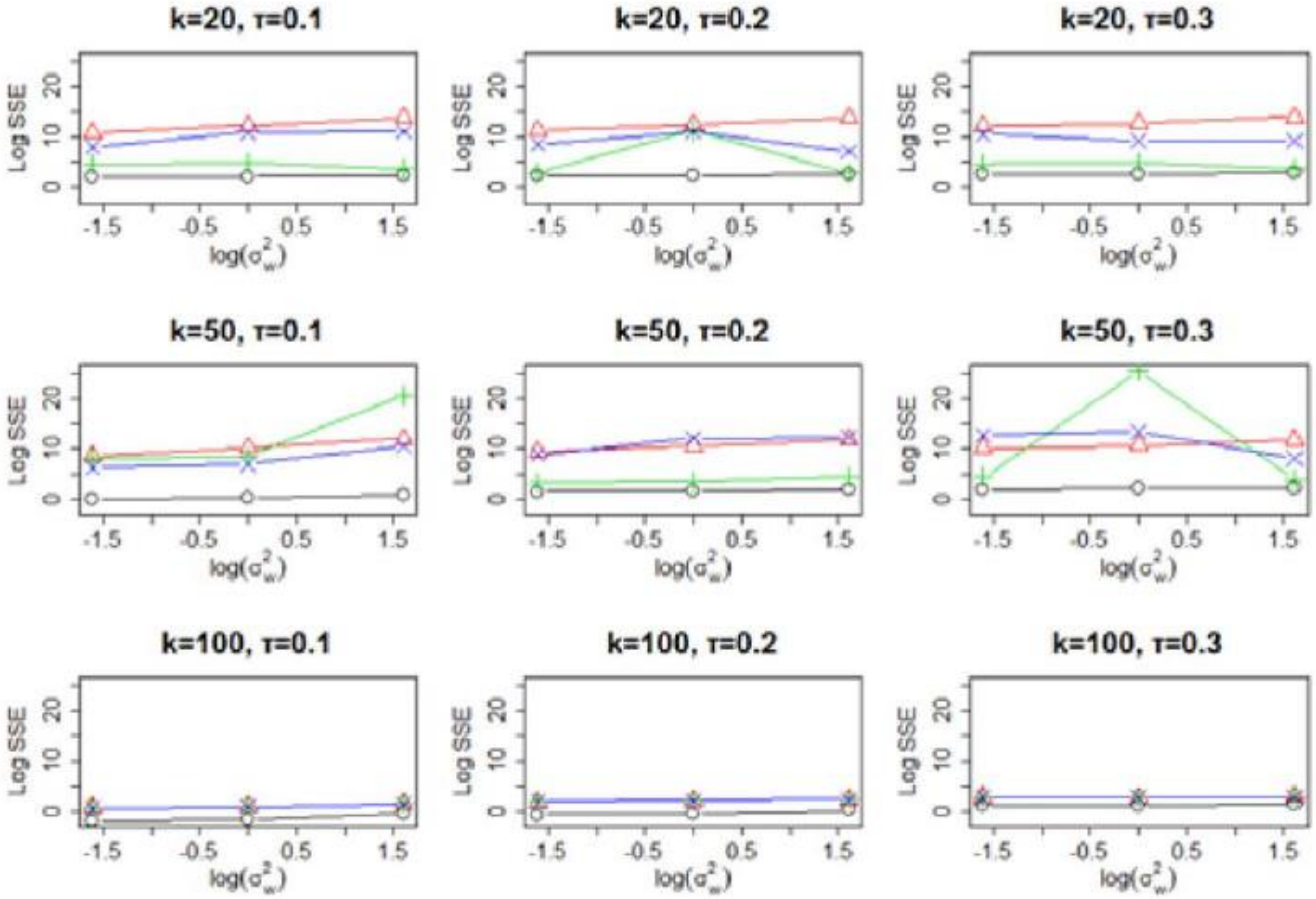
'□' SBPM-based root cause detection approach

'Δ' stepwise regression

'+' Lasso

'X' LARS

- nominal input case



V. VALIDATION (cont'd)

➢ *Validation*

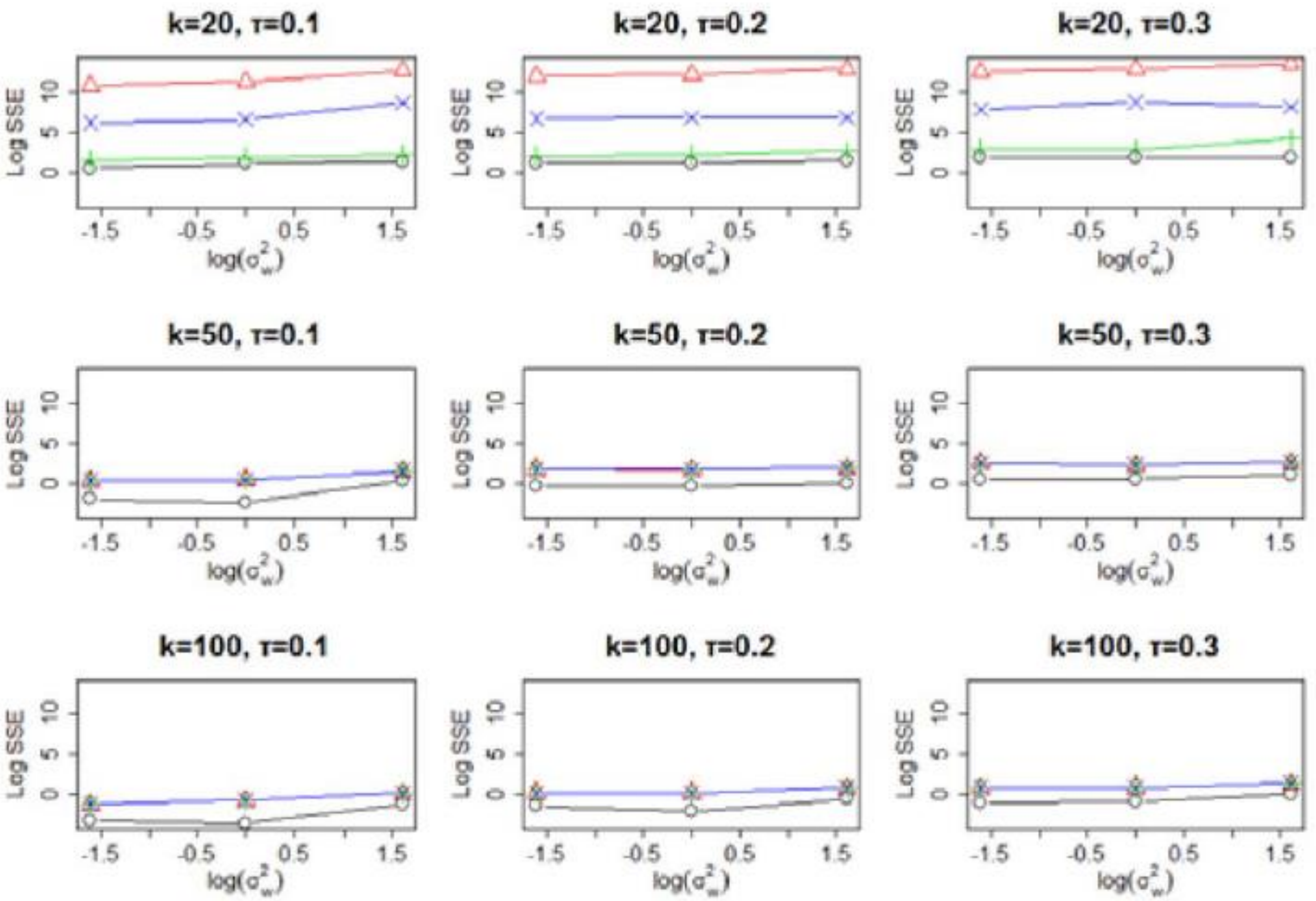
'□' SBPM-based root cause detection approach

'Δ' stepwise regression

'+' Lasso

'X' LARS

- mixed input case



V. VALIDATION (cont’d)

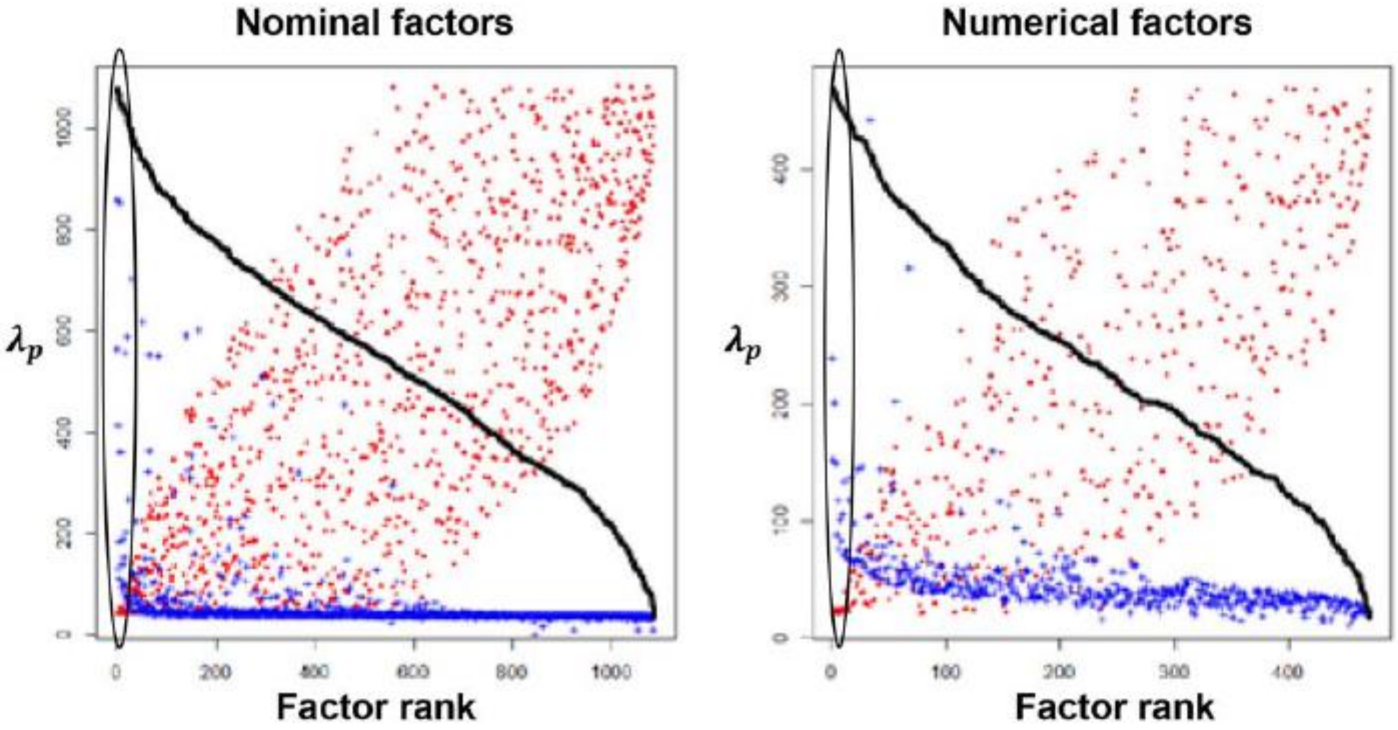
- *Empirical Study*
- *1) Problem Definition and Data Preparation*
- *2) Key Factor Screening:*
- *3) Model Construction and Evaluation*

V. VALIDATION (cont’d)

- ➢ *1) Problem Definition and Data Preparation*
- The response *Y* was the wafer CP yield,

V. VALIDATION (cont'd)

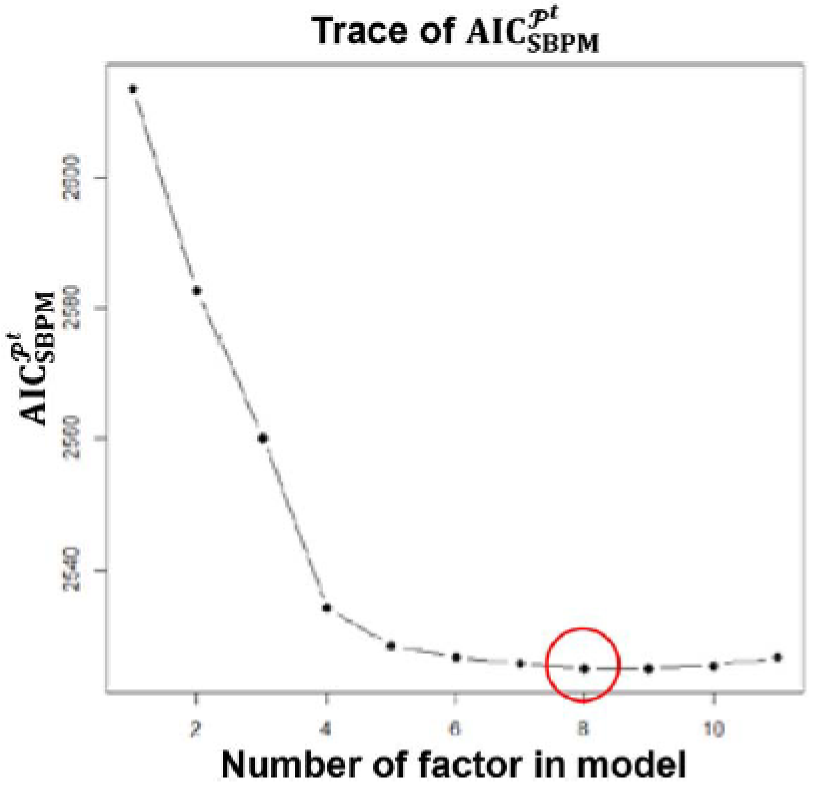
- *2) Key Factor Screening:*



V. VALIDATION (cont'd)

- *3) Model Construction and Evaluation*

$$\hat{Y}_{kn} = \beta_0 + \beta_1 S667\text{chrcp}_{kn} + \beta_2 S284\text{chid}_{kn} + \beta_3 S1230\text{chid}_{kn} + \beta_4 S1106\text{chrcp}_{kn} + \beta_5 WAT872_{kn} + \beta_6 WAT1067_{kn} + \beta_7 WAT99_{kn} + \beta_8 WAT1529_{kn},$$



V. VALIDATION (cont'd)

- *3) Model Construction and Evaluation*

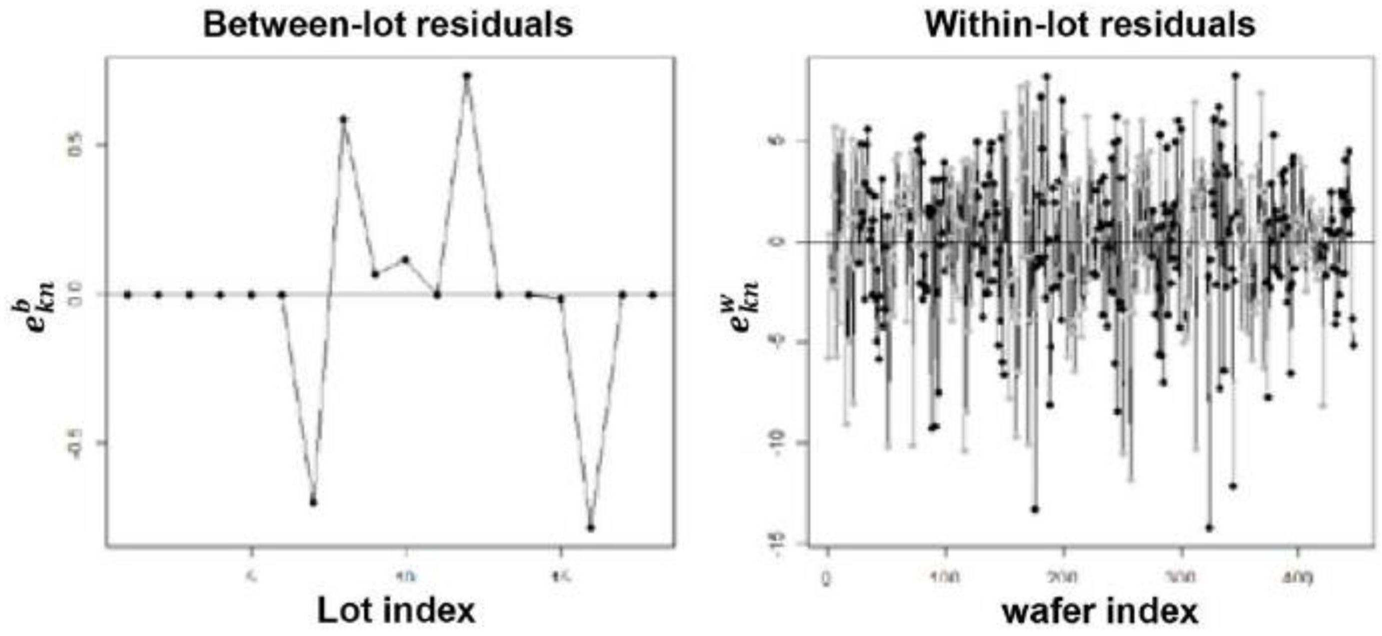


Fig. 21. Residuals. Left: between-lot residuals. Right: within-lot residuals. The colors of points are used to distinguish from lots.

V. VALIDATION (cont'd)

- *3) Model Construction and Evaluation*

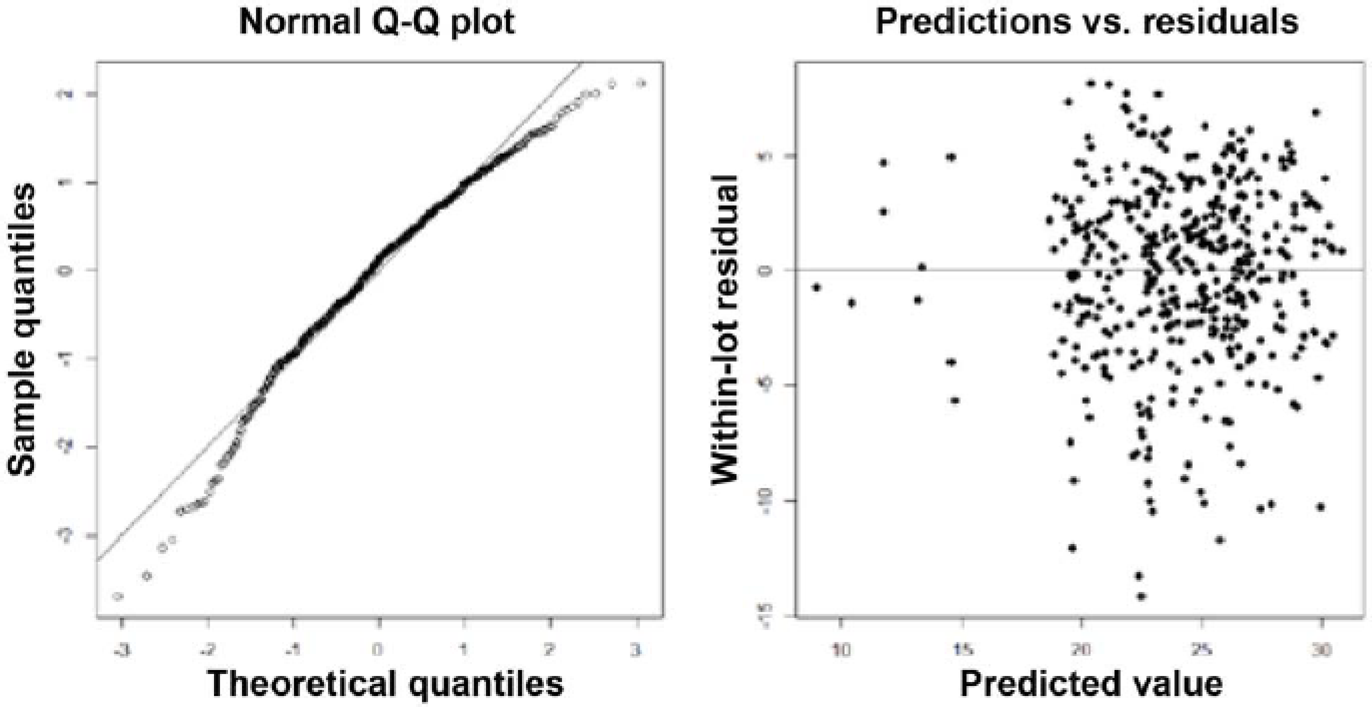


Fig. 22. Within-lot residuals diagnosis. Left: quintile–quentile plot of residuals of within-lot residuals. Right: expected values of response to residuals of within-lot residuals.

V. VALIDATION (cont'd)

- *3) Model Construction and Evaluation*

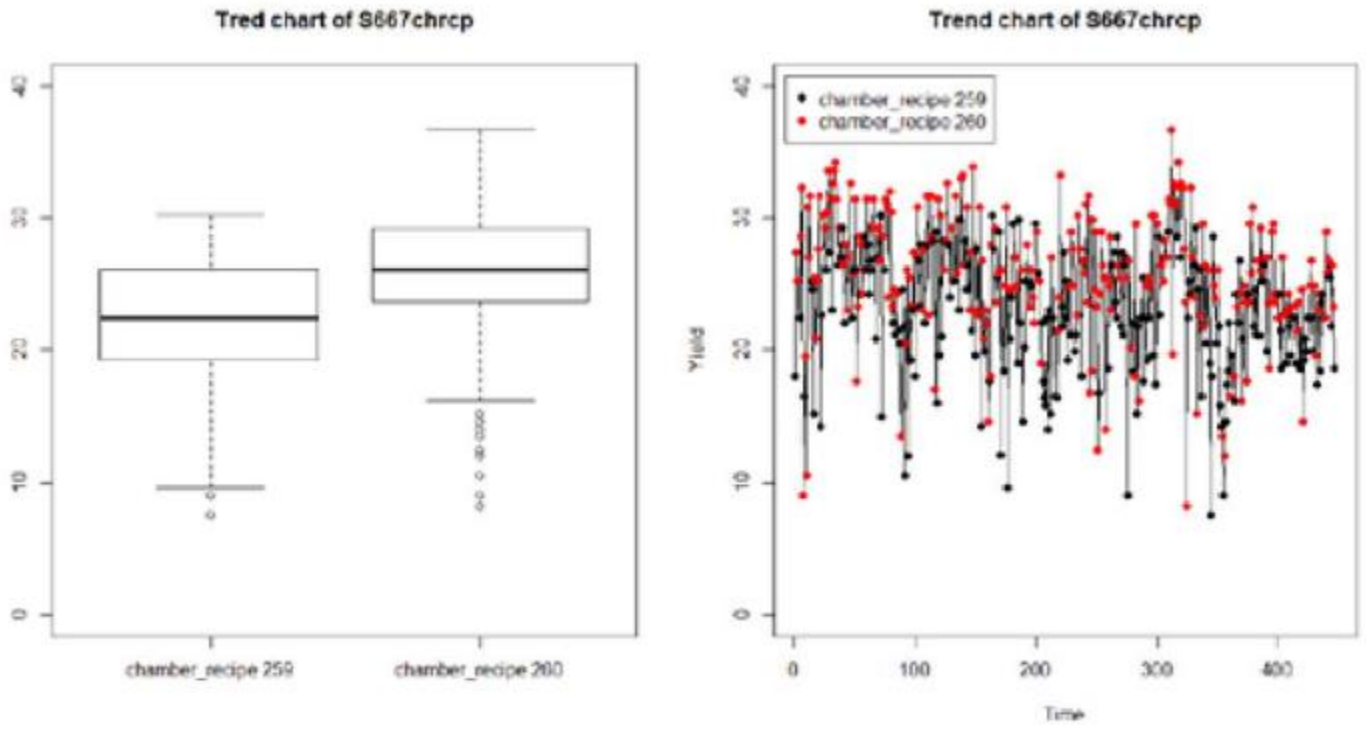


Fig. 23. Box plot and trend chart of S667chrcp.

V. VALIDATION (cont’d)

- *3) Model Cons*

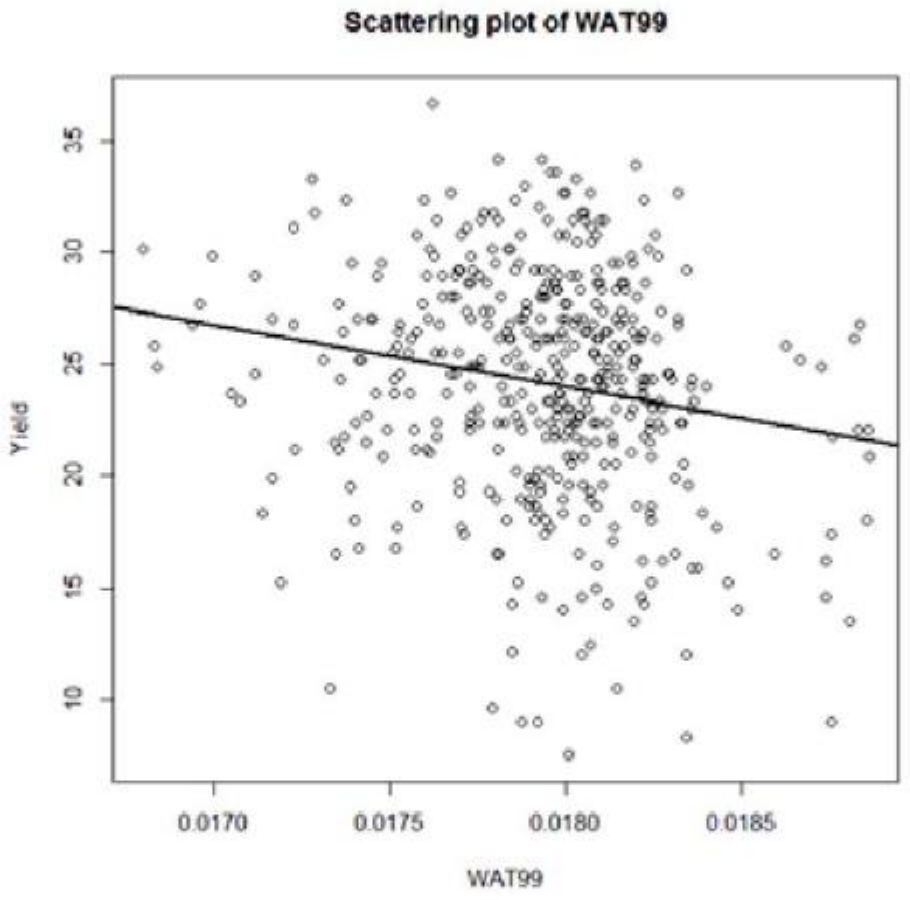


Fig. 24. Scattering plot of WAT99.

VI. CONCLUDING REMARKS

- This study proposed an effective framework to detect the root causes for sub-batch processing system for semiconductor manufacturing

- future study can be done to develop robust hypothesis testing when the sample size is small in the early stage of ramping advanced technology nodes.