Electronic Supplementary Material

Comparison of data driven and data-mechanism hybrid driven methods for key variables prediction based on datasets with different sample sizes and noises

Qihang Tan[1], Chao Wang (✉)[2], Wange Li[2], Jinghao Sun (✉)[1], Jun Zhao[2]

1 School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China

2 School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China

E-mails: chaowang688@163.com (Wang C); jhsun@dlut.edu.cn (Sun J)

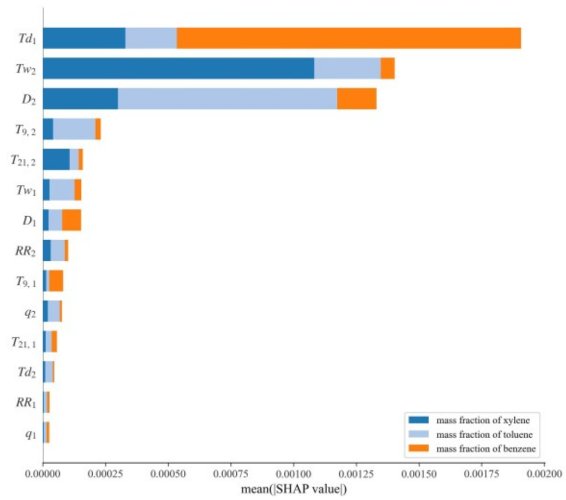

Fig. S1 Correlation analysis of case 1

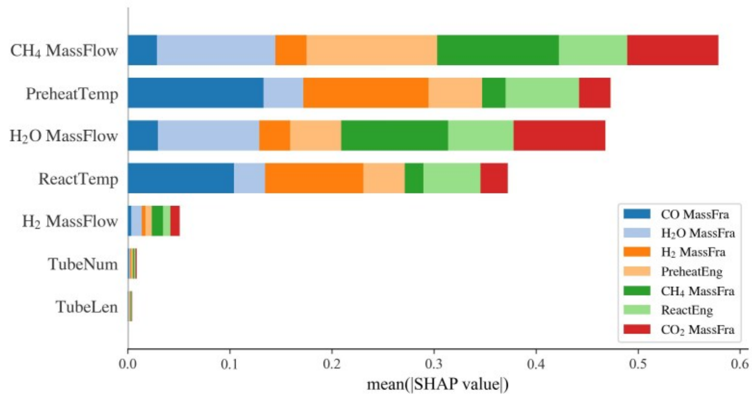

Fig. S2 Correlation analysis of case 2

Table S1 Input features selected of BTX distillation process

Name	Description	Name	Description
RR_1	The reflux ratio of C1	$T_{21,1}$	The temperature of the 21st plate in C1, ℃
D_1	The distillate rate of C1, kmol/h	q_1	The heat duty of reboiler in C1, W
RR_2	The reflux ratio of C2	Td_2	The top temperature of C2, ℃
D_2	The distillate rate of C2, kmol/h	Tw_2	The bottom temperature of C2, ℃
Td_1	The top temperature of C1, ℃	$T_{9,2}$	The temperature of the 9th plate in C2, ℃
Tw_1	The bottom temperature of C1, ℃	$T_{21,2}$	The temperature of the 21st plate in C2, ℃
$T_{9,1}$	The temperature of the 9th plate in C1, ℃	q_2	The heat duty of reboiler in C2, W

Table S2 Dataset size and noise intensity of BTX distillation process

Parameters	Details
Sample size	100, 200, 300, 400
Noise intensity	0.1, 0.15, 0.2

Table S3. Hyperparameters and optimization approach of BTX distillation process for different data-driven models

Models	Hyperparameters and optimization approach	Details
RF	Optimization approach	grid search
	n_estimators	from 10 to 1000, step 10
	max_feature	7 for BTX and 5 for SMR
XGBoost	Optimization approach	grid search
	n_estimators	from 10 to 500, step 10
	eta	from 0.01 to 0.3, step 0.02
	max_depth	from 3 to 7
	colsample_bytree	0.8
ANN	Optimization approach	Bayesian optimization
	learning rate	from 0.0001 to 0.01

batch size	from 5 to 40
α	from 0 to 1
β	from 0 to 1
hidden layer dimension	from 1 to 256

Table S4 “n _estimators” applied in benzene-toluene-xylene distillation process for random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	820	820	110	720
0.15	770	290	590	570
0.2	210	650	980	120

Table S5 “n _estimators” applied in benzene-toluene-xylene distillation process for method (a) containing random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	310	150	90	170
0.15	620	20	680	50
0.2	940	990	710	660

Table S6 “n _estimators” applied in benzene-toluene-xylene distillation process for method (b) containing random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	880	880	580	660
0.15	150	320	970	200
0.2	890	190	810	590

Table S7 “n _estimators” applied in benzene-toluene-xylene distillation process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	370	390	160	130
0.15	170	440	500	200
0.2	100	100	100	240

Table S8 “eta” applied in benzene-toluene-xylene distillation process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	0.05	0.03	0.05	0.13
0.15	0.13	0.03	0.01	0.05
0.2	0.31	0.17	0.07	0.03

Table S9 “max_depth” applied in benzene-toluene-xylene distillation process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	3	3	3	3
0.15	3	3	3	3
0.2	3	3	3	5

Table S10 “n _estimators” applied in benzene-toluene-xylene distillation process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	290	500	260	210
0.15	100	100	160	120
0.2	420	100	410	480

Table S11 “eta” applied in benzene-toluene-xylene distillation process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	0.03	0.01	0.03	0.03
0.15	0.05	0.07	0.03	0.05
0.2	0.01	0.05	0.01	0.01

Table S12 “max_depth” applied in benzene-toluene-xylene distillation process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	3	3	3	3
0.15	3	3	3	3
0.2	3	3	3	3

Table S13 “n _estimators” applied in benzene-toluene-xylene distillation process for method (b) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	100	300	190	160
0.15	490	490	100	230
0.2	340	100	450	180

Table S14 “eta” applied in benzene-toluene-xylene distillation process for method (b) containing extreme gradient boosting

Noise intensity	Sample size

	100	200	300	400
0.1	0.13	0.07	0.05	0.05
0.15	0.07	0.07	0.05	0.03
0.2	0.03	0.05	0.01	0.03

Table S15 "max_depth" applied in benzene-toluene-xylene distillation process for method (b) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	4	3	4	3
0.15	3	3	4	3
0.2	3	4	3	3

Table S16 Hidden layer dimension applied in benzene-toluene-xylene distillation process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[237, 34, 19]	[43, 206, 169]	[196, 151, 187]	[15, 104, 32]
0.15	[42, 77, 189]	[71, 183, 51]	[155, 110, 111]	[63, 157, 251]
0.2	[96, 97, 27]	[147, 58, 210]	[232, 226, 127]	[163, 108, 125]

Table S17 Learning rate applied in benzene-toluene-xylene distillation process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0033	0.0067	0.002	0.0002
0.15	0.0067	0.0057	0.0025	0.0016
0.2	0.0042	0.0053	0.001	0.0048

Table S18 Batch size applied in benzene-toluene-xylene distillation process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	6	37	38	21
0.15	27	11	37	21
0.2	13	18	9	29

Table S19 Hidden layer dimension applied in benzene-toluene-xylene distillation process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[59, 12, 255]	[45, 65, 101]	[27, 32, 8]	[90, 62, 92]
0.15	[202, 240, 210]	[143, 154, 54]	[13, 94, 86]	[114, 48, 147]
0.2	[126, 39, 54]	[118, 167, 109]	[30, 40, 106]	[30, 83, 104]

Table S20 Learning rate applied in benzene-toluene-xylene distillation process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0002	0.0001	0.0004	0.0001
0.15	0.0018	0.0001	0.0001	0.0001
0.2	0.0061	0.006	0.009	0.0003

Table S21 Batch size applied in benzene-toluene-xylene distillation process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	22	28	24	38
0.15	24	29	31	23
0.2	34	23	40	9

Table S22 Hidden layer dimension applied in benzene-toluene-xylene distillation process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[230, 159, 123]	[71, 118, 137]	[154, 192, 124]	[107, 246, 130]
0.15	[64, 145, 185]	[49, 8, 87]	[163, 139, 210]	[35, 202, 123]
0.2	[11, 92, 158]	[169, 55, 140]	[180, 120, 133]	[172, 247, 241]

Table S23 Learning rate applied in benzene-toluene-xylene distillation process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0027	0.006	0.0009	0.0008
0.15	0.0011	0.0007	0.0009	0.0046
0.2	0.0064	0.0006	0.0054	0.001

Table S24 Batch size applied in benzene-toluene-xylene distillation process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	17	20	21	26
0.15	26	39	16	28
0.2	13	20	27	20

Table S25 Hidden layer dimension applied in benzene-toluene-xylene distillation process for method (c) containing artificial neural network

Noise intensity	Sample size

	100	200	300	400
0.1	[108, 67, 99]	[197, 142, 58]	[212, 248, 141]	[9, 116, 9]
0.15	[73, 228, 218]	[64, 124, 134]	[188, 123, 176]	[93, 158, 64]
0.2	[238, 52, 122]	[21, 216, 69]	[155, 99, 94]	[68, 142, 221]

Table S26 Learning rate applied in benzene-toluene-xylene distillation process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0059	0.0096	0.0022	0.0002
0.15	0.0017	0.007	0.0063	0.0028
0.2	0.0037	0.0071	0.0023	0.0007

Table S27 Batch size applied in benzene-toluene-xylene distillation process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	15	26	30	6
0.15	20	31	22	32
0.2	20	27	26	24

Table S28 Alpha applied in benzene-toluene-xylene distillation process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.06	0.13	0.11	0.19
0.15	0.68	0.77	0.56	0.76
0.2	0.98	0.72	0.86	0.88

Table S28 Hidden layer dimension applied in benzene-toluene-xylene distillation process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[124, 23, 175]	[89, 175, 209]	[194, 176, 159]	[85, 107, 21]
0.15	[249, 209, 178]	[210, 256, 180]	[183, 208, 230]	[206, 240, 234]
0.2	[39, 173, 254]	[164, 37, 97]	[153, 240, 26]	[133, 228, 157]

Table S30 Learning rate applied in benzene-toluene-xylene distillation process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0019	0.0093	0.0013	0.0001
0.15	0.0027	0.0008	0.0017	0.001

0.2	0.0016	0.0085	0.0027	0.0008

Table S31 Batch size applied in benzene-toluene-xylene distillation process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	8	10	27	36
0.15	22	5	36	14
0.2	24	9	40	21

Table S32 Alpha applied in benzene-toluene-xylene distillation process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.1	0.97	0.42	0.61
0.15	0.98	0.59	0.53	0.44
0.2	0.35	0.02	0.07	0.58

Table S33 Parameters for kinetic terms and equilibrium constants of SMR process

reaction i	$A(k_i)$	E_{ai}/kJ·mol^{-1}	A_i	B_i
1	1.174×10^{15}mol·bar$^{0.5}$/(kg$_{cat}$·s)	240.1	-26830	30.136
2	5.4306×10^{5}mol/(kg$_{cat}$·s·bar)	67.13	4400	-4.063
3	2.833×10^{14}mol·bar$^{0.5}$/(kg$_{cat}$·s)	243.9	-22430	26.073

Table S34 Parameters for adsorption terms of SMR process

Component j	$A(A_j)$	$\Delta H_{A,J}$/kJ·mol^{-1}
CO	8.23×10^{-5} bar^{-1}	-70.65
H_2	6.12×10^{-9} bar^{-1}	-82.9
CH_4	6.65×10^{-4} bar^{-1}	-38.28
H_2O	1.77×10^{5}	88.68

Table S35 Input and output features selected for SMR process

Input features	Description	Output features	Description
CH_4 MassFlow	The mass flow rate of CH_4, kg/h	H_2O MassFra	The mass fraction of H_2O after reaction
H_2O MassFlow	The mass flow rate of H_2O, kg/h	CO_2 MassFra	The mass fraction of CO_2 after reaction
H_2 MassFlow	The mass flow rate of H_2, kg/h	CO MassFra	The mass fraction of CO after reaction

PreheatTemp	The preheat temperature, K	H_2 MassFra	The mass fraction of H_2 after reaction
ReactTemp	The react temperature, K	CH_4 MassFra	The mass fraction of CH_4 after reaction
TubeNum	The number of tubes in reactor	PreheatEng	Heat duty of preheating, W
TubeLen	The length of tubes in reactor, m	ReactEng	Heat duty of reacting, W

Table S36 “n_estimators” applied in Steam Methane Reforming process for random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	70	90	130	210
0.15	850	220	810	200
0.2	130	860	990	230

Table S37 “n_estimators” applied in Steam Methane Reforming process for method (a) containing random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	940	730	260	200
0.15	960	210	90	620
0.2	540	920	280	410

Table S38 “n_estimators” applied in Steam Methane Reforming process for method (b) containing random forest

Noise intensity	Sample size			
	100	200	300	400
0.1	290	330	970	40
0.15	330	540	850	570
0.2	90	980	440	80

Table S39 “n_estimators” applied in Steam Methane Reforming process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	500	290	150	130
0.15	330	500	250	210
0.2	100	500	500	500

Table S40 “eta” applied in Steam Methane Reforming process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	0.03	0.03	0.07	0.07
0.15	0.03	0.01	0.03	0.03
0.2	0.21	0.01	0.01	0.01

Table S41 “max_depth” applied in Steam Methane Reforming process for extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	3	3	3	3
0.15	3	3	3	3
0.2	3	3	3	3

Table S42 “n_estimators” applied in Steam Methane Reforming process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	390	400	450	460
0.15	460	130	500	350
0.2	200	350	340	370

Table S43 “eta” applied in Steam Methane Reforming process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	0.01	0.01	0.01	0.01
0.15	0.01	0.03	0.01	0.01
0.2	0.01	0.01	0.01	0.01

Table S44 “max_depth” applied in Steam Methane Reforming process for method (a) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	3	3	3	3
0.15	4	3	3	3
0.2	3	3	3	3

Table S45 “n_estimators” applied in Steam Methane Reforming process for method (b) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	240	100	100	500
0.15	100	500	500	130
0.2	110	150	380	120

Table S46 “eta” applied in Steam Methane Reforming process for method (b) containing extreme gradient boosting

Noise intensity	Sample size

	100	200	300	400
0.1	0.03	0.05	0.13	0.01
0.15	0.11	0.01	0.01	0.03
0.2	0.03	0.03	0.01	0.03

Table S47 “max_depth” applied in Steam Methane Reforming process for method (b) containing extreme gradient boosting

Noise intensity	Sample size			
	100	200	300	400
0.1	3	3	3	3
0.15	3	3	3	3
0.2	3	3	3	3

Table S48 Hidden layer dimension applied in Steam Methane Reforming process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[97, 8, 157]	[211, 12, 137]	[200, 144, 103]	[134, 124, 194]
0.15	[210, 38, 253]	[214, 29, 18]	[164, 16, 158]	[49, 85, 69]
0.2	[190, 41, 106]	[205, 227, 187]	[19, 198, 159]	[182, 94, 229]

Table S49 Learning rate applied in Steam Methane Reforming process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.002	0.0025	0.0012	0.0019
0.15	0.0016	0.0001	0.0009	0.0001
0.2	0.0016	0.0067	0.001	0.0005

Table S50 Batch size applied in Steam Methane Reforming process for artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	14	18	31	16
0.15	32	36	16	26
0.2	35	20	13	13

Table S51 Hidden layer dimension applied in Steam Methane Reforming process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[236, 253, 124]	[101, 62, 96]	[21, 16, 151]	[88, 55, 219]
0.15	[149, 91, 190]	[179, 67, 153]	[41, 103, 59]	[56, 85, 27]
0.2	[91, 39, 77]	[203, 12, 158]	[29, 226, 18]	[13, 47, 195]

Table S52 Learning rate applied in Steam Methane Reforming process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0029	0.0003	0.0001	0.0001
0.15	0.0056	0.0001	0.0003	0.0004
0.2	0.001	0.0001	0.0009	0.0004

Table S53 Batch size applied in Steam Methane Reforming process for method (a) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	18	26	12	39
0.15	14	28	35	12
0.2	35	26	19	22

Table S54 Hidden layer dimension applied in Steam Methane Reforming process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[234, 153, 104]	[144, 52, 234]	[220, 80, 15]	[150, 214, 132]
0.15	[159, 121, 90]	[118, 237, 148]	[190, 106, 148]	[8, 21, 33]
0.2	[223, 149, 128]	[30, 70, 48]	[168, 220, 21]	[255, 30, 215]

Table S55 Learning rate applied in Steam Methane Reforming process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0011	0.0048	0.0001	0.0035
0.15	0.0011	0.0014	0.0068	0.0007
0.2	0.0023	0.008	0.005	0.0014

Table S56 Batch size applied in Steam Methane Reforming process for method (b) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	12	21	15	29
0.15	13	16	21	18
0.2	19	13	17	5

Table S57 Hidden layer dimension applied in Steam Methane Reforming process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400

0.1	[255, 139, 256]	[250, 113, 242]	[204, 35, 95]	[223, 232, 151]
0.15	[238, 237, 112]	[72, 116, 29]	[116, 205, 77]	[213, 76, 171]
0.2	[6, 239, 248]	[247, 46, 223]	[55, 195, 155]	[172, 18, 146]

Table S58 Learning rate applied in Steam Methane Reforming process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0082	0.0083	0.008	0.0007
0.15	0.0046	0.0056	0.0002	0.0006
0.2	0.0035	0.0034	0.001	0.0016

Table S59 Batch size applied in Steam Methane Reforming process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	21	28	23	15
0.15	28	27	40	11
0.2	31	36	15	14

Table S60 Alpha applied in Steam Methane Reforming process for method (c) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.82	0.43	0.65	0.2
0.15	0.98	0.51	0.35	0.61
0.2	0.92	0.59	0.77	0.8

Table S61 Hidden layer dimension applied in Steam Methane Reforming process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	[31, 223, 58]	[182, 126, 99]	[211, 97, 146]	[125, 95, 221]
0.15	[22, 112, 155]	[178, 59, 124]	[20, 22, 160]	[27, 124, 69]
0.2	[245, 220, 161]	[22, 250, 21]	[52, 184, 224]	[215, 95, 113]

Table S62 Learning rate in Steam Methane Reforming process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0064	0.003	0.0048	0.0004
0.15	0.0024	0.0015	0.0005	0.0026
0.2	0.006	0.0002	0.009	0.0017

Table S63 Batch size in Steam Methane Reforming process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	21	25	15	7
0.15	10	10	38	21
0.2	34	34	7	29

Table S64 Alpha in Steam Methane Reforming process for method (d) containing artificial neural network

Noise intensity	Sample size			
	100	200	300	400
0.1	0.69	0.53	0.54	0.49
0.15	0.9	0.53	0.33	0.52
0.2	0.86	0.46	0.79	0.61

Table S65 Hidden layer dimension in Steam Methane Reforming process for double mechanism-data hybrid driven method

Noise intensity	Sample size			
	100	200	300	400
0.1	[232, 184, 41]	[235, 154, 130]	[149, 218, 129]	[237, 27, 120]
0.15	[74, 58, 193]	[186, 67, 219]	[56, 8, 166]	[106, 91, 36]
0.2	[13, 193, 218]	[159, 65, 250]	[88, 182, 45]	[10, 87, 35]

Table S66 Learning rate in Steam Methane Reforming process for double mechanism-data hybrid driven method

Noise intensity	Sample size			
	100	200	300	400
0.1	0.0044	0.0004	0.01	0.0004
0.15	0.009	0.001	0.0084	0.0037
0.2	0.0015	0.0019	0.0053	0.0014

Table S67 Batch size in Steam Methane Reforming process for double mechanism-data hybrid driven method

Noise intensity	Sample size			
	100	200	300	400
0.1	25	19	33	35
0.15	18	34	26	27
0.2	37	38	21	25

Table S68 Alpha in Steam Methane Reforming process for double mechanism-data hybrid driven method

Noise intensity	Sample size			
	100	200	300	400
0.1	0.77	0.17	0.34	0.42
0.15	0.69	0.3	0.5	0.67
0.2	0.8	0.56	0.42	0.39

Table S69 Beta in Steam Methane Reforming process for double mechanism-data hybrid driven method

Noise intensity	Sample size			
	100	200	300	400
0.1	0.09	0.55	0.12	0.27
0.15	0.23	0.03	0.002	0.07
0.2	0.45	0.01	0.19	0.05