

Identification of a servo-motor model from experimental data

By

Yue Dang

Thesis submitted in Partial Fulfillment of the Requirements for the
Degree of

Master of Science

in the Automotive Engineering

College of Mechanical, Aerospace, Automotive and Production
Engineering

Guide by Prof. Carlo Novara

2019

Politecnico di Torino

Acknowledgement

I would like to express my great appreciation to my professor Carlo Novara, that during the long run of this thesis, he guides me with a great patient.

To my wife and to my parents, great thanks for the support and encouragement during the past hard years.

Contents

Abstract	4
1 Introduction	4
2 System Identification method applied to simulated servo-motor system	8
3 System identification method applied to the real measurement.....	16
4 Conclusion	72
5 Reference	73
6 Appendix	74

Abstract

The thesis applied a system identification method named nonlinear Hammerstein-Wiener to identify a servo-motor system based on the limited measurement of input current and the output displacement. The servo-motor is a servo motor embedded in the arc robot, by define the current, the arc robot moves the corresponding arc head with certain displacement. As long as the motor system is not a linear system, and the number of measurements is limited, the thesis tries to find an effective way to identify a reliable model based on the measurement combination and decreasing of the cost indicator RMSE.

1 Introduction

The thesis mainly applied the system identification method to identify the mathematic model of an unknown servo-motor model in the arc-welding robot. With the input signal and output signal data, the system identification method is able to find out the black box model. The system identification method was firstly tested in a simulated servo-motor model. Then, following the theoretical model, the system identification method was applied to the real servo-motor in an arc-welding robot from Comau Spa. Several non-linear models were built by using the method, the best models were finally selected by optimization.

As long as the measurements are limited with the numbers, the combination of the measurements is applied to improve the identification process, therefore, the identified model.

1.1 Introduction to the nonlinear system identification

For the nonlinear system identification issue, one of the most popular method is the nonlinear Hammerstein-Wiener model. The Hammerstein model was discussed date back to 1966 by Narendra and Gallman [1]. As mentioned in the paper from Dr. Bai [2], the existing method for Hammerstein model identification could be roughly divided into four categories: the iterative method (Narendra & Gallman, 1966; Stoica, 1981), the over-parameterization method (Bai, 1998; Boutayeb, Rafaralahy, & Darouach, 1996; Chang & Luus, 1971; Hsia, 1976), the stochastic method (Bilings & Fakhouri, 1978; Greblicki, 1992; Pawlak, 1991) and the separable least-squares method [2].

In this paper, the Hammerstein-Wiener Model is applied to identify a nonlinear behavior servo-motor model. Since the NLHW model is applied to modeling of electro-mechanical systems successfully, according to Han [3].

The nonlinear Hammerstein-Wiener model (NLHW) could be considered as a combination of linear blocks and nonlinear blocks, as shown in the Fig1.1-1, Fig1.1-2 [3] and Fig1.1-3 [4] the Hammerstein model, the Wiener model, and the NLHW model.

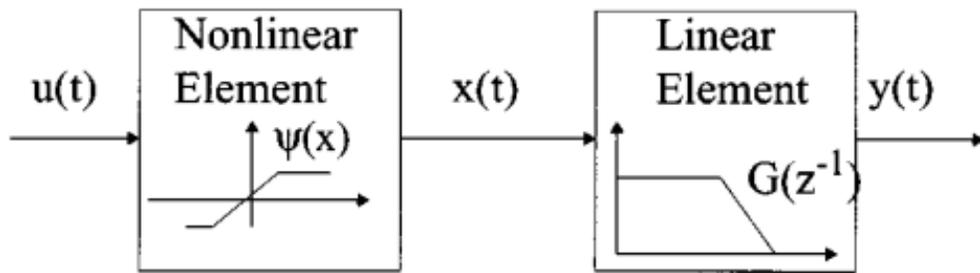


Fig1.1-1. Hammerstein Model.

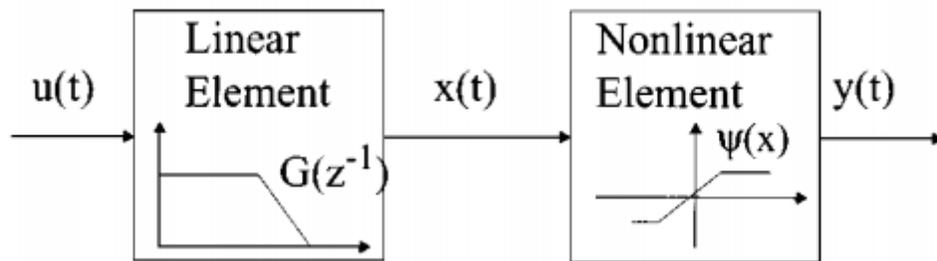


Fig1.1-2. Wiener Model.

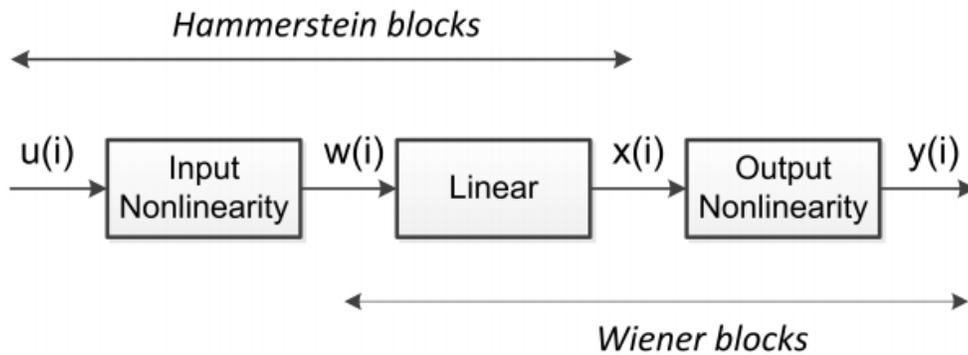


Fig1.1-3. Hammerstein-Wiener Model.

The Hammerstein-Wiener system is consisting of three main blocks: an input static nonlinearity block [5], a linear dynamic system in the middle, and followed with an output static nonlinearity block [5]. In the paper [7], Bai gave a solution named as “optimal two stage identification algorithms” . Where from Han [3], the Hammerstein Wiener system is summarized as a very well-known way:

$$\begin{aligned}
 x_1(t) &= f_1(u(t)) \\
 x_2(t) &= G(q)x_1(t) \\
 y(t) &= f_2(x_2(t)) + v(t)
 \end{aligned}$$

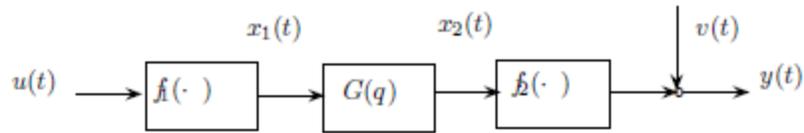


Fig1.1-4. Hammerstein Wiener system

And in this paper, a Hammerstein wiener identification tool box from MATHLAB is used, which has the same structure way of the above system, and use three parameters to represent the gain of the linear block $G(q)$:

$$G = \frac{nb}{nq}$$

And the factor n_k is used to describe the delay [8]. This model will be used as a basic tool in the later identification works, and the detailed will be further explained.

1.2 Introduction to the spot-welding gun system and the servo-motor

In this thesis, the main task is aiming to identify a high-accuracy mathematic model for the spot-welding gun system by using the input current and output displacement. The spot-welding gun system was provided by Comau S.p.a. And the spot-welding gun was driving by a linear stepper motor to close the “clamp”. The “linear” represent the movement of the motor is a line compare to the normal motor with rotation movement. And the general structure could be seen from the Fig1.2-1 in below:

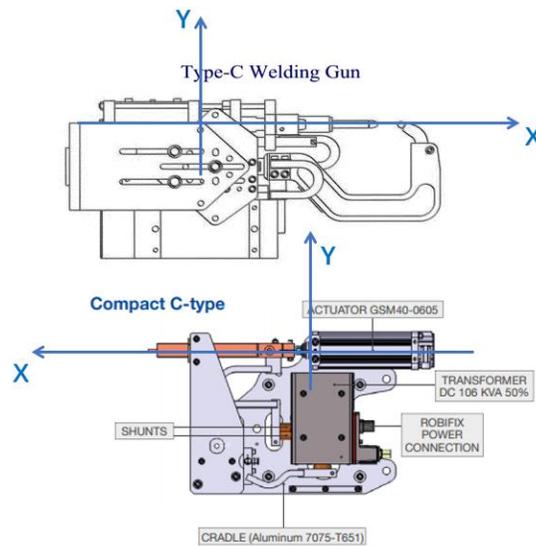


Fig1.2-1. A Welding Gun with the linear stepper motor

The linear stepper motor with a current excitation will move the front end towards, along the positive direction of the x axle. As in the current paper, the entire welding gun along with the stepper motor are considered as a black box, ignoring the complexity of the entire system, the input current is the only changing factor which determined the displacement of the front end.

The input current and output displacement are recorded correspondingly, and feed to the nonlinear Hammerstein Wiener algorithms to identify a nonlinear model. Further, the stepper motor model will be firstly simplified in the Chapter 2 with a simulation model and test with the nonlinear Hammerstein Wiener identification method for the first step theoretical verification.

2 System Identification method applied to simulated servo-motor system

2.1 Simulated servo-motor system build up

For the very first step, a simplified spot-welding gun system will be set up, aiming to test the system identification method and detect the optimized model for the next steps.

The principle for the simplified model, should be considered all the time, as the simplified model should be as lean as possible, meanwhile contains the necessary elements as well as the real spot-welding gun system contains.

After the careful testing and research, the model consists of an electric circuit, an electronic-motor, the wheel-axle system, and the spring mass system which represent the work piece under compression load, is chosen for the first step testing.

In general, the spot-welding gun or servo-motor system shall contain the following parts, the base electric circuit, electric motor, wheel-axis junction, and a spring-mass subsystem, which are shown in the following Fig2.1-1:

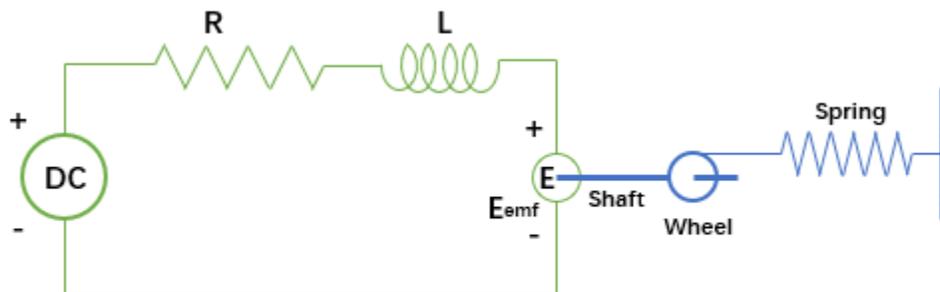


Fig2.1-1

Where the system in green represent the electric circuit, including DC power (DC), Resister (R), Inductor (L), and a DC motor (E). The system in blue represent the mechanical part: the wheel-axis junction (Shaft, Wheel) and the spring-mass subsystem.

Further the model is built in MATLAB Simulink with a similar structure Fig2.1-2. The excitation current to the electrical motor and the displacement of the spring will be further used in the theory pre-validation section.

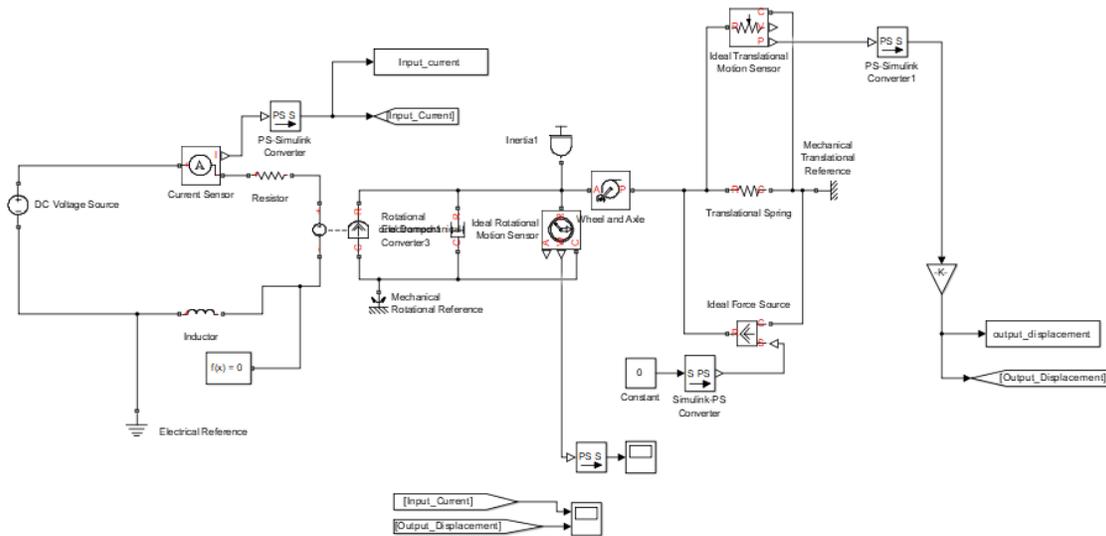


Fig2.1-2

The simplified model parameters could be checked in the following table T2.1-1.

Short Name	Long Name	Value	Units
R	Resister	100	Ohm
V0	DC Voltage Source	220	V
I	Inductor	1	H
EMF	E-motor Back EMF	7.2×10^{-5}	V/rpm
Re	E-motor Resistance	3.9	Ohm
le	E-motor Inductance	1.2×10^{-5}	H
Inertia	Inertia of the connection shaft	0.3	kg*m ²
Rdw	Wheel Radius	1	m
k	Spring Rate	293.1	N/mm

T2.1-1

2.2 System identification of the simulated system

Based on the simulation model, we could generate a set of measurements. Each set of the measurements consists of current input and the displacement output. We try to modify the input current by modifying the resister (R):

$$i = V0/R$$

Of course, consider the inductor in the loop and the servo-motor inductance and resistance, the actual current will be smaller than the ideal situation. Still we are able to measure the actual current value in the loop, as well as the displacement of the spring.

In the following chapters, the system identification of simulated system will follow the process with 4 steps as shown in the Fig2.2-1:

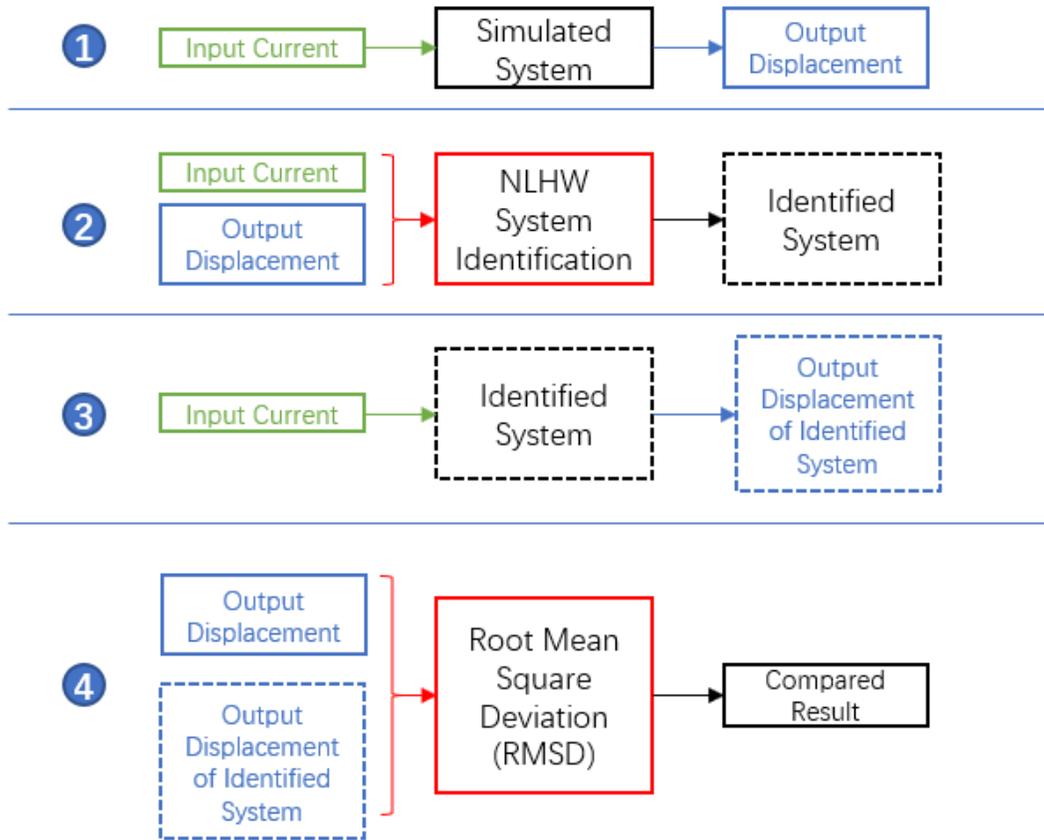


Fig2.2-1

Step1. Measurement of the input and output from the simulated target system.

In this step, the input current will be feed-in to the Simulink model, and measure the output of the displacement of the spring. In the Fig2.2-1, the result of this step will be the output displacement.

Step2. System identification based on the measurement sets.

The measurement of input current and output displacement will be considered as one measurement set. We use the Non-linear Hammerstein Wiener identification model to identify a system. In the Fig2.2-1, the result of this step will be an identified system with dash line.

Step3. Generate displacement based on the same current input and the identified system.

The identified system from step 2 will be feed by the same input current as in the step 1. And a result of the displacement from the identified system will be measured.

Step4. Compare the result of the original displacement measurement and the identified measurement.

Here we use the root mean square error to evaluate the displacement from the simulated system and the identified system. Therefore, to evaluate the result of the identified system.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N \|D_{orig} - D_{ident}\|^2}$$

2.3 Result of the simulated system

As discussed in the last chapter 2.2, we designed a group of input current (T2.3-1):

Data Set Number	1	2	3	4	5
Input Current [A]	1	2	3	4	5

T2.3-1

And in the following sub chapters, the 4 steps explained in the Fig2.2-1 will be applied with the simulated model, and the result of each step will be shown.

2.3.1 Measurement of the input and output from the simulated target system.

As we control the input current from the table T2.3-1 to the simulated target system Fig2.1-2.

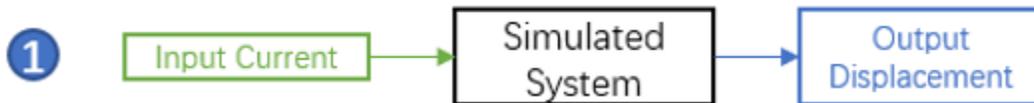


Fig2.3.1-1

we generate the following measurement of the input current and output displacement:

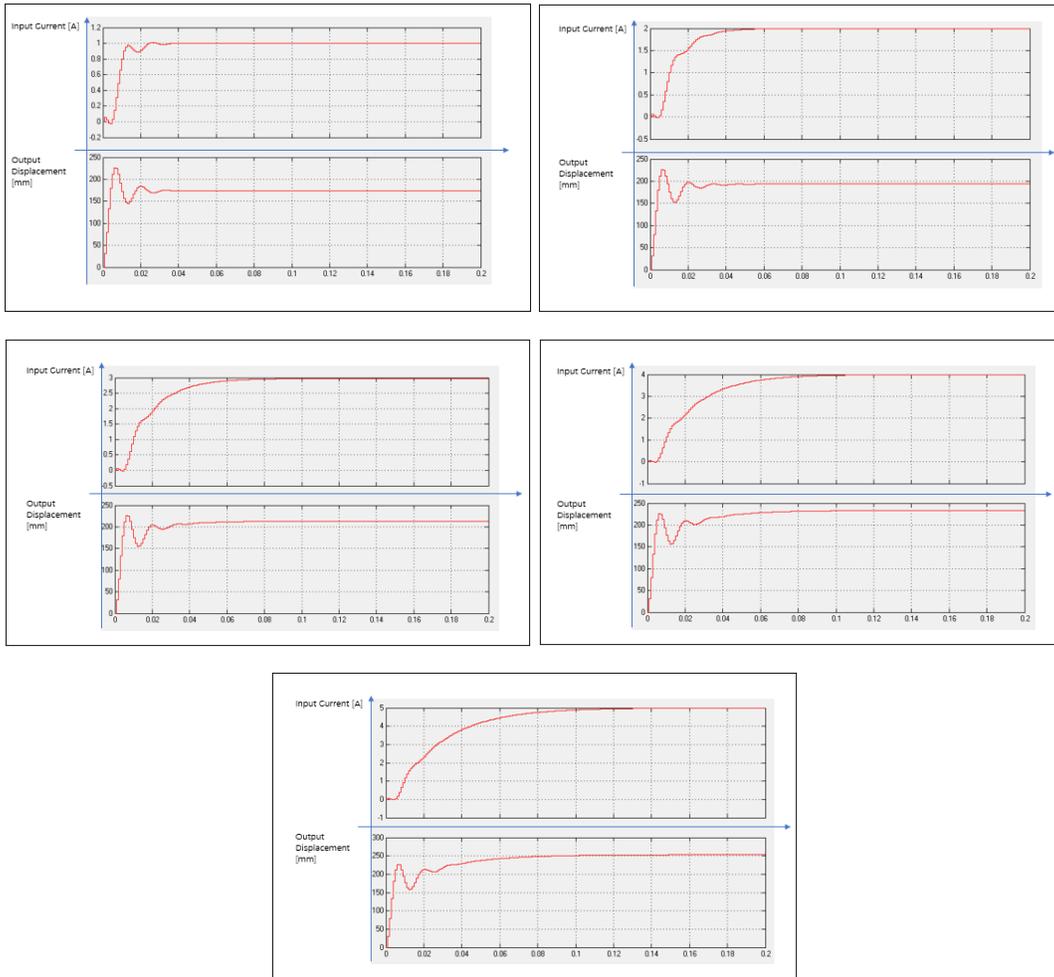


Fig2.3.1-2

We could observed that the relation between the input current is not directly linear. Therefore, the simulated system is good to try with the Non-linear Hammerstein Wiener model. And the simulation result (displacement) and input current will be grouped as data set and used for the next step.

2.3.2 System identification based on the measurement sets.

Based on the data set from last step, the input current (i_{simu}) and the output displacement (d_{simu}) of the simulated system, we try to identify a NLHW system. Hereby, the simulated data sets are firtly grouped and then feed into the NLHW with different parameters combination. Since we use the MATLAB NHLW system identification funtions, there are 2 parameters could be modified to generate the system: n_b and n_k , where the parameters was described in the chapter 1.1, as gain of the linear part of the NLHW block sets and the delay parameter.

Hereby, we consider the parameter range of [1,10] for and . Therefore, the combination could generate 10x10, a 100 systems based on one measurement sets T2.3.2-1.

System Number	n_b	n_k
1	1	1
2	1	2
3	1	3
...		
56	5	6
...		
100	10	10

T2.3.2-1

In the Appendix 7.1.1, the code for system identification will be shown there.

By this step, we mark the identified model as $SYS_{ident}(n)$, where n represent the system number.

And for each measurement, we have identified 100 models.

2.3.3 Generate displacement based on the same current input and the identified system.

As we already have the identified models, hereby, we simulate the identified model $SYS_{ident}(n)$ with the same current input (i_{simu}), and try to obtain the identification model displacement ($d_{ident}(n)$) of each identified model. The illustration table T2.3.3-1 is shown as below, the full scale of the identification result will be shown in the attached files illustrated in Appnedix 7.1.

1	45.43779	118.1152	149.2415	152.2934	...	256.3159	256.3159	256.3159	256.3159
2	244.6509	245.0383	245.1333	245.4133	...	252.9694	252.9694	252.9694	252.9695
...									
100	131.2844	72.71037	155.7699	142.9627	...	253.2419	253.2421	253.2422	253.242

T2.3.3-1

In the Appendix 7.1.2, the code for simulation of the identified model will be post.

And in the table, each row represent the simulated displacement of the identified model with the corresponding number. E.g. the displacement result of identified model 1-3-1, Fig.2.3.3-1:

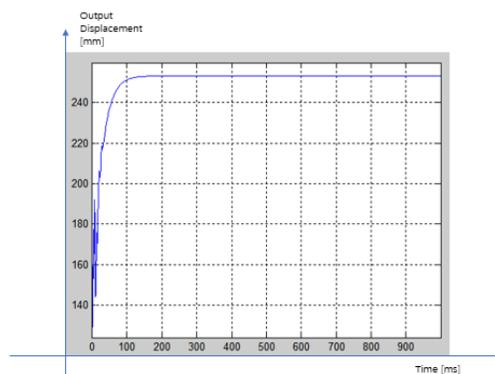


Fig.2.3.3-1

2.3.4 Compare the result of the original displacement measurement and the identified measurement.

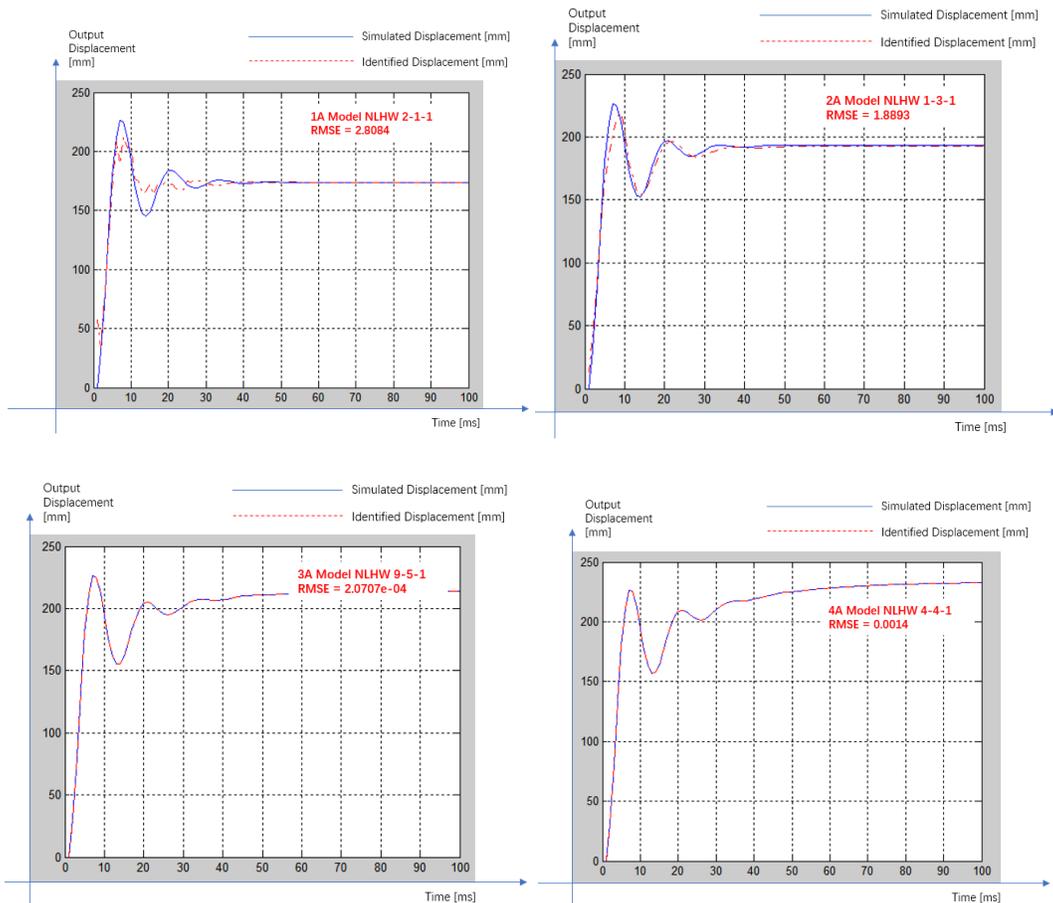
From the last step, we already obtained the displacement, hereby, we apply the RMSE calculation to the simulated displacement and the identified displacement. Further, the lower the RMSE, the better result it is.

Remember that the RMSE is applied in the way:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N \|D_{simu} - D_{ident}\|^2}$$

We could obtain a 100x1 array of the RMSE result of each identified model data set. And we select the least deviation (low error) models. The code for RMSE calculation is attached in the Appendix 7.1.3.

For the 5 groups of data sets, we obtained 5 corresponding result and shown in the following figure Fig.2.3.4-1, where the blue line represent the original simulated displacement from the system, and the red dash line represent the displacement from the identified system with best RMSE result of each group.



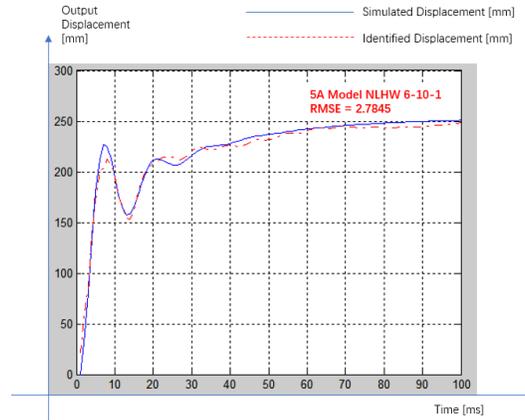


Fig.2.3.4-1

In this step, the result already shows a good identified result. From the theoretical point of view, we could continue to use the NLHW model for the real measurement from the spot-welding gun system.

However, it should be noticed that, in this chapter, the very simplified NLHW identification method is applied. Still, a further grouping of the data sets and cross check of the identified model is missing. In the next chapters, we will consider on more comprehensive way to generate a well grouped data set and cross verification experiment, so that the final identified system is neither under fitting nor over fitting.

3 System identification method applied to the real measurement

In this big chapter, the system identification method of NLHW is applied to the real measurement. And we firstly obtain the data measurement from the designed experiments, and calibrated. Then based on the measured three types of signals, the step signal, sinusoidal signal and random signal, the identification is separate into two big steps: firstly, the identification based on the single measurement and verified with the other measurements. Then based on the result of the single identification system, the identification based on the grouped measurements is conduct to find the best identified model, and cross check with the testing data sets.

3.1 Experiment in real spot-welding gun system and data pre-calibration

In this chapter, the experiment in the real spot-welding gun system will be illustrated, as well as the pre-calibration of the data sets.

3.1.1 Experiment in real spot-welding gun system

Similar like in the previous chapter, the real measurement is targeting to obtain the current input to the welding gun, and the displacement of the head (ending) of the welding gun.

In the Comau workshop, the spot-welding gun was tested with 3 different types of current signals as input. The 3 types are step signal, sinusoidal signal and random signal. Each type of signal was segmented with different amplitude. And for sinusoidal signal, not only the amplitude, but also the frequency is segmented with 5 different values.

In the table below, the 3 signal groups can be easily seen:

For the Random signal, we conduct 5 tests with different amplitude:

Amplitude	1	2	3	4	5
Test No.	#1	#2	#3	#4	#5

T3.1.1-1 Random signal

For the step signal, we conduct 15 tests, with 3 different delays and 5 different amplitudes:

Delay\Ampl.	2	4	6	8	10
0	#1	#2	#3	#4	#5
100	#6	#7	#8	#9	#10
300	#11	#12	#13	#14	#15

T3.1.1-2 Step signal

For the sinusoidal signal, we planned to conduct 25 tests, with 5 different frequency and 5 different amplitudes, however, as in certain frequency area, the motor face a strong resonance and we shut down 2 tests, considering the safety reasons.

W\A	1	2	3	4	5
0.5	#1	#2	#3	#4	#5
0.34	#6	#7	#8	#9	#10
0.22	#11	#12	#13	#14	#15
0.11	#16	#17	#18	R	R
0.005	#19	#20	#21	#22	#23

T3.1.1-3 Sinusoidal signal

Where in the table, w represents the frequency, and A represent the amplitude. And R represent the resonance area, and the tests was not performed and left blank.

The sinusoidal signal is conducted following the function:

$$i = A \cdot \sin(w * t) \quad (1)$$

where:

$$w = 0.005/T_s \quad (2)$$

According to the upper input current signals, the displacement of each input is measured, and group as corresponding data sets. And the raw measurement data sets need the further calibration, which will be illustrated in the next chapter 3.1.2.

3.1.2 Data pre-calibration

As the raw data are collected from the Comau spot welding gun, the current data and displacement data are of different length, with noise, offset and even with small mistakes. Therefore, the first step is to calibrate the raw data, and choose the useful part for the identification in the next steps.

As we have said in the previous section, the experiments are conducted with 3 groups. The step signal is conducted with the certain delay. Therefore, the offset and the total length should be calibrated. For the Random signal, only the total length should be calibrated. And the Sinusoidal signals should be calibrated with both the length and the offset.

3.1.2.1 Calibration of Step signals

The step signals are little bit different from the data of sinusoidal signals. It requires

The calibration follows the steps as below:

1. Cut the length of the current data, meanwhile cut the output displacement as same length.
2. Calibrate the offset of the displacement data, as the starting displacement should set to 0. (The offset value is chosen by the average value in the first period, see Fig.3.1.2.1-1, -2)

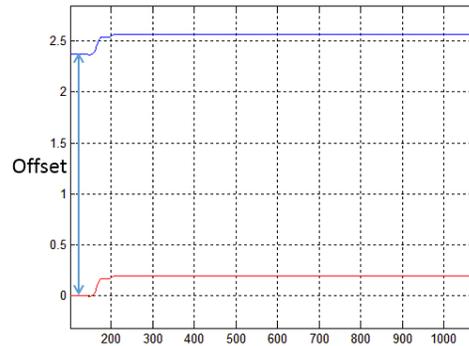


Fig.3.1.2.1-1 The displacement for sinusoidal signals

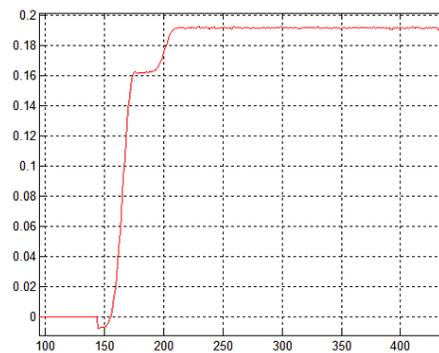


Fig.3.1.2.1-2 The data after calibration.

3.1.2.2 Calibration of Sinusoidal signals

For sinusoidal current inputs:

$$i = A \cdot \sin(\omega \cdot t) \quad (1)$$

And all the current inputs are named after: Sin01, Sin02, Sin03, ... Sin23.

Correspondingly, the displacement is calibrated as well.

The calibration follows the steps as below:

1. Cut the length of the current data, meanwhile cut the output displacement as same length.
2. Calibrate the offset of the input current, as well as the displacement data, as the starting displacement should set to 0. (The offset value is chosen by the average value in the first period T, see Fig.3.1.2.2-1, Fig.3.1.2.2-2, Fig.3.1.2.2-3, Fig.3.1.2.2-4)
3. Cut the length of the current data again, and meanwhile cut the output displacement as same length

And the calibration MATLAB program code can be seen in the Appendix Data Calibration.
In the figure below, the data before calibration and after calibration are shown.

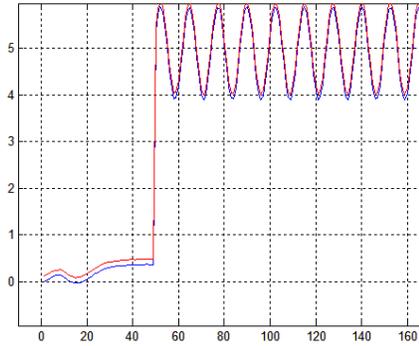


Fig.3.1.2.2-1 Offset calibration of the sinusoidal signals

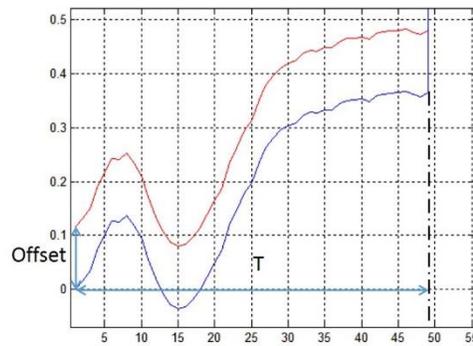


Fig.3.1.2.2-2 Offset calibration of the sinusoidal signals.

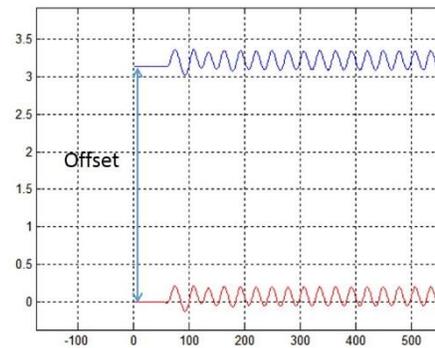


Fig.3.1.2.2-3 Offset calibration of the sinusoidal signals output displacement.

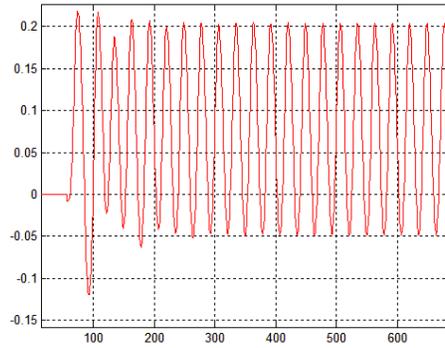


Fig.3.1.2.2-4 Offset calibration of the sinusoidal signals output displacement.

3.1.2.3 Calibration of Random signals

The random signals just need the length cut to modify the data. See Fig 3.1.2.3-1 and -2.

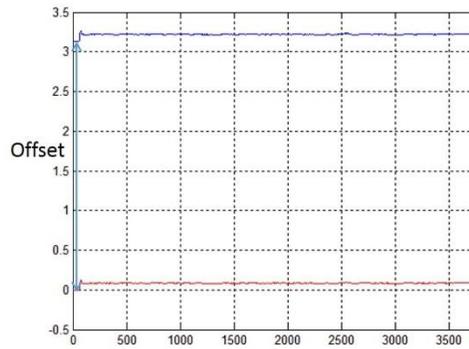


Fig 3.2.3 The displacement data before calibration and after calibration.

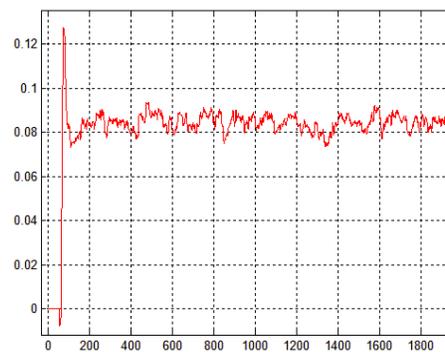


Fig 3.2.3 The displacement data after calibration.

3.2 Single measurement system identification

In this step, the NLHW identification method is applied to single step measurement. Since there are no further model optimizing steps, the identified model will be only cross checked with the step measurement sets.

3.2.1 Method of single measurement system identification

Similar like in the chapter 2 of the simulated system, but more adapted to the application case, hereby the identification of single step measurements is also following the steps:

1 Identification of the model:

- 1.1 Using single data set to identify models based on the combination of parameter n_b and n_k . For example, the identified model is named as "NLHW-Step-n- n_b - n_k -1", to represent the model identified by data set n, using the parameter n_b and n_k . Shown in Fig.3.2.1-1, step 1.
- 1.2 Self-cross check of the identified model (NLHW-Step-n- n_b - n_k -1) with the same measurement data set (n), and choose the best one or several identified models. (NLHW-Step-n- n_b - n_k -1-Chosen, shown in Fig.3.2.1-1, step 2 and 3.)

2 Verification of the model:

Based on the identified model (NLHW-Step-n- n_b - n_k -1-Chosen), verified with all the step datasets.

- 2.1 Applying the input current of each step datasets (m), and obtain the identified model displacement output (Displacement--Step-n- n_b - n_k -1-m, shown in Fig.3.2.1-1, step 4).
- 2.2 Verification of the result with RMSD method based on measured displacement and the identified model displacement. (Shown in Fig.3.2.1-1, step 5)

In the following chapter, the detailed result of each steps will be illustrated.

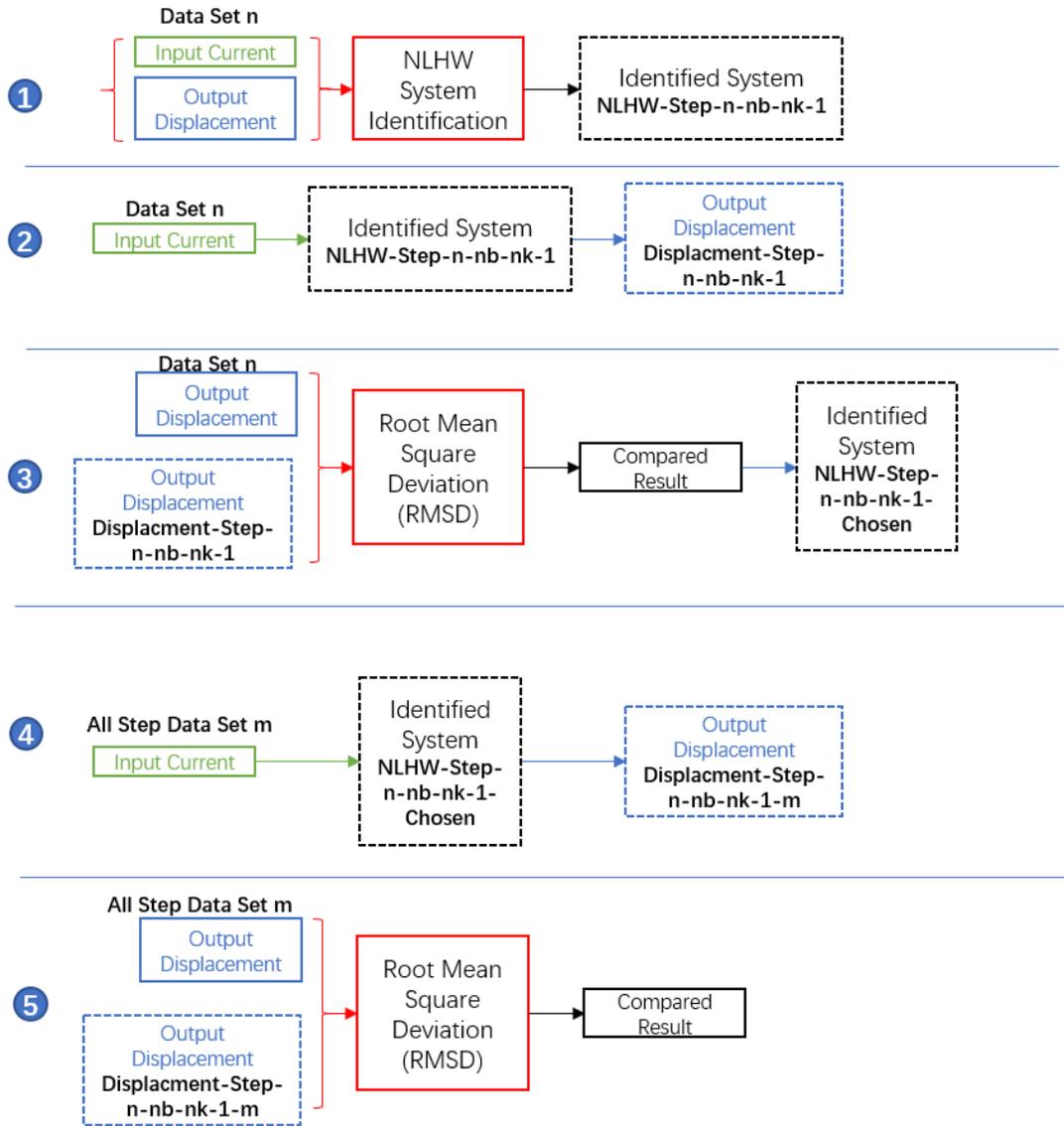


Fig.3.2.1-1

3.2.2 Result of single measurement system identification

3.2.2.1 Result of single measurement with self-cross check

Based on the process designed in the previous chapter, the result of each step can be checked in the following, and the code for step 1 to 3, is attached in Appendix 7.1.4.1.

In the step 1, for 15 step data sets, considering the parameter n_b and n_k combination:

$$n_b \in [1,10], n_k \in [1,10]$$

we could identify models of 1500 (15x10x10), which we named as “NLHW-Step-n- n_b - n_k -1”.

After the 1500 models are identified, according to step 2, the 1500 models are firstly stimulated with the same current input ($i_{meas}(n)$), which are used also for the identification of the

measurement n. And therefore, the $X_{mdl}(m)$ ($m \in [1,1500]$) are generated in the step 2. Considering the size of each displacement from the identified model (1500 x time series of 7500 steps) will generate a large memory size data result, we set the step 2 as an intermediate step. And the result will not be saved separately, but directly goes to the step 3 for the RMSD result calculation.

And in the step 3, the RMSD result is firstly generated with a matrix of 100x15, where 100 is a result of the combination of parameter n_b and n_k , 15 is a result of the number of the measurement data set.

Still, the calculation is based on the RMSD equation:

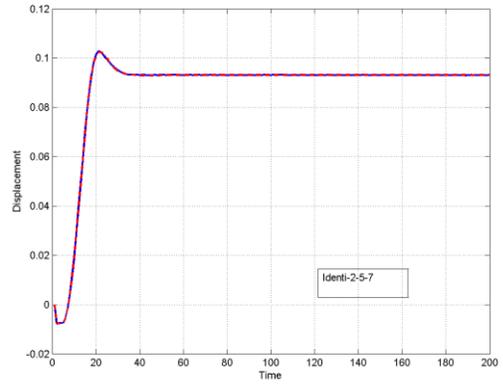
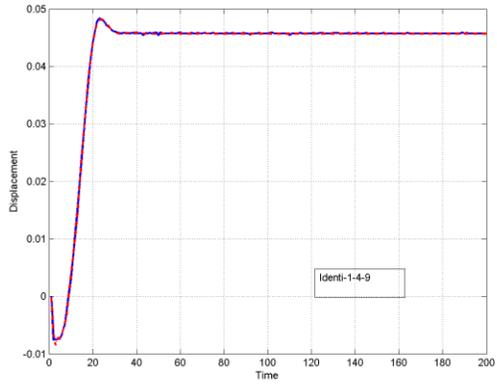
$$RMSD_{n-nb-nk-1} = \frac{\|x_{mdl-n-nb-nk-1} - x_{meas-n}\|}{\sqrt{length_x}}$$

And then, for each data set, the top three model with lowest RMSD result (15x3) is selected (NLHW-Step-n-nb-nk-1-Chosen), which is shown as below (T3.2.2-1).

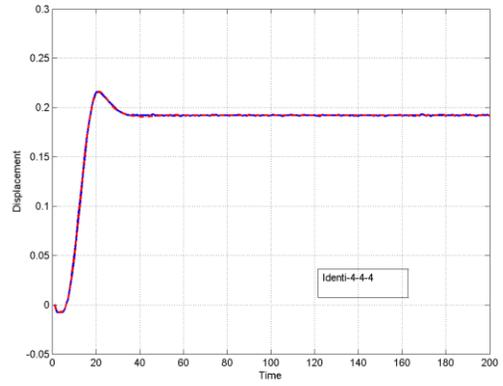
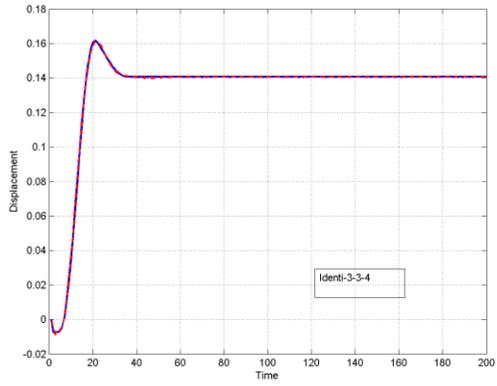
Data Set	RMSE 1	RMSE 2	RMSE 3
1	0.000171	0.000193	0.000244
2	0.000114	0.000131	0.000151
3	0.000295	0.000371	0.00038
4	0.000462	0.000502	0.000539
5	0.000226	0.000231	0.000396
6	0.000106	0.000242	0.00051
7	0.000415	0.000571	0.000595
8	8.82E-05	0.000176	0.000184
9	0.000416	0.000501	0.000568
10	0.000305	0.000335	0.000422
11	0.000168	0.000224	0.000354
12	0.000327	0.000371	0.000457
13	0.000197	0.000245	0.000509
14	0.000244	0.000607	0.000612
15	0.000555	0.000746	0.001047

T3.2.2-1

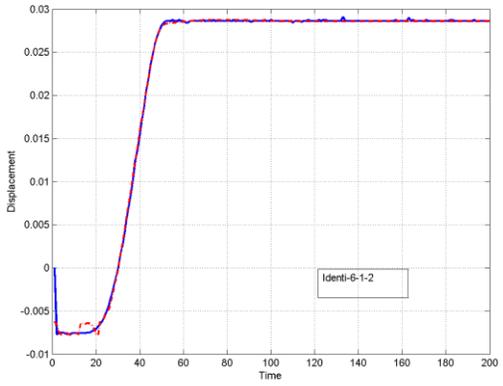
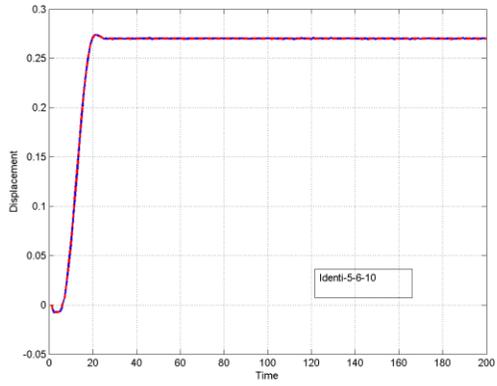
In the following page, the compare of the displacement from the low RMSE value simulated system and the original measurement displacement of the best RMSD result will be shown (Fig.3.2.2-1).



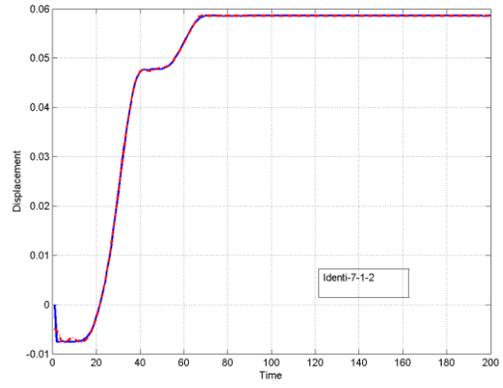
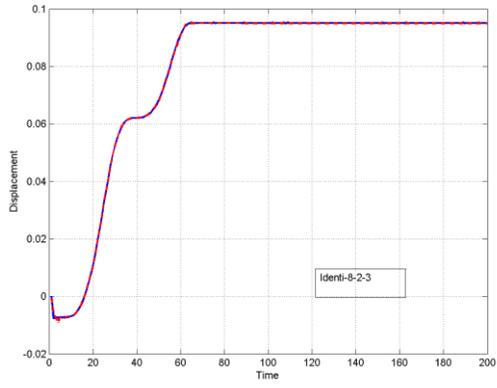
(1) (2)



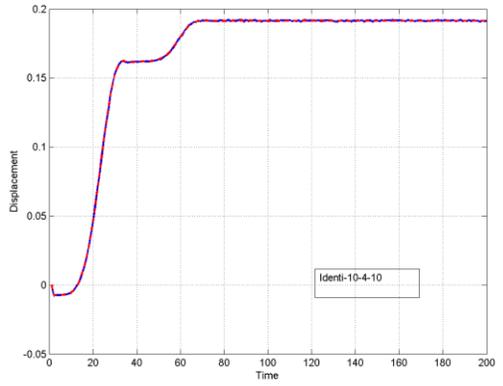
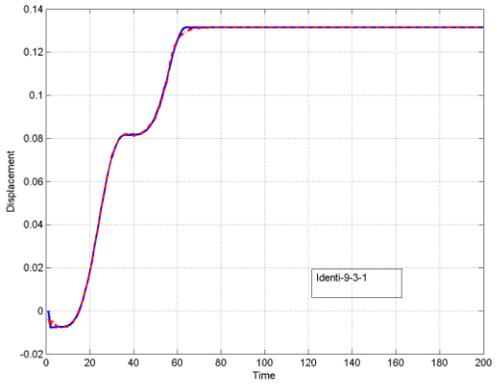
(3) (4)



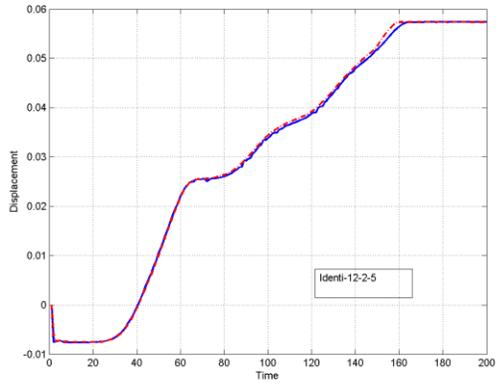
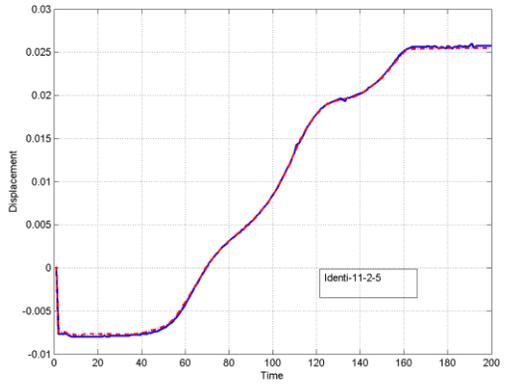
(5) (6)



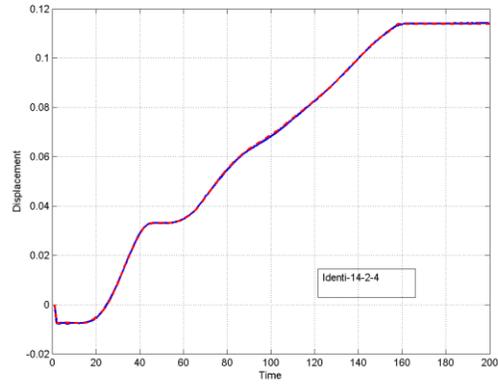
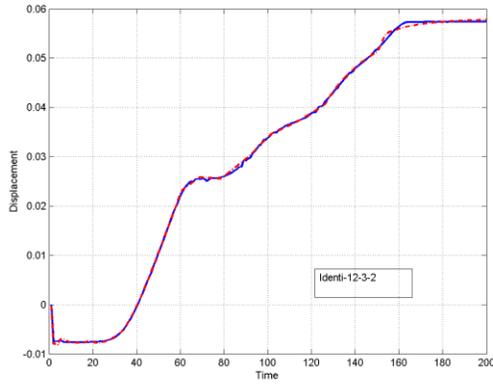
(7) (8)



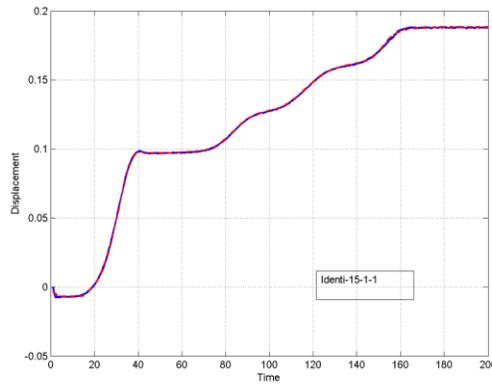
(9) (10)



(11) (12)



(13) (14)



(15)

Fig.3.2.2-1

From the figure above, the Non-linear Hammerstein Wiener method shows a very perfect fitting in identification result. Both in the transition situation and the station situation.

However, there could be a potential risk of overfitting. That in the case of overfitting, the identified model may fit perfectly to the data set which used for identification, but quite poor with the other data sets. For example, with the identified model based on data set ($m=10$), the best fitted model is identified with parameter of $nb = 4$, $nk = 10$. But when the model is applied to the other data set ($m=8$), the result shows a very bad fitting (Fig.3.2.2-2):

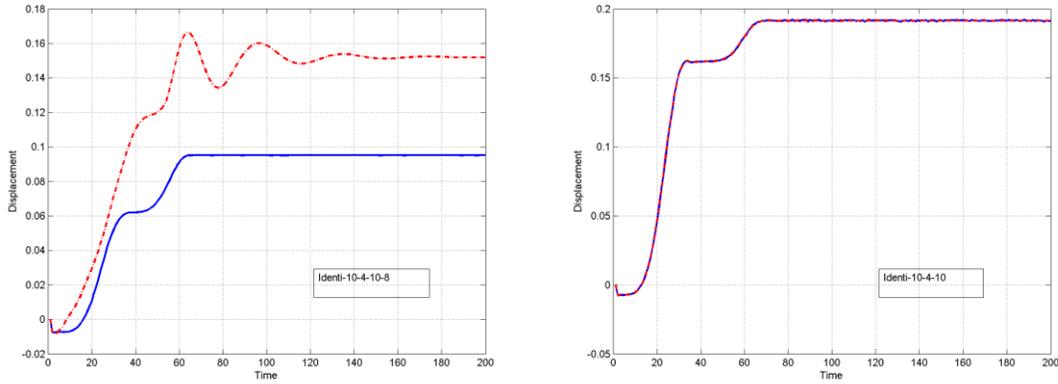


Fig.3.2.2-2

Therefore, the single best fit could come to the overfitting, and a general cross check with the other dataset is necessary.

3.2.2.2 Result of single measurement with over all cross check

In order to have a further full coverage check, and to generate the displacement of the identified model in the step 2 and to make it comparable to all the other data sets, it is needed to stimulate the identified system (1500) with the measured current input of each data sets (15), and it is resulting in a size of 1500x15 data sets. Where each data set is consisting of a time series signal (7500). Considering both from the memory consumption and calculation efficiency side, it is not necessary to create such an intermediate huge matrix (1500x15x7500). It is designed to directly calculate the RMSD value (1500x15) of each model stimulated displacement result and the measured displacement result.

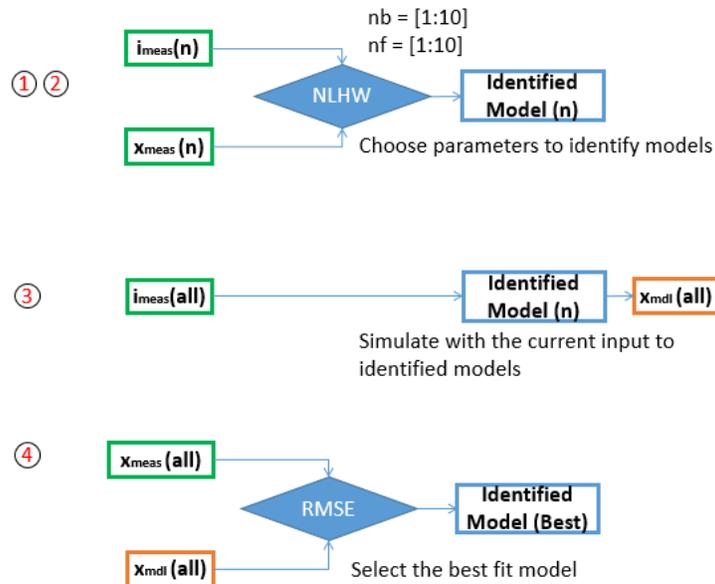


Fig.3.2.2.2-1

And the simulated displacement of the identified model of 1500x15 data sets are calculated with the RMSD functions:

$$\text{RMSD}_{n-nb-nk-m} = \frac{\|x_{mdl-n-nb-nk-m} - x_{meas-m}\|}{\sqrt{\text{length}_x}}$$

Where the n represents the step data set number, which are used to identify the model. The n_b and n_k as we have already mentioned, the parameter of the identified model. The combination of n- n_b - n_k -m is of 15x10x10x15. $X_{mdl-n-nb-nk-m}$ represent the displacement of the identified model based on data set n, with the same current input of data set m. The X_{meas-m} represent the displacement of the measured value of data set m. Here, the m and n represent 15 data sets. The n represents the data set which used to identify the model, and m represent the test/stimulate data set.

And the full result is grouped in an exempld way in Matrix T3.2.2.2-1 (1500x15):

identify data set (n)			test data set (m)							
	nb	nk	1	2	3	4	...	13	14	15
1	1	1	0.001434	30.5893	61.26105	91.92479	...	38.99844	59.76257	80.79977
1	1	2	0.000479	2.03936	4.0759	6.108788	...	8.278666	7.072215	6.832956
1	1	3	0.001124	0.008017	0.01543	0.019241	...	0.04593	0.06264	0.050647
...
4	5	4	0.595397	0.553589	0.18921	0.001079	...	0.458872	0.392213	0.319247
...
15	10	9	0.018531	0.021615	0.021223	0.047953	...	0.044332	0.042272	0.016405
15	10	10	0.030358	0.055255	0.09093	0.136843	...	0.037791	0.03502	0.029765

T3.2.2.2-1

Hence, based on the RMSD result in the table, we could further select the identified model which satisfied the RMSD threshold and could cover most of the data sets. This would mean that we need a criterial to evaluate that, up to which level the RMSD result is good enough.

After considering the compare the result, a threshold of RMSD is selected as 0.01. See the plot (Fig.3.2.2.2-2) in below:

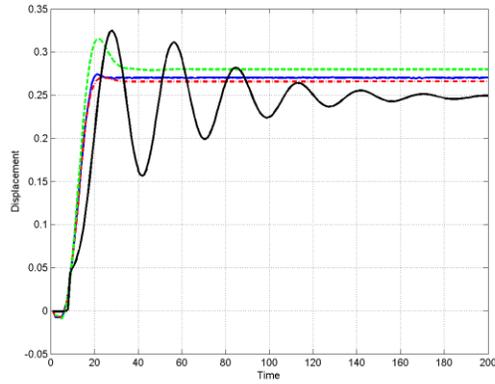


Fig.3.2.2.2-2

Where the blue line represents the original measurement displacement of the data set ($m = 5$). The red line represents the displacement from an identified model of RMSD value equals to 0.004. The green line represents the displacement from an identified model of RMSD value equals to 0.01. And the black line represents the displacement from an identified model of RMSD value equals to 0.04.

It could be clearly seen that the RMSD threshold of 0.004 may lead to an overfit situation, and 0.04 is far too away from the accurate level. Therefore, a threshold of 0.001 is selected for the RMSD. And the following table T3.2.2.2-2 is generated:

Identification Data Set n	nb	nf	Num of Data set RMSD value under 0.01
6	3	7	7
7	7	1	7
9	6	3	7
9	4	2	7
13	8	2	7
14	10	9	7
14	6	9	7
14	5	1	7
14	7	7	7
14	7	6	7
14	3	9	7
14	1	3	7
14	5	4	7
14	10	6	7
14	10	5	7

7	10	1	6
8	1	2	6
8	6	4	6
12	8	2	6
13	10	8	6
13	7	1	6
13	6	10	6
13	8	6	6
13	9	6	6
13	9	3	6
13	6	7	6
13	9	4	6
13	6	1	6
13	3	6	6
13	9	5	6
13	10	3	6
13	9	1	6
13	7	8	6
13	8	8	6
13	7	6	6
13	10	7	6
13	5	8	6
13	1	1	6
13	8	3	6
13	8	4	6
13	6	4	6
13	5	3	6
14	1	1	6
14	2	9	6
14	8	10	6
14	9	9	6
14	3	7	6
14	5	3	6
14	2	10	6
14	7	8	6
14	6	2	6
14	10	4	6
14	2	8	6
14	4	10	6
14	6	10	6
14	8	3	6
14	7	3	6

14	7	5	6
14	6	7	6
14	5	6	6
14	3	1	6
14	1	5	6
14	1	2	6
14	9	6	6
14	6	6	6
14	1	9	6
14	2	7	6
14	1	7	6
14	9	4	6

T3.2.2.2-2

Where we could find out that, based on the data set: Step-6, Step-7, Step-8, Step-9, Step-12, Step-13, Step-14, the identified model could cover a large number of data sets. Therefore, we select the above 7 data sets as the basis for the next grouping identification base.

Additionally, if we use the Table (T3.2.2.2-3) to show the coverage situation or grouping situation, we could find out that, although the data sets are listed based on the ramping time and the target current amplitude, the grouping of the data sets are not ramping time-current amplitude based.

Identification data set \ Validation data set	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Number under Threshold
1	4E-04	0.026	0.052	0.081	0.134	0.016	0.009	0.006	0.019	0.06	0.022	0.015	0.011	0.009	0.045	4
2	0.013	0.006	0.034	0.069	0.129	0.016	0.033	0.016	0.011	0.052	0.02	0.03	0.028	0.022	0.038	1
3	0.02	0.012	7E-04	0.002	0.022	0.007	0.02	0.04	0.055	0.052	0.011	0.015	0.034	0.05	0.043	3
4	0.039	0.032	0.01	7E-04	0.025	0.022	0.011	0.033	0.054	0.052	0.016	0.015	0.028	0.047	0.048	2
5	0.01	0.031	0.019	0.006	0.001	0.02	0.014	0.025	0.057	0.072	0.036	0.024	0.028	0.048	0.053	3
6	0.017	0.034	0.051	0.07	0.113	1E-04	0.004	0.007	0.01	0.043	0.007	0.003	0.004	0.005	0.038	7
7	0.019	0.042	0.049	0.064	0.102	0.005	0.007	0.005	0.008	0.03	0.005	0.007	0.008	0.012	0.032	7
8	0.023	0.039	0.055	0.115	0.206	0.005	0.004	0.004	0.048	0.118	0.003	0.004	0.006	0.02	0.08	6
9	0.01	0.032	0.046	0.066	0.139	0.009	0.006	0.005	0.003	0.065	0.011	0.01	0.01	0.012	0.044	7
10	0.009	0.044	0.018	0.038	0.076	0.01	0.066	0.046	0.033	0.001	0.009	0.049	0.051	0.048	0.02	4
11	0.025	0.046	0.067	0.09	0.137	0.005	0.009	0.018	0.027	0.065	0.001	0.006	0.008	0.011	0.054	5
12	0.018	0.041	0.044	0.05	0.08	0.003	0.007	0.006	0.014	0.025	0.005	0.004	0.008	0.017	0.031	6
13	0.008	0.029	0.05	0.085	0.144	0.009	0.008	0.008	0.023	0.068	0.01	0.006	0.006	0.004	0.051	7
14	0.026	0.04	0.052	0.077	0.126	0.009	0.008	0.009	0.016	0.049	0.006	0.006	0.006	0.005	0.04	7
15	0.027	0.01	0.016	0.035	0.078	0.038	0.04	0.033	0.029	0.008	0.031	0.04	0.041	0.041	7E-04	3

T3.2.2.2-3

In the table, where the “Identification data set” column represents the model which identified by the corresponding data group. And in the “Validation data set” row represent the verification data set. In the table, the RMSD value validated based on the validation data set, and the model is identified based on the identification data set. And, where the last column “Num under Threshold” shows the number of the verification data sets the corresponding model covers (the RMSD value over 0.01).

Based on the above observation, we could come to the following conclusion:

1. Considering the data sets covered by the identified models, it could be observed that the data set 15 is quite a unique measurement, which may due to either wrong measurement or extreme working area of the arc-robot. And this could be verified later on in the grouping identification and verification and may be excluded from the later on verification.
2. Further, it could be observed that data set 6, 7, 8, 9, 12, 13 and 14 are either mostly covered by the other models, or the models identified base on these data sets could cover a most of the other data sets.
3. It could be easily seen that it is not possible to have one model based on single data set to cover all the situation. This might due to the reason of the measured data sets could only

cover small range of the motor/arc-robot. And in different motor operation area, the performance of the motor/arc-robot differs a lot.

Based on the above conclusion, it is necessary to group the measurements and identify the model based on the grouped data sets.

3.3 Grouped measurement system identification

3.3.1 Pre-definition of the grouping identification

As it is already explained in the previous chapter, the target of this chapter is to group the measured data sets and identify a few models which could cover the most of the situation.

In below, the grouping considerations will be listed as the guidance for the later on grouping and identification works, it is necessary to group the data sets from single data set and add up one by one.

However, this could lead to a huge number of grouping possibilities:

$$\text{Number of grouping} = 15^{15} \approx 4.37 * 10^{17}$$

Consider if each time the data sets for grouping shall not be reused:

$$\text{Number of grouping} = 15! \approx 1.31 * 10^{12}$$

And each time if we do a verification, this could lead to a very inefficient and endless work load.

To optimize the identify process and quickly find out the possible model, the following identification path is defined:

1. To working in a scientific way, the grouping of the data sets shall be started from the very single data set and add up one by one.
2. Each time after a group, the identified model shall be verified with all the step data sets to check if the grouped model could cover more data sets (to have more data sets satisfied the RMSD threshold)
3. Based on the step2, select the groups which covers more data sets, and again add the new data set to the group.
4. Repeat the above path until the verification result won't increase any more.

For a better understanding, the identification path is explained as a flow chart (Fig.3.3.1-1) as below:

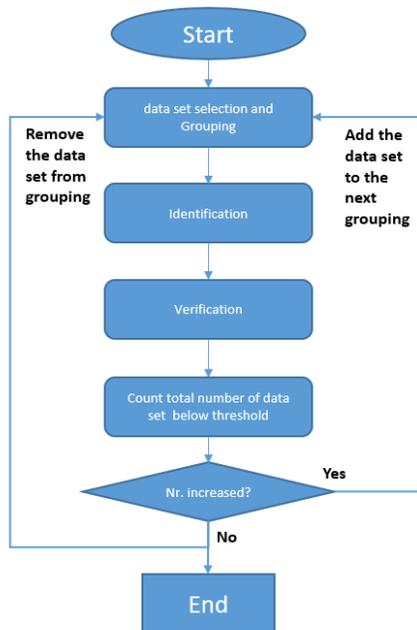


Fig.3.3.1-1

Consider a more mathematical way and more system way to explain the process, hereby we define the process as following:

Each measurement dataset is represented by $Meas(n)$,

where $n(\in [1,15])$ is the measurement dataset number, for here we have the dataset of Step signals measurement 1 to 15.

Each measurement dataset consists of two sub-datasets:

the Input current: $i_{meas}(n)$

the output displacement: $x_{meas}(n)$

For grouping of the measurement datasets, the following symbol is used: $G_k(M)$

where $k(\in [1,15])$ is the number/level of measurement dataset in combination group number. In the later on phase, this could also represent the level of iteration number.

M is the measurement dataset.

For example:

$G_3(Meas(1), Meas(5), Meas(2))$ is representing the Grouping on 3rd level/iteration, with the ordered measurement dataset 1, 5 and 2.

$G_3(Meas(1), Meas(2), Meas(5))$ is representing the Grouping on 3rd level/iteration with the ordered measurement dataset 1, 2 and 5.

The following figure (Fig.3.3.1-2) could be a good explanation of the sub-dataset, dataset and grouping relation:

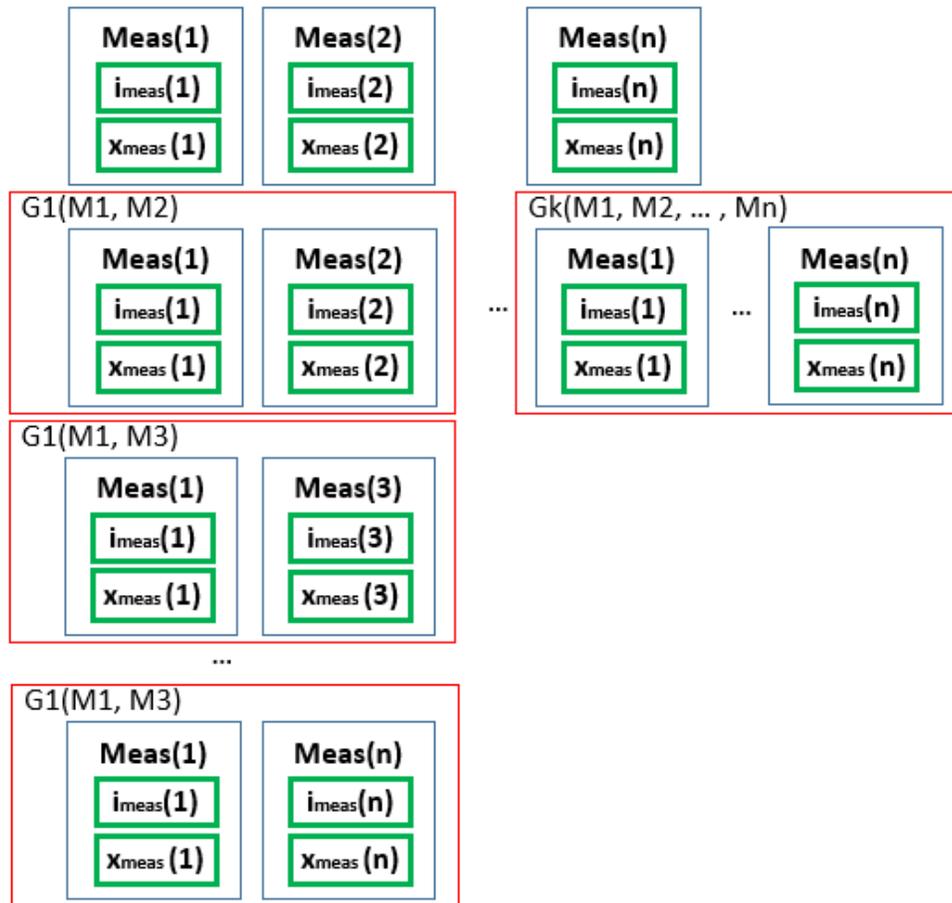


Fig.3.3.1-2

3.3.2 Grouping identification loops

As it is already defined in the last step, from this chapter the detailed computational process will be explained.

In general, the identification uses an iteration calculation:

each time increase the grouping level (number of measurements $Meas(n)$), validate the RMSD value, find the smallest RMSD combination and use the combination for the next round of calculation:

3.3.2.1 Pre-definition and starting point:

As we have measurement data sets in total of kn . ($k_n \in R$):

$$\text{Measurement_Total} = \{Meas(1), Meas(2), \dots, Meas(k_n)\}.$$

For the 1st iteration, the grouping is no different than the single measurement identification:

$$\text{Group}_1 = \{\text{Meas}(k_1)\},$$

where k_1 represent the measurement, which identified a system $\text{System}(k_1)$ and via validation with the rest of the measurements, generates the most RMSD threshold passed situation.

3.3.2.2 j^{th} iteration calculation:

3.3.2.2.1 Grouping:

For j^{th} iteration, the grouping of the measurements is:

$$\text{Group}_j(k_{jr}) = \{\text{Meas}(k_1), \dots, \text{Meas}(k_i), \dots, \text{Meas}(k_{jr})\}$$

Where, $\text{Meas}(k_{jr})$ represent the j th element of the grouped data. And the $\text{Meas}(k_{jr})$ could be the r^{th} measurement dataset in the rest of the un-selected measurements.

This means: $\text{Meas}(k_{jr})$ should be selected from the Measurement other than the existing measurements in the grouping:

$$\text{Meas}(k_{jr}) \in ! \{\text{Meas}(k_1), \dots, \text{Meas}(k_i), \dots, \text{Meas}(k_{(j-1)r})\}$$

($\in !$ Represent not belonging)

3.3.2.2.2 Identification:

System identification of the $\text{Group}_j(k_{jr})$, with full range combination of zeros and poles. Consider for each grouping, the system identified will be a number of 100 (10x10, as the number of n_b (zeros), n_k (poles)):

$$\text{System}_{k_{jr}}^{nb,nf}(i), \text{ where } n_b \in [1,10], n_k \in [1,10]$$

3.3.2.2.3 RMSD calculation and filter of each identified system:

3.3.2.2.3.1 RMSD calculation

For each identified system $\text{System}_{k_{jr}}^{nb,nf}(i)$, a RMSD is firstly calculated with the measurement ($\text{Meas}(k_{jr})$), which used for identification:

$$\text{RMSD}_{k_{jr}}^{nb,nf}(x_{k_{j_{idnt}}}, x_{k_{jr}}) = \frac{\|x_{k_{j_{idnt}}} - x_{k_{jr}}\|}{\sqrt{\text{length}_x}}$$

Where the $x_{k_{j_{idnt}}}$ is calculated by:

$$x_{k_{j_{idnt}}} = \text{System}_{k_{jr}}^{nb,nf}(i_{kj}), \text{ where } n_b \in [1,10], n_k \in [1,10]$$

And the length_x , is the total number of the $x_{k_{j_{idnt}}}$ or $x_{k_{jr}}$.

Remembering that the RMSD number is calculated based on the identified system of the grouped measurement data set with $\text{Meas}(k_{jr})$, therefore it shall obtain again a 100 of RMSD numbers corresponding to the System identified numbers.

3.3.2.2.4 RMSD filter

3.3.2.2.4.1 Self-RMSD Filtering

As mentioned in the last chapter, for each identified system, we obtained a RMSD matrix of 10x10. Therefore, for the current iteration, there are total $k_j \times 10 \times 10$ number of System/RMSD. Considering such a large number of systems shall not be applied directly to cross check with the other measurement set, we shall firstly apply a RMSD filter to limit the unpassed system.

To filter the unpassed system, we should focus on the target to obtain a threshold of RMSD (ThresholdRMSD), which will be applied to filter out the inaccurate systems.

The first step, we would like to filter out the systems which will not even pass the measurement for the identification inputs. And since the number of the systems will be largely decreased with small calculation time, this step improves to save the whole system calculation time.

And a matrix with the passed RMSD is obtained with the recorded qualified identified system numbers: $Idx_RMSD = [k_{jr}, n_{bq}, n_{kq}]$.

And the corresponding qualified system is called:

$System_{k_{jr}}^{nbq, n_{kq}}(i)$, where n_{bq} , n_{kq} are the qualified system zeros and poles.

The qualified system shall satisfy the following criteria:

$$RMSD_{k_{jr}}^{nbq, n_{kq}}(x_{k_{j}^{r_{idnt}}}, x_{k_{j}^r}) < Threshold_{RMSD}$$

Only the system whose RMSD value could first pass the threshold can be selected for further filtering.

Please note that after this first step filtering, the huge amount of the identified systems will be largely decreased, and further facilitate the next step.

3.3.2.2.4.2 Cross RMSD verification Filtering

In this section, the identified system with the first filtering shall be further filtered with the rest of the verification data sets. For a better understanding, the difference between the step 1 filter and step 2 verification filter, are mainly regarding the verified data sets are different. The step 1 filter is more like a self-checking step to filter out the poor identified system. The step 2 verification is targeting to verified the identified system with more coverage other than self-overfitting system.

Hereby, the system $System_{k_{jr}}^{nbq, n_{kq}}(i)$ from last step is verified with all the selected verification data sets: $MeasValid(k_m)$, where $k_m \leq k_n$ (Total data set numbers).

The verification filtering step is similar as the previous steps:

- i. Calculation of the RMSD based on the identified system with the verification data sets:

$$RMSD_{km}^{nbv,nfv} (x_{km^{Ident}}, x_{km^{Valid}}) = \frac{\|x_{km^{Valid}}^{Ident} - x_{km^{Valid}}\|}{\sqrt{length_x}}$$

Where the $x_{km^{Valid}}^{Ident}$ is calculated by:

$$x_{km^{Valid}}^{Ident} = System_{kj^r}^{nbq,nfq}(i_{km^{Valid}}^{Ident}), \text{ where } nb \in [1,10], nk \in [1,10]$$

As, $i_{km^{Valid}}^{Ident}$ is the input current of the validation data set.

$System_{kj^r}^{nbq,nfq}(i)$ is the system which passed the step 1 filter.

- ii. Compare the RSMD value with the threshold value. The system could pass the following criteria would be selected for the next step:

$$RMSD_{km}^{nbv,nfv} (x_{km^{Valid}}^{Ident}, x_{km^{Valid}}) < Threshold_{RMSD}$$

After the steps above, the filtered system will be named as $System_{kj^r}^{nbv,nfv}(i)$.

3.3.2.2.5 Selection of the data set based on validation filter result.

Hereby, in order to guarantee the independency of the test data set respect to the identification data set and validation data set, the selection of the system for the next iteration grouping and identification is also using the validation data set.

For selection of the data set for the next iteration grouping, a simple method is applied to score the after-validation systems.

During the validation process, for each system, there will be a number of validation data sets (number = k_m) to be validated and calculating the RMSD number. By counting the number of the data sets, which passed the RMSD threshold, we could easily rank the system from covering the most data set to the system covering the least data set. Therefore, the highest rank system could cover up to a number of k_m data sets, and lowest rank system could cover at least 1 data set, since the system which could not cover any validation data set has been filtered out in the step 3.2.3.

In order to guarantee a potential possibility of the system which could be further improved in the next iteration, hereby, the top 3 system with highest rank will be selected.

After this step, the top 3 systems are selected. Remembering that the 3 systems are identified with a certain grouping of the data sets, and those combination of the data sets will be used as the bases for the next iteration.

For the next iteration, it starts again at the step 3.3.2.2, with the grouping identification loops.

3.3.2.2.6 Testing of the identified system with the testing data sets

After the identification iterations, the identified system come to a situation that could not further increase the covered data set. Then the identification process is finished, and the identified system shall be checked with the testing data sets to fine out the final performance of the identified system.

It shall be noticed that, the testing data sets shall be different from the identified data sets and the validation data sets, in order to guarantee an objective result.

3.3.3 Group identification Result

Remember that from the last chapter, the identification result based on the single data set is illustrated in the table T3.2.2.2-3, the data set Step-6, Step-7, Step-8, Step-9, Step-12, Step-13, Step-14 have the most covered RMSD value above the threshold. The Grouping identification start with the above 7 data sets, and combined with another data sets. Hereby, we use "iteration" to represent the grouping steps. And as we have discussed, the step measurement data sets are split into three big groups: identification data sets, cross validation data sets and the test data sets. The data sets are listed in the table T3.3.3-1.

Identification Data Sets & Cross check Data Sets	Test Data Sets
Step-1	Step-5
Step-2	Step-10
Step-3	Step-11
Step-4	
Step-6	
Step-7	
Step-8	
Step-9	
Step-12	
Step-13	
Step-14	
Step-15	

T3.3.3-1

The identification data sets and cross check data sets are used to group and identify the model, and cross validate for the next iteration calculation. The test data sets are only used by the end of the group identification for the final result check.

3.3.3.1 The 1st iteration result of group identification

As explained in the previous chapter, the 1st iteration of grouping is based on the following table T3.3.3.1-1:

Grouping Add-on Data Set \ Grouping Base Data Set	6	7	8	9	12	13	14
1	M-6-1	M-7-1	M-8-1	M-9-1	M-12-1	M-13-1	M-14-1
2	M-6-2	M-7-2	M-8-2	M-9-2	M-12-2	M-13-2	M-14-2
3	M-6-3	M-7-3	M-8-3	M-9-3	M-12-3	M-13-3	M-14-3
4	M-6-4	M-7-4	M-8-4	M-9-4	M-12-4	M-13-4	M-14-4
6	-	M-7-6	M-8-6	M-9-6	M-12-6	M-13-6	M-14-6
7	M-6-7	-	M-8-7	M-9-7	M-12-7	M-13-7	M-14-7
8	M-6-8	M-7-8	-	M-9-8	M-12-8	M-13-8	M-14-8
9	M-6-9	M-7-9	M-8-9	-	M-12-9	M-13-9	M-14-9
12	M-6-12	M-7-12	M-8-12	M-9-12	-	M-13-12	M-14-12
13	M-6-13	M-7-13	M-8-13	M-9-13	M-12-13	-	M-14-13
14	M-6-14	M-7-14	M-8-14	M-9-14	M-12-14	M-13-14	-
15	M-6-15	M-7-15	M-8-15	M-9-15	M-12-15	M-13-15	M-14-15

T3.3.3.1-1

And as explained in the previous chapter, the calculation process is similar:

- (1) Identify the model based on the grouped data sets.
- (2) Cross validate the model with the data sets input and generate displacement output $X_{ident}(n1-n2-nb-nk-m)$
- (3) Calculate the RMSD value between the measurement displacement of the validation data set (m) and the identified displacement.
- (4) Count the numbers of data sets whose RMSD value is below the threshold line (0.01).
- (5) Select the most covered data set groups for the next iteration

After the above steps, the following table (T3.3.3.1-2) which counting the numbers of cross check data sets (Step1 to 15, without Step5, 10 and 11) over the RMSD threshold value of 0.01 is listed.

Grouping Add-on Data Set \ Grouping Base Data Set	6	7	8	9	12	13	14
1	6	7	5	6	6	5	6
2	4	3	4	4	3	5	5
3	4	5	4	5	5	3	5
4	5	6	7	3	4	6	5
6	x	7	7	7	7	7	6
7	6	x	7	7	7	7	7
8	7	7	x	7	7	7	7
9	7	8	7	x	7	7	7
12	7	7	7	7	x	7	6
13	6	6	7	7	6	x	6
14	6	6	7	7	6	7	x
15	3	2	5	2	4	4	5

T3.3.3.1-2

From the table above, we could see that the combination of data set 7 with 9 shows a very outstanding coverage of the cross-check data sets. And some other combination such as data set 6-8, 6-9, 6-12, 7-1, 7-6, 7-8, 7-12, 8-4, 8-6, 8-7, etc., are also cover the cross-check data set 7 out of 12 with RMSD value under 0.01.

Therefore, for the next step of grouping, the above combination is considered as base data sets.

3.3.3.2 The 2nd iteration result of group identification

Based on the 1st iteration, the 2nd iteration is grouping based on the following table T3.3.3.2-1:

Grouping Add-on Data Set \ Grouping Base Data Set		1	2	3	4	6	7	8	9	12	13	14	15
6	8	6-8-1	6-8-2	6-8-3	6-8-4	6-8-6	6-8-7	6-8-8	6-8-9	6-8-12	6-8-13	6-8-14	6-8-15
6	9	6-9-1	6-9-2	6-9-3	6-9-4	6-9-6	6-9-7	6-9-8	6-9-9	6-9-12	6-9-13	6-9-14	6-9-15
6	12	6-12-1	6-12-2	6-12-3	6-12-4	6-12-6	6-12-7	6-12-8	6-12-9	6-12-12	6-12-13	6-12-14	6-12-15
7	1	7-1-1	7-1-2	7-1-3	7-1-4	7-1-6	7-1-7	7-1-8	7-1-9	7-1-12	7-1-13	7-1-14	7-1-15
7	6	7-6-1	7-6-2	7-6-3	7-6-4	7-6-6	7-6-7	7-6-8	7-6-9	7-6-12	7-6-13	7-6-14	7-6-15
7	8	7-8-1	7-8-2	7-8-3	7-8-4	7-8-6	7-8-7	7-8-8	7-8-9	7-8-12	7-8-13	7-8-14	7-8-15
7	9	7-9-1	7-9-2	7-9-3	7-9-4	7-9-6	7-9-7	7-9-8	7-9-9	7-9-12	7-9-13	7-9-14	7-9-15
7	12	7-12-1	7-12-2	7-12-3	7-12-4	7-12-6	7-12-7	7-12-8	7-12-9	7-12-12	7-12-13	7-12-14	7-12-15
8	4	8-4-1	8-4-2	8-4-3	8-4-4	8-4-6	8-4-7	8-4-8	8-4-9	8-4-12	8-4-13	8-4-14	8-4-15
8	6	8-6-1	8-6-2	8-6-3	8-6-4	8-6-6	8-6-7	8-6-8	8-6-9	8-6-12	8-6-13	8-6-14	8-6-15
8	7	8-7-1	8-7-2	8-7-3	8-7-4	8-7-6	8-7-7	8-7-8	8-7-9	8-7-12	8-7-13	8-7-14	8-7-15
8	9	8-9-1	8-9-2	8-9-3	8-9-4	8-9-6	8-9-7	8-9-8	8-9-9	8-9-12	8-9-13	8-9-14	8-9-15
8	12	8-12-1	8-12-2	8-12-3	8-12-4	8-12-6	8-12-7	8-12-8	8-12-9	8-12-12	8-12-13	8-12-14	8-12-15
8	13	8-13-1	8-13-2	8-13-3	8-13-4	8-13-6	8-13-7	8-13-8	8-13-9	8-13-12	8-13-13	8-13-14	8-13-15
8	14	8-14-1	8-14-2	8-14-3	8-14-4	8-14-6	8-14-7	8-14-8	8-14-9	8-14-12	8-14-13	8-14-14	8-14-15
9	6	9-6-1	9-6-2	9-6-3	9-6-4	9-6-6	9-6-7	9-6-8	9-6-9	9-6-12	9-6-13	9-6-14	9-6-15
9	7	9-7-1	9-7-2	9-7-3	9-7-4	9-7-6	9-7-7	9-7-8	9-7-9	9-7-12	9-7-13	9-7-14	9-7-15
9	8	9-8-1	9-8-2	9-8-3	9-8-4	9-8-6	9-8-7	9-8-8	9-8-9	9-8-12	9-8-13	9-8-14	9-8-15
9	12	9-12-1	9-12-2	9-12-3	9-12-4	9-12-6	9-12-7	9-12-8	9-12-9	9-12-12	9-12-13	9-12-14	9-12-15
9	13	9-13-1	9-13-2	9-13-3	9-13-4	9-13-6	9-13-7	9-13-8	9-13-9	9-13-12	9-13-13	9-13-14	9-13-15
9	14	9-14-1	9-14-2	9-14-3	9-14-4	9-14-6	9-14-7	9-14-8	9-14-9	9-14-12	9-14-13	9-14-14	9-14-15
12	6	12-6-1	12-6-2	12-6-3	12-6-4	12-6-6	12-6-7	12-6-8	12-6-9	12-6-12	12-6-13	12-6-14	12-6-15
12	7	12-7-1	12-7-2	12-7-3	12-7-4	12-7-6	12-7-7	12-7-8	12-7-9	12-7-12	12-7-13	12-7-14	12-7-15
12	8	12-8-1	12-8-2	12-8-3	12-8-4	12-8-6	12-8-7	12-8-8	12-8-9	12-8-12	12-8-13	12-8-14	12-8-15
12	9	12-9-1	12-9-2	12-9-3	12-9-4	12-9-6	12-9-7	12-9-8	12-9-9	12-9-12	12-9-13	12-9-14	12-9-15
13	6	13-6-1	13-6-2	13-6-3	13-6-4	13-6-6	13-6-7	13-6-8	13-6-9	13-6-12	13-6-13	13-6-14	13-6-15
13	7	13-7-1	13-7-2	13-7-3	13-7-4	13-7-6	13-7-7	13-7-8	13-7-9	13-7-12	13-7-13	13-7-14	13-7-15
13	8	13-8-1	13-8-2	13-8-3	13-8-4	13-8-6	13-8-7	13-8-8	13-8-9	13-8-12	13-8-13	13-8-14	13-8-15
13	9	13-9-1	13-9-2	13-9-3	13-9-4	13-9-6	13-9-7	13-9-8	13-9-9	13-9-12	13-9-13	13-9-14	13-9-15
13	12	13-12-1	13-12-2	13-12-3	13-12-4	13-12-6	13-12-7	13-12-8	13-12-9	13-12-12	13-12-13	13-12-14	13-12-15
13	14	13-14-1	13-14-2	13-14-3	13-14-4	13-14-6	13-14-7	13-14-8	13-14-9	13-14-12	13-14-13	13-14-14	13-14-15
14	7	14-7-1	14-7-2	14-7-3	14-7-4	14-7-6	14-7-7	14-7-8	14-7-9	14-7-12	14-7-13	14-7-14	14-7-15
14	8	14-8-1	14-8-2	14-8-3	14-8-4	14-8-6	14-8-7	14-8-8	14-8-9	14-8-12	14-8-13	14-8-14	14-8-15
14	9	14-9-1	14-9-2	14-9-3	14-9-4	14-9-6	14-9-7	14-9-8	14-9-9	14-9-12	14-9-13	14-9-14	14-9-15

T3.3.3.2-1

The same calculation procedure is applied as shown in the 1st iteration. And we obtained a RMSD coverage table T3.3.3.2-2:

Grouping Add-on Data Set \ Grouping Base Data Set		1	2	3	4	6	7	8	9	12	13	14	15
6	8	8	4	4	6	-	7	-	7	7	7	6	5
6	9	8	6	3	3	-	7	7	-	7	7	7	6
6	12	6	3	4	4	-	7	7	7	-	6	6	6
7	1	-	2	6	4	8	-	7	8	7	5	6	3
7	6	8	3	5	5	-	-	7	7	7	6	6	4
7	8	7	5	4	6	7	-	-	8	7	8	7	4
7	9	8	6	4	5	7	-	8	-	7	7	8	4
7	12	7	5	4	5	7	-	7	7	-	6	6	5
8	4	5	5	3	-	6	7	-	5	7	7	6	3
8	6	8	5	3	6	-	7	-	7	7	7	7	5
8	7	7	6	4	7	7	-	-	7	7	7	7	5
8	9	8	5	3	4	7	7	-	-	7	8	8	3
8	12	7	8	5	7	7	7	-	7	-	7	6	5
8	13	7	7	4	7	7	7	-	8	7	-	7	6
8	14	7	6	4	6	6	7	-	8	7	7	-	6
9	6	7	4	5	5	-	7	7	-	7	7	7	4
9	7	6	6	5	6	7	-	7	-	7	7	8	6
9	8	7	4	4	6	7	7	-	-	7	7	7	3
9	12	7	4	5	5	7	7	7	-	-	7	7	6
9	13	8	5	5	5	7	7	7	-	7	-	7	6
9	14	7	6	7	6	7	7	8	-	7	7	-	6
12	6	7	2	4	4	-	7	7	7	-	7	7	4
12	7	7	6	5	5	6	-	7	7	-	7	7	4
12	8	8	5	4	6	7	7	-	8	-	7	6	6
12	9	8	6	4	5	7	7	7	-	-	7	7	7
13	6	7	4	3	6	-	7	7	7	6	-	6	6
13	7	5	5	4	6	7	-	7	7	6	-	7	5
13	8	8	7	5	7	7	7	-	8	7	-	7	5
13	9	7	5	7	6	7	8	7	-	7	-	8	6
13	12	6	5	4	6	5	6	7	7	-	-	6	4
13	14	7	5	4	5	6	6	7	7	7	-	-	6
14	7	6	6	6	5	6	-	7	7	6	6	-	6
14	8	6	6	5	6	6	7	-	8	7	7	-	6
14	9	8	7	6	5	7	8	8	-	7	7	-	6

T3.3.3.2-2

Based on the result table, we could observe that although the maximum number of cross-check data set whose RMSD number below the threshold keep the same as 8. However, the identified model which cover the cross-check data set below RMSD with the number of 8, has largely increased. Therefore, to further check the result, the combination which covers 8 data sets has been chosen for the next 3rd iteration identification.

3.3.3.3 The 3rd iteration result of group identification

In the 3rd iteration identification, the following combination table (T3.3.3.3-1) is chosen, based on the 2nd iteration.

Grouping Add-on Data Set \ Grouping Base Data Set			1	2	3	4	6	7	8	9	12	13	14	15
6	8	1	-	6-8-1-2	6-8-1-3	6-8-1-4	-	6-8-1-7	-	6-8-1-9	6-8-1-12	6-8-1-13	6-8-1-14	6-8-1-15
6	9	1	-	6-9-1-2	6-9-1-3	6-9-1-4	-	6-9-1-7	6-9-1-8	-	6-9-1-12	6-9-1-13	6-9-1-14	6-9-1-15
7	1	6	-	7-1-6-2	7-1-6-3	7-1-6-4	-	-	7-1-6-8	7-1-6-9	7-1-6-12	7-1-6-13	7-1-6-14	7-1-6-15
7	1	9	-	7-1-9-2	7-1-9-3	7-1-9-4	7-1-9-6	-	7-1-9-8	-	7-1-9-12	7-1-9-13	7-1-9-14	7-1-9-15
7	6	1	-	7-6-1-2	7-6-1-3	7-6-1-4	-	-	7-6-1-8	7-6-1-9	7-6-1-12	7-6-1-13	7-6-1-14	7-6-1-15
7	8	9	7-8-9-1	7-8-9-2	7-8-9-3	7-8-9-4	7-8-9-6	-	-	-	7-8-9-12	7-8-9-13	7-8-9-14	7-8-9-15
7	8	13	7-8-13-1	7-8-13-2	7-8-13-3	7-8-13-4	7-8-13-6	-	-	7-8-13-9	7-8-13-12	-	7-8-13-14	7-8-13-15
7	9	1	-	7-9-1-2	7-9-1-3	7-9-1-4	7-9-1-6	-	7-9-1-8	-	7-9-1-12	7-9-1-13	7-9-1-14	7-9-1-15
7	9	8	7-9-8-1	7-9-8-2	7-9-8-3	7-9-8-4	7-9-8-6	-	-	-	7-9-8-12	7-9-8-13	7-9-8-14	7-9-8-15
7	9	14	7-9-14-1	7-9-14-2	7-9-14-3	7-9-14-4	7-9-14-6	-	7-9-14-8	-	7-9-14-12	7-9-14-13	-	7-9-14-15
8	6	1	-	8-6-1-2	8-6-1-3	8-6-1-4	-	8-6-1-7	-	8-6-1-9	8-6-1-12	8-6-1-13	8-6-1-14	8-6-1-15
8	9	1	-	8-9-1-2	8-9-1-3	8-9-1-4	8-9-1-6	8-9-1-7	-	-	8-9-1-12	8-9-1-13	8-9-1-14	8-9-1-15
8	9	13	8-9-13-1	8-9-13-2	8-9-13-3	8-9-13-4	8-9-13-6	8-9-13-7	-	-	8-9-13-12	-	8-9-13-14	8-9-13-15
8	9	14	8-9-14-1	8-9-14-2	8-9-14-3	8-9-14-4	8-9-14-6	8-9-14-7	-	-	8-9-14-12	8-9-14-13	-	8-9-14-15
8	12	2	8-12-2-1	-	8-12-2-3	8-12-2-4	8-12-2-6	8-12-2-7	-	8-12-2-9	-	8-12-2-13	8-12-2-14	8-12-2-15
8	13	9	8-13-9-1	8-13-9-2	8-13-9-3	8-13-9-4	8-13-9-6	8-13-9-7	-	-	8-13-9-12	-	8-13-9-14	8-13-9-15
8	14	9	8-14-9-1	8-14-9-2	8-14-9-3	8-14-9-4	8-14-9-6	8-14-9-7	-	-	8-14-9-12	8-14-9-13	8-14-9-14	8-14-9-15
9	7	14	9-7-14-1	9-7-14-2	9-7-14-3	9-7-14-4	9-7-14-6	-	9-7-14-8	-	9-7-14-12	9-7-14-13	-	9-7-14-15
9	13	1	-	9-13-1-2	9-13-1-3	9-13-1-4	9-13-1-6	9-13-1-7	9-13-1-8	-	9-13-1-12	-	9-13-1-14	9-13-1-15
9	14	8	9-14-8-1	9-14-8-2	9-14-8-3	9-14-8-4	9-14-8-6	9-14-8-7	-	-	9-14-8-12	9-14-8-13	9-14-8-14	9-14-8-15
12	8	1	-	12-8-1-2	12-8-1-3	12-8-1-4	12-8-1-6	12-8-1-7	-	12-8-1-9	-	12-8-1-13	12-8-1-14	12-8-1-15
12	8	9	12-8-9-1	12-8-9-2	12-8-9-3	12-8-9-4	12-8-9-6	12-8-9-7	-	-	-	12-8-9-13	12-8-9-14	12-8-9-15
12	9	1	-	12-9-1-2	12-9-1-3	12-9-1-4	12-9-1-6	12-9-1-7	12-9-1-8	-	-	12-9-1-13	12-9-1-14	12-9-1-15
13	8	1	-	13-8-1-2	13-8-1-3	13-8-1-4	13-8-1-6	13-8-1-7	-	13-8-1-9	13-8-1-12	-	13-8-1-14	13-8-1-15
13	8	9	13-8-9-1	13-8-9-2	13-8-9-3	13-8-9-4	13-8-9-6	13-8-9-7	-	-	13-8-9-12	-	13-8-9-14	13-8-9-15
13	9	7	13-9-7-1	13-9-7-2	13-9-7-3	13-9-7-4	13-9-7-6	-	13-9-7-8	-	13-9-7-12	-	13-9-7-14	13-9-7-15
13	9	14	13-9-14-1	13-9-14-2	13-9-14-3	13-9-14-4	13-9-14-6	13-9-14-7	13-9-14-8	-	13-9-14-12	-	-	13-9-14-15
14	8	9	14-8-9-1	14-8-9-2	14-8-9-3	14-8-9-4	14-8-9-6	14-8-9-7	-	-	14-8-9-12	14-8-9-13	-	14-8-9-15
14	9	1	-	14-9-1-2	14-9-1-3	14-9-1-4	14-9-1-6	14-9-1-7	14-9-1-8	-	14-9-1-12	14-9-1-13	-	14-9-1-15
14	9	7	14-9-7-1	14-9-7-2	14-9-7-3	14-9-7-4	14-9-7-6	-	14-9-7-8	-	14-9-7-12	14-9-7-13	-	14-9-7-15
14	9	8	14-9-8-1	14-9-8-2	14-9-8-3	14-9-8-4	14-9-8-6	14-9-8-7	-	-	14-9-8-12	14-9-8-13	-	14-9-8-15

T3.3.3.3-1

The same calculation procedure is applied as shown in the previous iteration. And we obtained a RMSD coverage table T3.3.3.3-2:

Grouping Add-on Data Set \ Grouping Base Data Set			1	2	3	4	6	7	8	9	12	13	14	15
6	8	1	-	4	4	7	-	8	-	8	8	7	7	6
6	9	1	-	6	4	4	-	8	8	-	8	8	8	6
7	1	6	-	4	5	6	-	-	8	8	7	7	7	4
7	1	9	-	5	4	6	8	-	6	-	8	8	7	5
7	6	1	-	4	5	7	-	-	8	8	8	7	7	4
7	8	9	6	7	7	5	7	-	-	-	7	8	8	6
7	8	13	7	5	5	7	7	-	-	7	7	-	6	6
7	9	1	-	5	5	6	8	-	7	-	8	8	7	4
7	9	8	7	6	7	6	8	-	-	-	7	7	8	7
7	9	14	7	6	5	6	7	-	8	-	7	7	-	7
8	6	1	-	5	4	7	-	8	-	7	8	8	7	5
8	9	1	-	4	3	5	7	6	-	-	7	7	7	4
8	9	13	8	7	5	6	7	8	-	-	8	-	8	6
8	9	14	7	7	5	6	7	8	-	-	7	8	-	6
8	12	2	5	-	2	5	5	6	-	7	-	6	6	3
8	13	9	8	6	6	6	7	8	-	-	8	-	8	6
8	14	9	8	7	5	6	7	8	-	-	7	7	-	6
9	7	14	7	6	6	6	7	-	8	-	7	7	-	7
9	13	1	-	5	4	6	8	7	7	-	8	-	8	6
9	14	8	7	5	6	6	7	8	-	-	7	8	-	5
12	8	1	-	5	4	7	8	7	-	7	-	7	7	6
12	8	9	7	6	6	5	7	7	-	-	-	8	7	7
12	9	1	-	7	4	5	8	7	7	-	-	8	8	6
13	8	1	-	6	6	6	8	7	-	8	8	-	7	5
13	8	9	8	6	7	6	7	7	-	-	7	-	8	6
13	9	7	8	6	5	6	7	-	8	-	7	-	7	7
13	9	14	7	6	4	6	7	8	8	-	7	-	-	8
14	8	9	7	6	6	7	7	7	-	-	8	7	-	6
14	9	1	-	6	5	6	8	8	7	-	7	8	-	5
14	9	7	7	7	6	7	7	-	8	-	8	7	-	7
14	9	8	7	6	6	6	7	8	-	-	7	8	-	7

T3.3.3.3-2

It could be observed that, the maximum number of models which have RMSD below threshold still keep the same as 8. This indicate that the identification may approximate the best solution situation. And it is worthy to further deeply trace the model with the test data sets.

3.3.3.4 The test result of group identification

In this sub-chapter, the target is to verify the identification model with the pre-selected data sets. The identification model is chosen with the following table T3.3.3.4-1:

Identification Model\Test Data Set	nb	nk	Test 5	Test 10	Test 11
6	3	7	0.168	0.0904	0.0054
6-8	4	3	0.094	0.0383	0.0253
6-8-1	1	6	0.114	0.0431	0.0077
6-8-1-12	1	5	0.067	0.0276	0.0068

T3.3.3.4-1

In the table T3.3.3.4-1, the first column represents the data set used for model identification. And the column n_b and n_k represent the identification parameter. The Test represents the data set used for testing. And in the table, we could see the RMSD result calculated.

Hereby, it is necessary to emphasize that the test data set is only tested after the model is identified. The process of grouping from single data sets (e.g. data set 6) to the combination of four data sets with order (e.g. data set group 6-8-1-12) has no influence from the test data sets. This is very important for the independency of the result.

We could see from the result RMSD value that: with the grouping of the data sets, the identified model shows a decreasing trend of RMSD, which represents an increasing trend of accuracy (Fig.3.3.3.4-1).

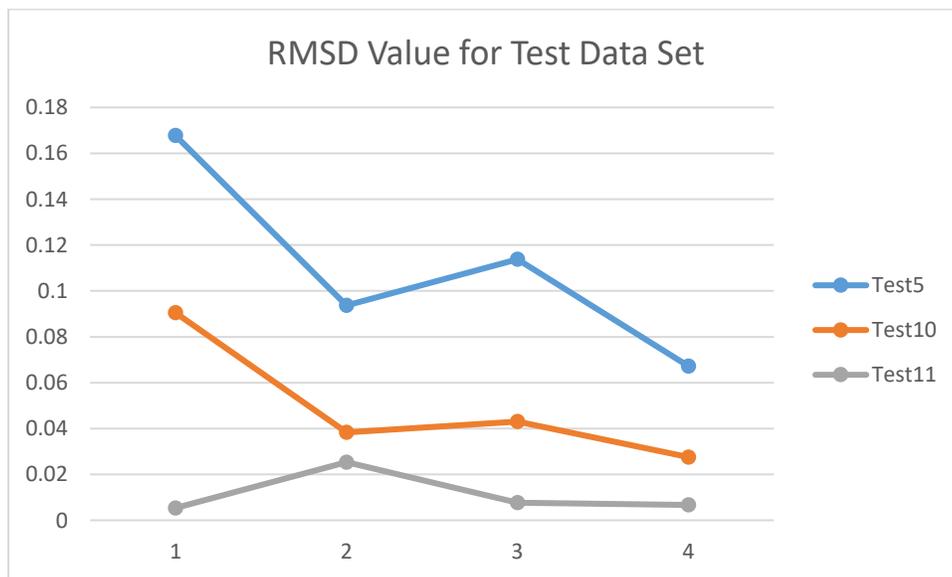


Fig.3.3.3.4-1

And for a more directly over view of the result, in the Fig.3.3.3.4-2, the displacement of the identified model against the displacement from real measurement for the test data set 5, 10 and 11 is shown.

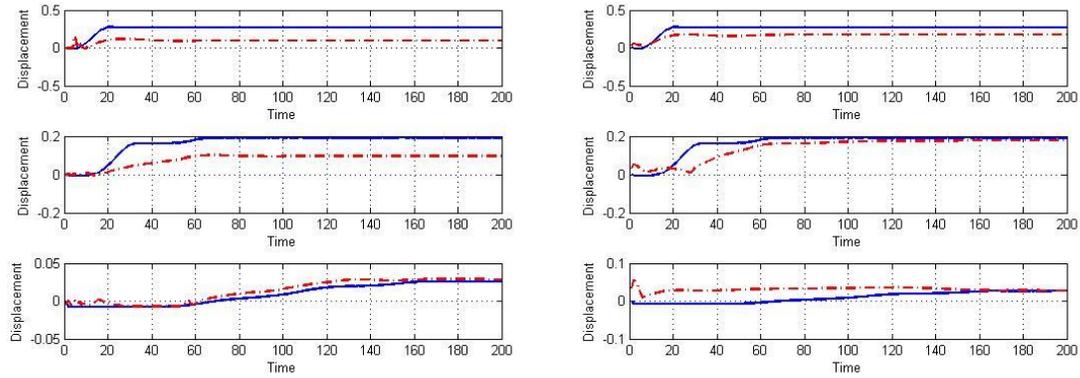


Fig.3.3.3.4-2(1)-(2)

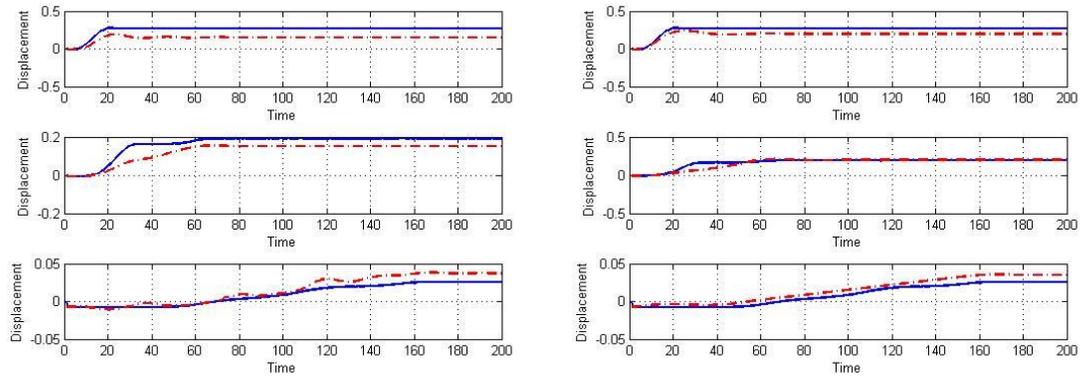


Fig.3.3.3.4-2(3)-(4)

In the plot we could see that the accuracy of the model is largely increased, as the identified model result (red dash line) is approaching the measurement result (blue line). And the also the deviation of the final stable value is also narrowing from the sub-plot (1) to (4), which represent the model 6 to 6-9-1-12.

Up to here, we could conclude that the combination of the data sets for identification could largely help to improve the coverage of the identified model.

Further, we would like to check how is the model result test against all the data sets, in order to understand if the model could cover all the situation. And in the next 2 chapter (Chapter 3.4.1 and 3.4.2), we will identify two combination models based the step data sets (Combination Model 1), and sinusoidal and step data sets (Combination Model 2).

3.4 Grouped measurement system identification

In the last chapter, the grouped identification method is proved to largely improve the coverage of the model to the data sets. In this chapter, we would like to stepwise check the group identified model against all the data sets, in order to further understand the improvement process, and the detailed coverage.

3.4.1 Step based grouped identified system

Same in the chapter 3.3, we applied the grouping method to identified a model, which we named as *Combination Model 1*. And the model is actually identified stepwise based on the combination of datasets: 6, 9, 1, 8.

	Step Data Coverage
6	6
6-9	7
6-9-1	8
6-9-1-8	8

T3.4.1-1

From the table T3.4.1-1, we could see that with the increasing of the data set combination, the coverage of the step data sets is increasing (Fig.3.4.1-1). However, after the iteration 3, the coverage is saturated. This means, the maximum coverage of the identified model is up to its limitation, considering the minimum requirement of the RMSE value (accuracy).

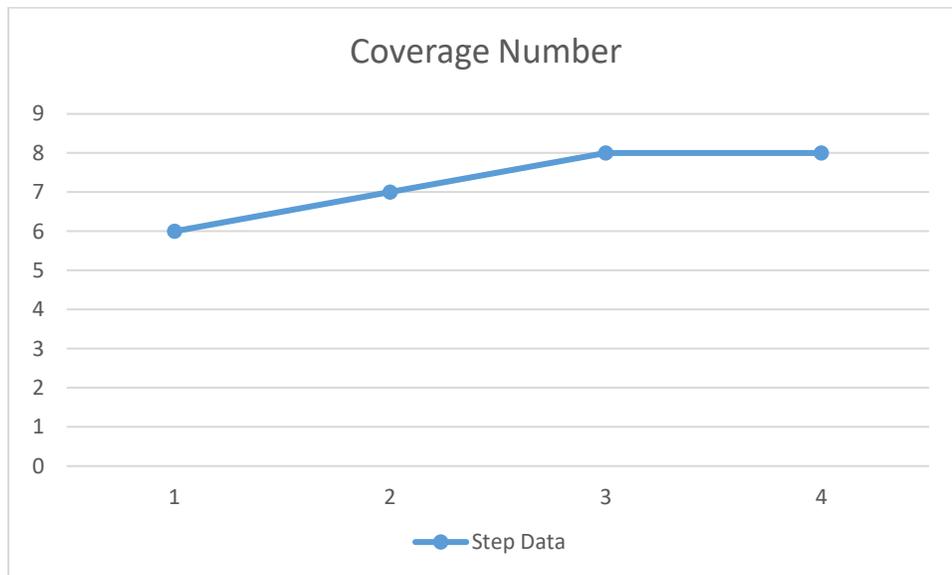


Fig.3.4.1-1

To detailly understand the trend, the RMSE values of the 4 models against each data sets are listed in the following table T3.4.1-2,3:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
6	0.017041463	0.033887637	0.050753572	0.070468079	0.113190061	0.000106189	0.003778575	0.006694683	0.010341803	0.042854148	0.007286747	0.002637115	0.004300515	0.005409193	0.037543559
6-9	0.016688242	0.026020847	0.043220432	0.063691854	0.108298822	0.002420408	0.007225938	0.003842788	0.005964418	0.0384747	0.001919067	0.007522354	0.009173517	0.009706264	0.033750345
6-9-1	0.007463	0.025534	0.042588	0.060849	0.10304	0.00775	0.007622	0.003875	0.004861	0.034787	0.009165	0.008193	0.009678	0.011065	0.030792
6-9-1-8	0.007295512	0.024772831	0.042670237	0.063569911	0.108151586	0.008123489	0.008499559	0.004105536	0.005488927	0.039521207	0.009209917	0.00890172	0.009935087	0.009749277	0.033236488

T3.4.1-2

And for Sinusoidal data sets:

	1	2	3	4	5	6	7	8	9	10	11	12
6	0.026011653	0.020540931	0.031365056	0.041262816	0.056660472	0.026122504	0.021813263	0.039765336	0.053880109	0.072408547	0.010967865	0.017432715
6-9	0.019764191	0.011279217	0.013617633	0.017981558	0.022075052	0.019854854	0.010593411	0.008957487	0.013288128	0.016778805	0.011450924	0.018013519
6-9-1	0.019146214	0.005394033	0.004794244	0.007684831	0.010931175	0.019420342	0.006669698	0.007096035	0.012307929	0.016380135	0.007719219	0.010556885
6-9-1-8	0.017028956	0.003693736	0.004015428	0.005651887	0.007700901	0.017246544	0.005529444	0.008062551	0.011704686	0.014353761	0.008450916	0.01316499

	13	14	15	16	17	18	19	20	21	22	23
6	0.068382329	0.088391862	0.103975208	0.018735767	0.01747712	0.045309965	0.030287721	0.023571335	0.022725503	0.041262816	0.022746709
6-9	0.032959805	0.050775458	0.065686674	0.016403894	0.016499237	0.031503153	0.025094159	0.020324969	0.022011123	0.017981558	0.02334468
6-9-1	0.056578444	0.067908015	0.016161805	0.014146467	0.03207536	0.024791549	0.02016834	0.020991435	0.007684831	0.025194769	0.056578444
6-9-1-8	0.033936222	0.044618256	0.051911729	0.013985778	0.011519017	0.035899631	0.02319617	0.019677712	0.020140539	0.005651887	0.02368942

T3.4.1-3

It would be clearly seen that the data sets which has the lower RMSE (higher accuracy) in the beginning of the grouping model, are increasing its RMSE value (decreasing in accuracy) (e.g. the step data set Nr.6, 14) compared to the end of the grouping model, to compromise with more coverage of the data sets (e.g. the step data sets Nr.1, 9, sin data sets Nr. 16, 22, etc.). And even though, some data sets are still not covered with the threshold value (e.g. the step data sets Nr.3, 4, 5, sin data sets 1, 6, 10, etc.). Based on this, it could be more interesting to check the figure directly.

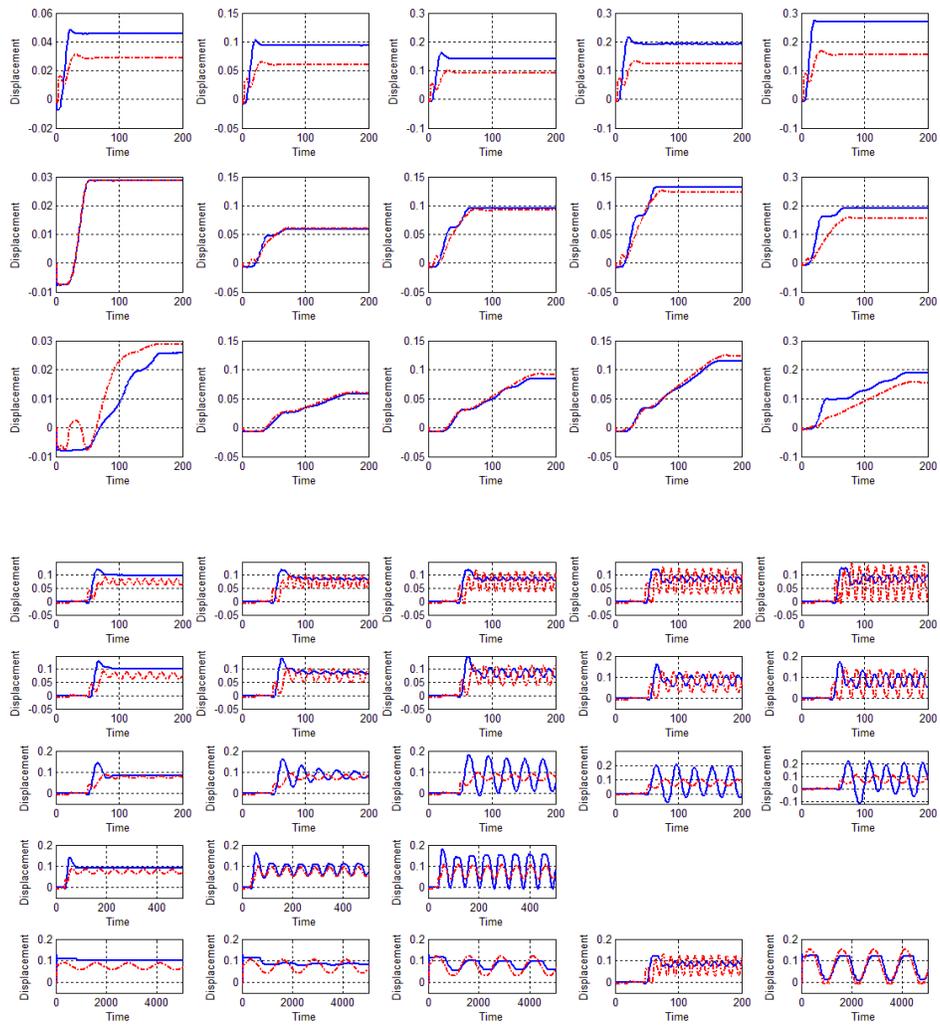


Fig.3.4.1-2(1)

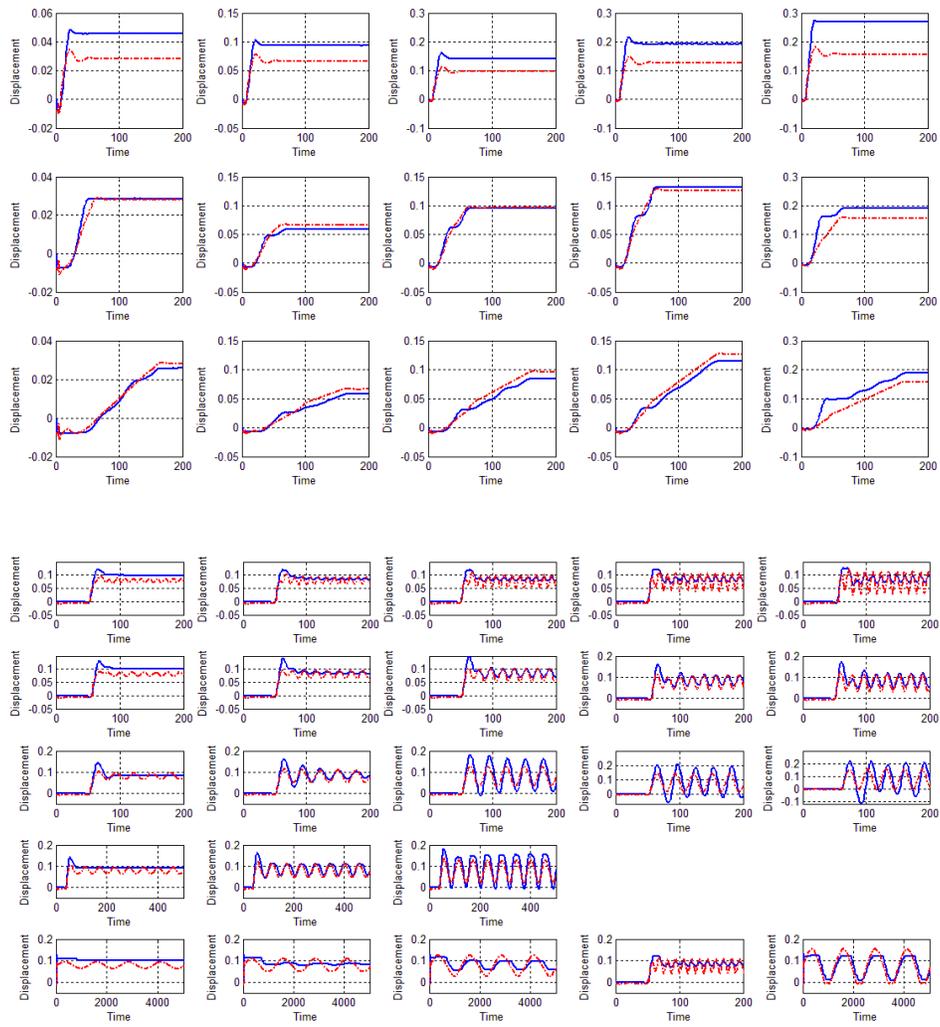


Fig.3.4.1-2(2)

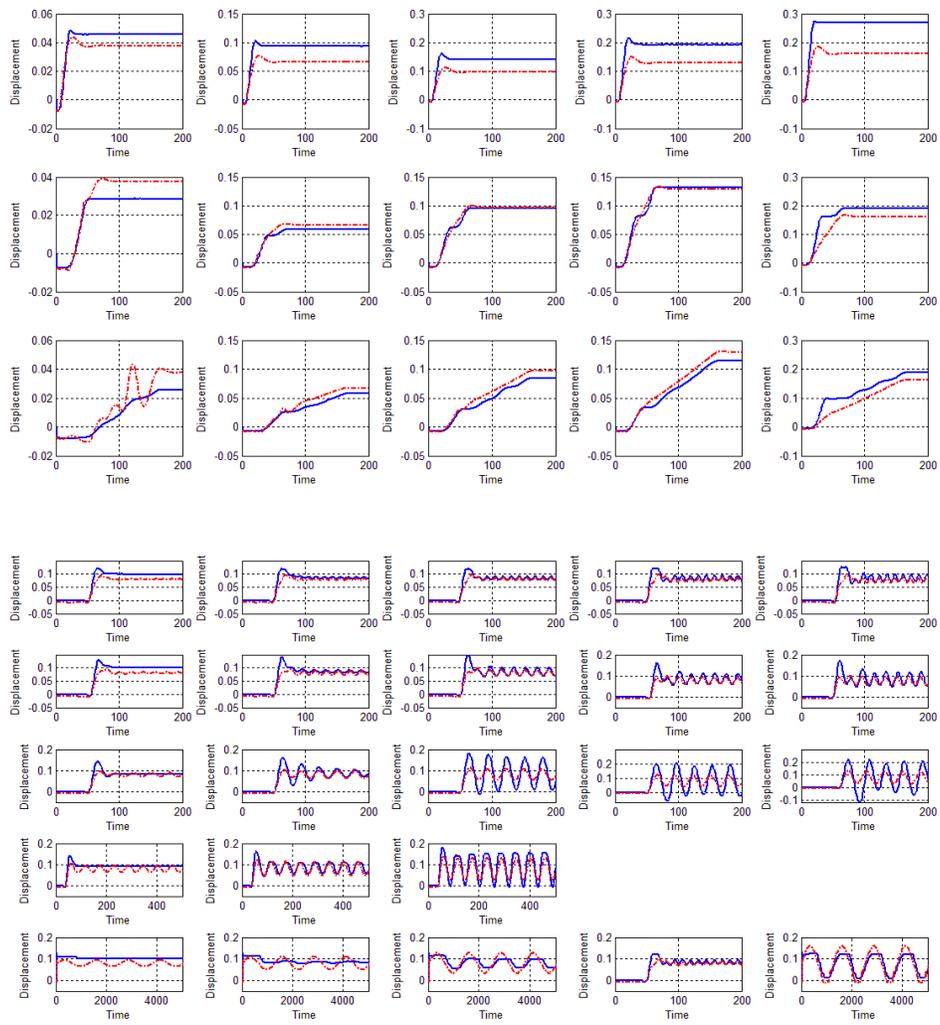


Fig.3.4.1-2(3)

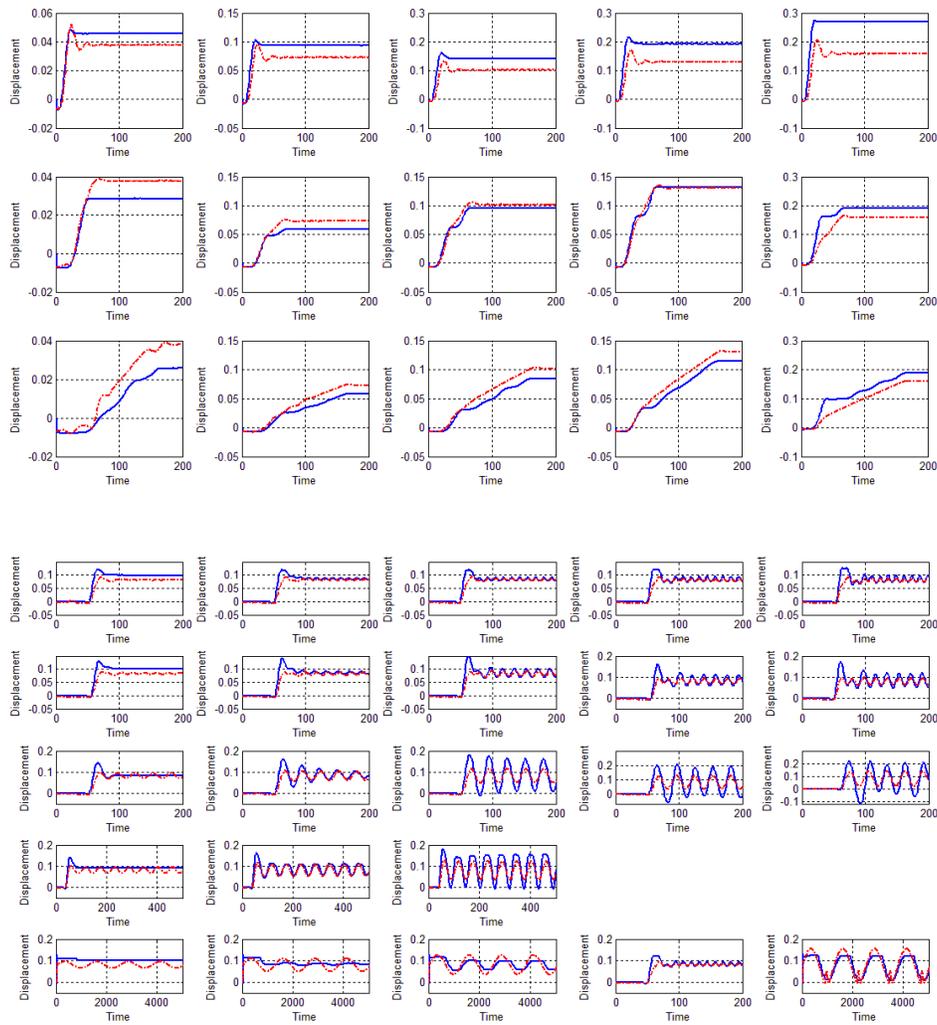


Fig.3.4.1-2(4)

From the figures above, we could observe the following trend:

1. As designed, the coverage of the cross-check data sets is increasing with more identify data sets combination.
2. Although the coverage is increasing and finally saturated with the number of coverages, it is still not enough to cover all the cross-check data sets.
3. With the increasing of the identify data sets combination, some cross-check data sets show a slightly decrease in accuracy (increasing in RMSE value), which could be considered as a compromise for coverage.

In the next sub-chapter, an identified combination model (*Combination Model 2*) based on both step and sinusoidal data sets will be shown.

3.4.2 Step and sinusoidal based grouped identified system

In this chapter, the result of identified combination model (*Combination Model 2*) based on step and sinusoidal data sets will be shown and discussed.

Same as last chapter, the model 2 is also identified based the same procedure, with the combination of data sets: 20, 16, 9, 35, 7. The data sets 20, 16, 35 are sinusoidal data sets, and the 7 and 9 are step data sets.

The coverage of the model stepwise result is shown in table T3.4.2-1:

	Step & Sin Data Coverage
20	14
20-16	18
20-16-9	19
20-16-9-5	21
20-16-9-5-35	23
20-16-9-5-35-7	23

T3.4.2-1

And the increasing trend of number of data sets coverage could be shown as Fig.3.4.2-1:

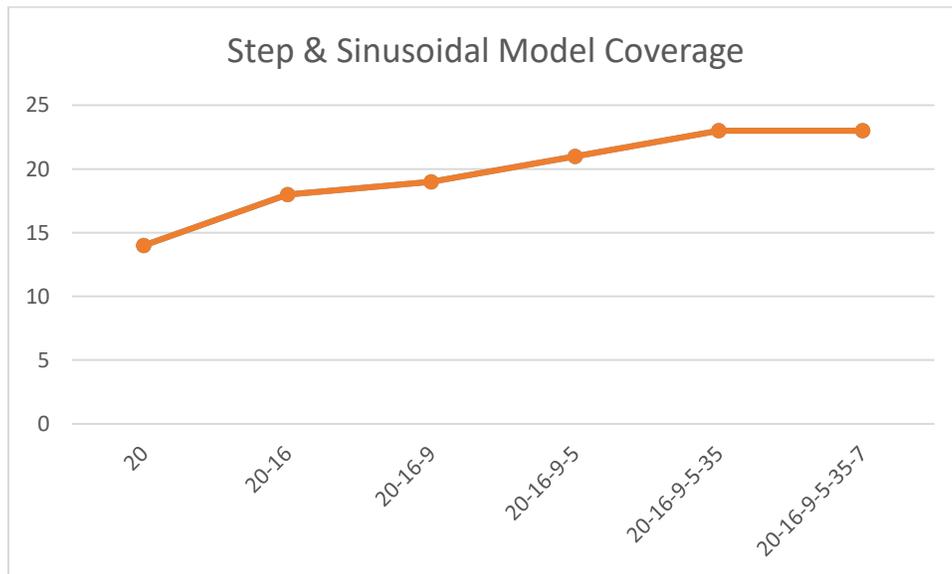


Fig.3.4.2-1

And in the following table T3.4.2-1, the RMSE value decreasing trend could be observed:

Step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
20	0.009	0.02	0.037	0.053	0.059	0.008	0.014	0.01	0.009	0.025	0.008	0.013	0.016	0.018	0.025
20-16	0.045	0.008	0.044	0.047	0.078	0.026	0.035	0.016	0.016	0.018	0.017	0.026	0.023	0.026	0.031
20-16-9	0.01	0.007	0.045	0.062	0.096	0.008	0.032	0.01	0.007	0.027	0.01	0.022	0.019	0.017	0.027
20-16-9-5	0.008	0.004	0.047	0.07	0.045	0.021	0.031	0.009	0.011	0.036	0.02	0.023	0.021	0.017	0.033
20-16-9-5-35	0.006	0.002	0.046	0.059	0.005	0.01	0.031	0.009	0.005	0.069	0.009	0.027	0.022	0.02	0.047
20-16-9-5-35-7	0.01	0.003	0.046	0.06	0.006	0.007	0.028	0.009	0.004	0.068	0.006	0.024	0.02	0.02	0.045

Sin	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
20	0.011	0.003	0.002	0.002	5E-04	0.012	0.006	0.004	0.003	0.003	0.017	0.024	0.015	0.03	0.052
20-16	7E-04	0.007	0.003	0.002	9E-04	0.002	0.009	0.006	0.005	0.004	0.014	0.019	0.021	0.041	0.061
20-16-9	0.002	0.003	0.002	0.003	0.002	0.004	0.005	0.004	0.003	0.003	0.012	0.014	0.028	0.046	0.065
20-16-9-5	0.001	0.003	0.004	0.006	0.003	0.002	0.004	0.004	0.01	0.009	0.015	0.005	0.043	0.06	0.071
20-16-9-5-35	0.004	0.006	0.003	0.007	0.002	0.005	0.005	0.004	0.009	0.005	0.011	0.006	0.045	0.06	0.075
20-16-9-5-35-7	0.004	0.006	0.005	0.006	0.002	0.005	0.006	0.004	0.009	0.008	0.01	0.006	0.041	0.058	0.071

	16	17	18	19	20	21	22	23
20	0.017	0.021	0.029	0.019	0.019	0.021	0.002	0.037
20-16	0.008	0.018	0.021	0.005	0.021	0.036	0.002	0.038
20-16-9	0.006	0.016	0.029	0.007	0.017	0.021	0.003	0.032
20-16-9-5	0.009	0.013	0.039	0.007	0.016	0.019	0.006	0.031
20-16-9-5-35	0.005	0.017	0.04	0.007	0.01	0.017	0.007	0.046
20-16-9-5-35-7	0.005	0.016	0.033	0.008	0.01	0.023	0.006	0.038

T3.4.2-1

And in the following figure Fig.3.4.2-1, we could see the progress with the cross-check data sets step and sinusoidal.

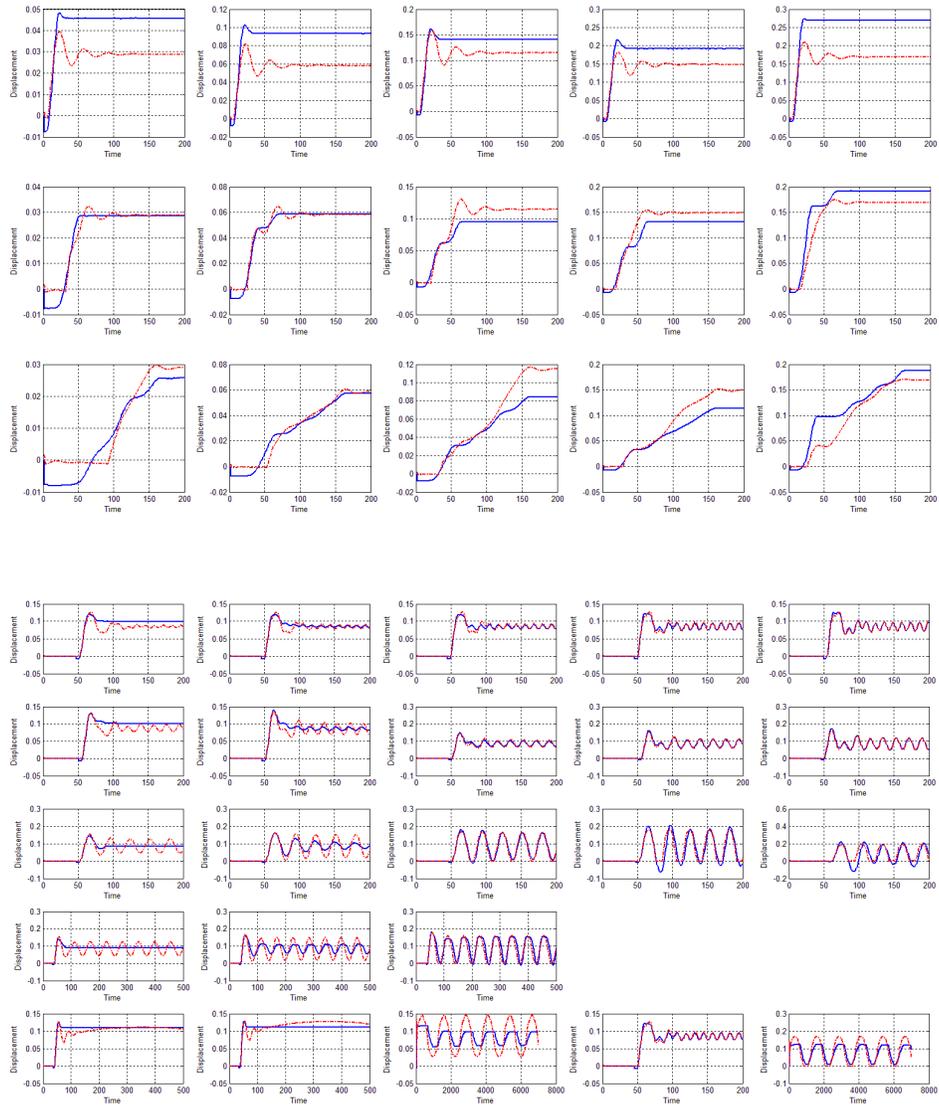


Fig.3.4.2-1(1)

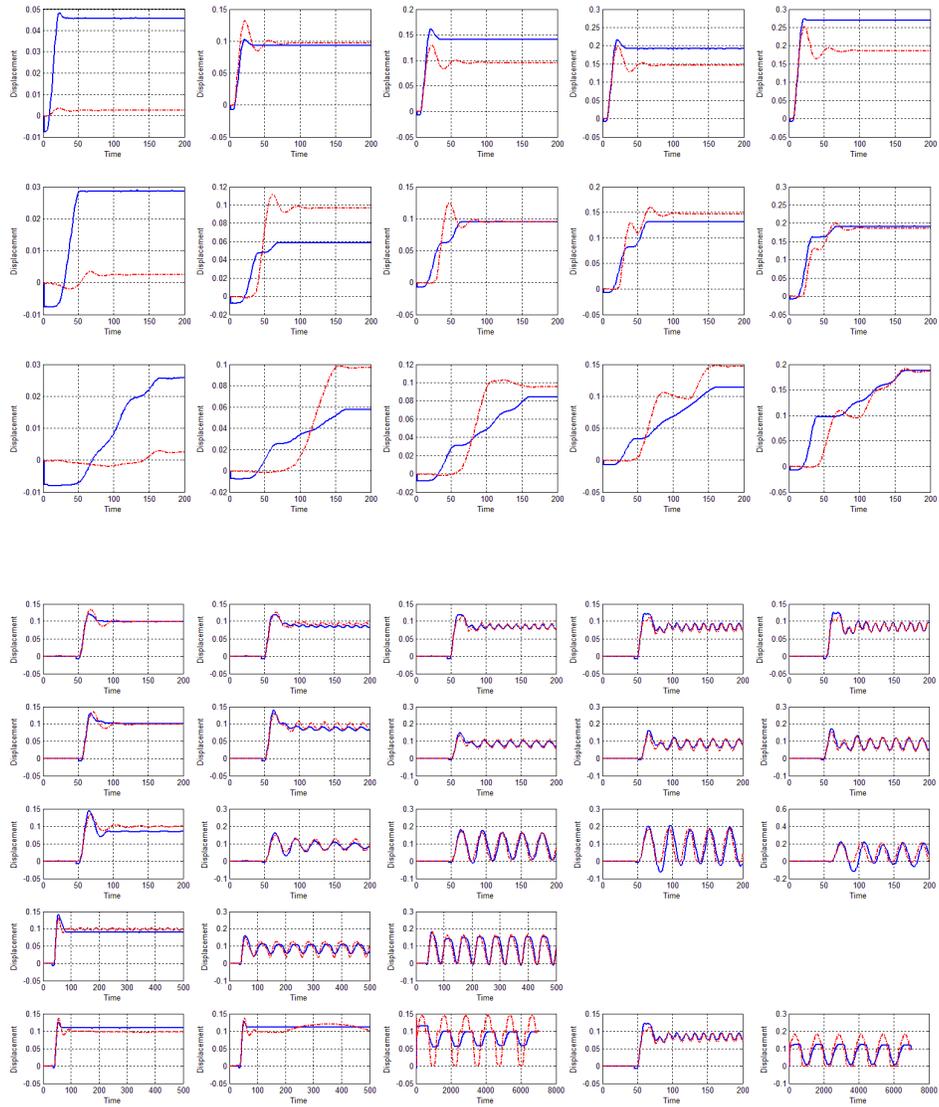


Fig.3.2.2-1(2)

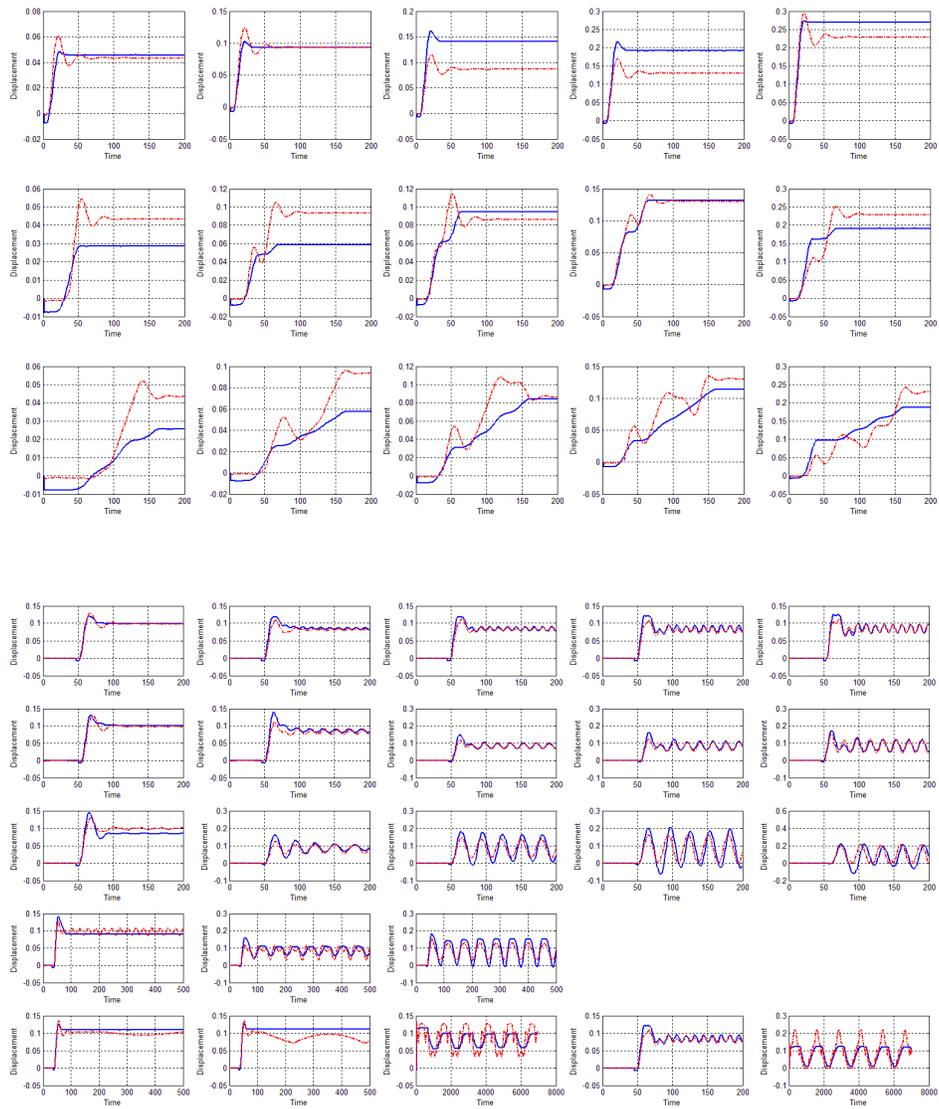


Fig.3.4.2-1(3)

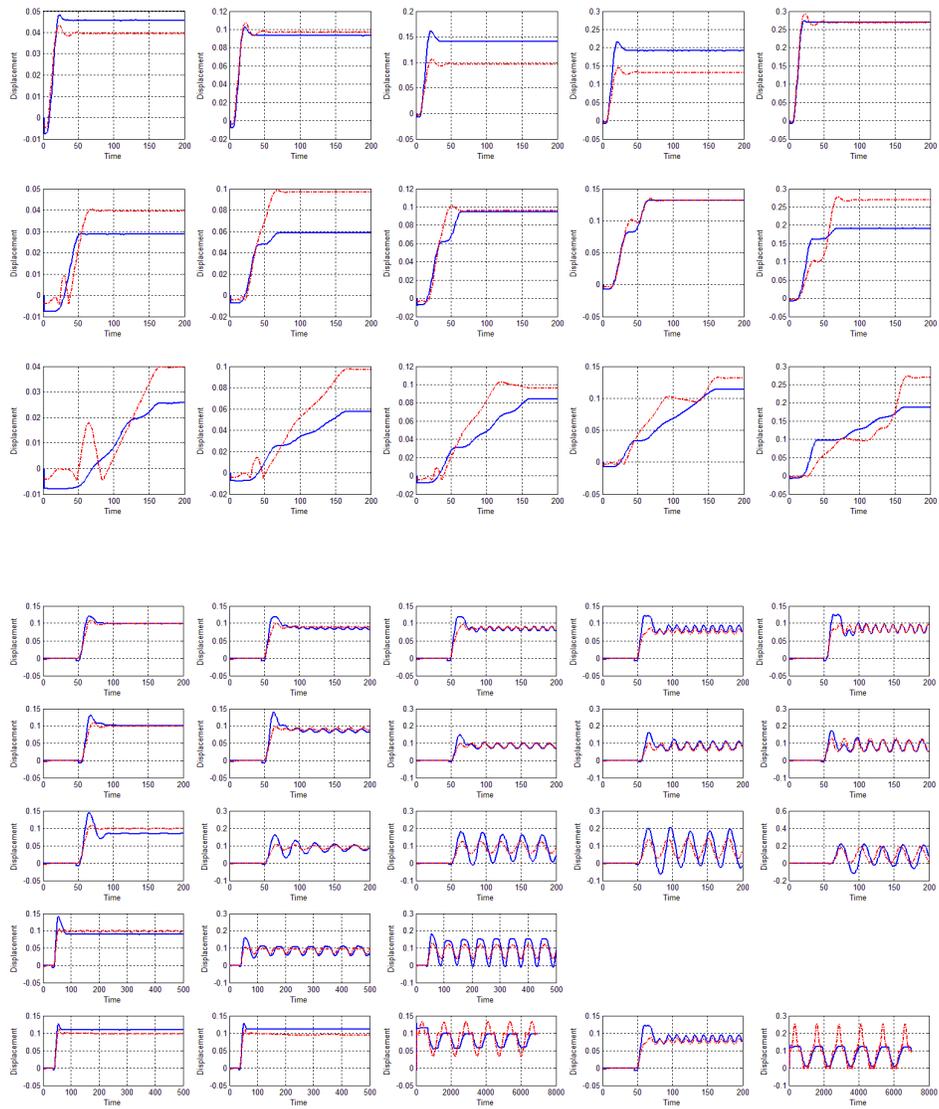


Fig.3.2.2-1(4)

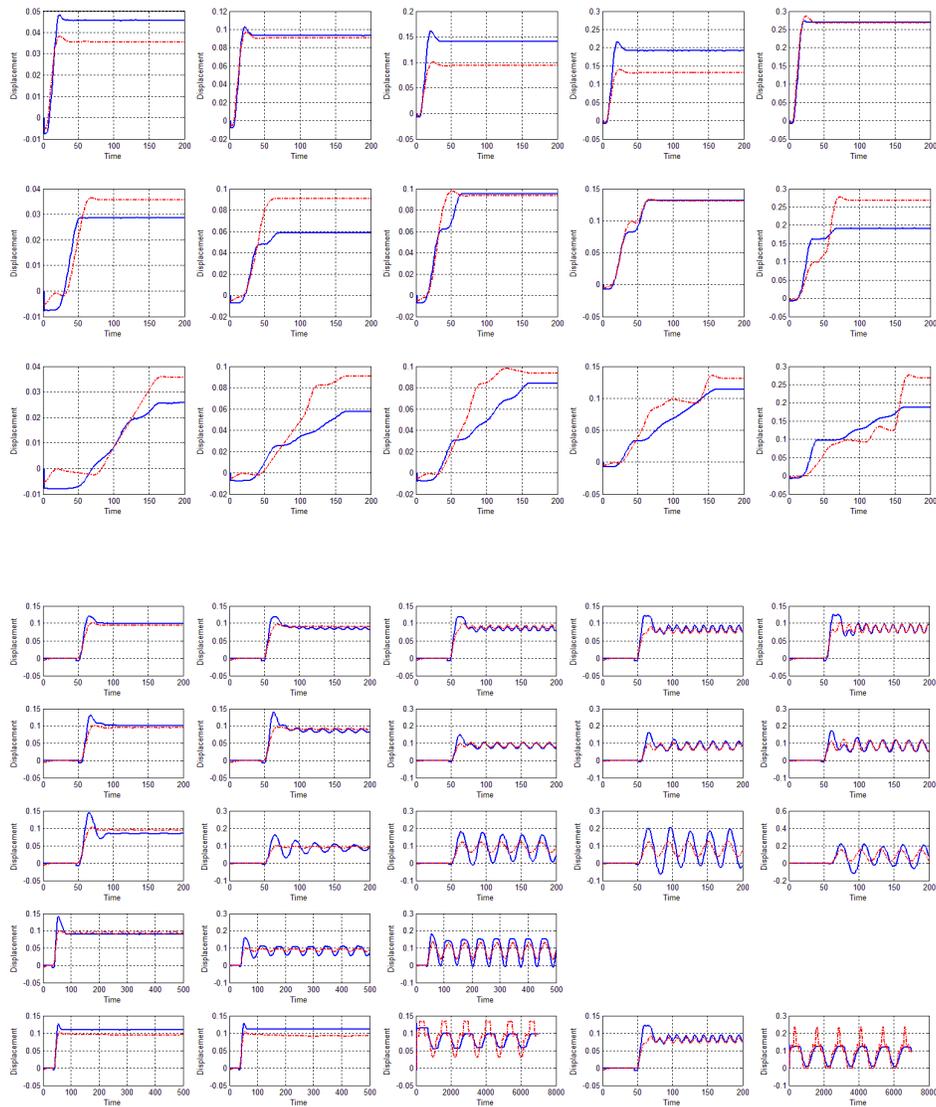


Fig.3.2.2-1(5)

Same as the previous subchapter, the trend is also clear:

1. The coverage of the *Combination Model 2* is increasing as designed.
2. The coverage number of the *Combination Model* come to saturation, even though more identification data sets are added.
3. With coverage number increasing, some data sets are shown a decreasing in accuracy, at the same time more data sets are shown as increasing in accuracy. Therefore, a compromise situation is generated.

In the next subchapter, the discussion of the combination will be given.

3.4.3 Discussion in grouped identified system

Based on the previous chapters, the following points could be observed and summarized as:

1. The data set combination method is working well to identify a model to cover more data sets.
2. With the combination of identified data sets increasing, the cross-check data sets show a different trend: the uncovered data sets are increasing in accuracy (Fig.3.4.3-1(1) (2)), the covered data sets are decreasing in accuracy (Fig.3.4.3-2(1) (2)). And it comes to saturation, as no more coverage could be reached, even though not all the data sets are covered.
3. Some of the data sets show a great different than the other data sets, where they could not be covered by any of the model, which means, not a unique model could cover all the data sets, even the compromising of limited accuracy is applied.

The support evidence for point 2 could be seen in the following figures.

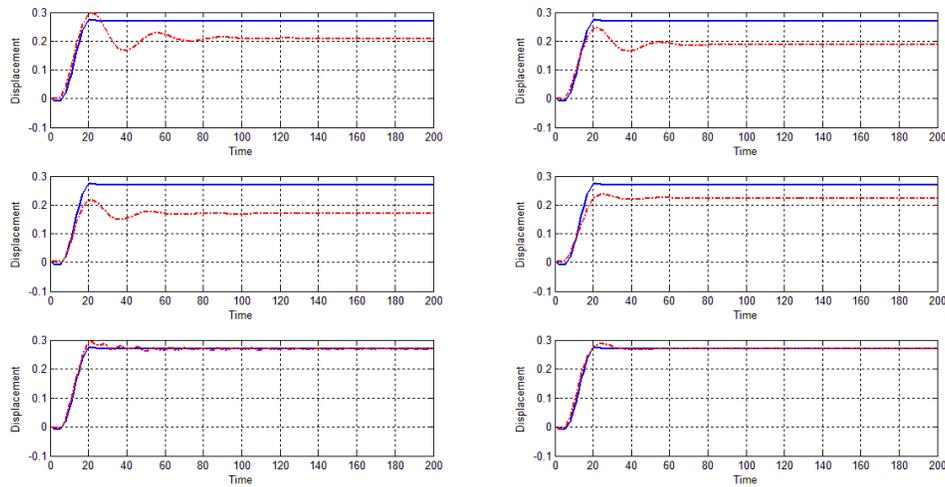


Fig.3.4.3-1(1) Increasing of accuracy for data set 5, against increasing of combination models

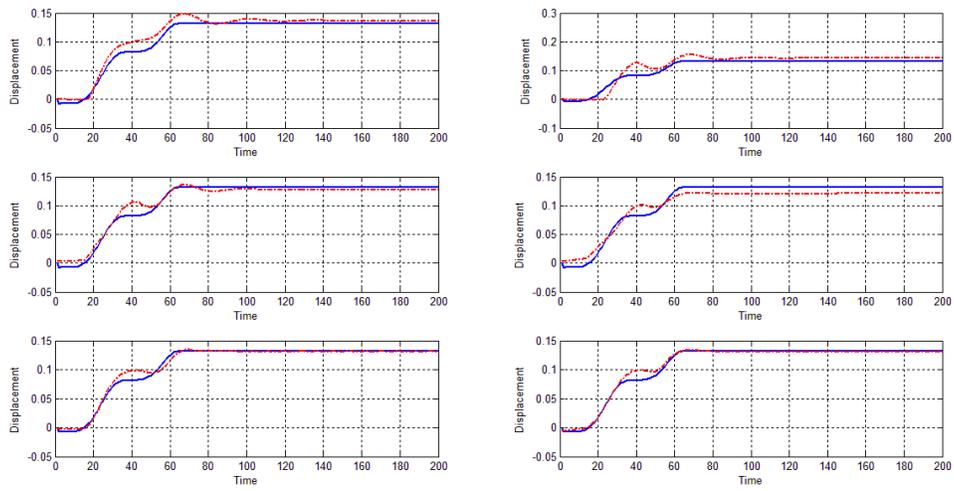


Fig.3.4.3-1(2) Increasing of accuracy for data set 9, against increasing of combination models

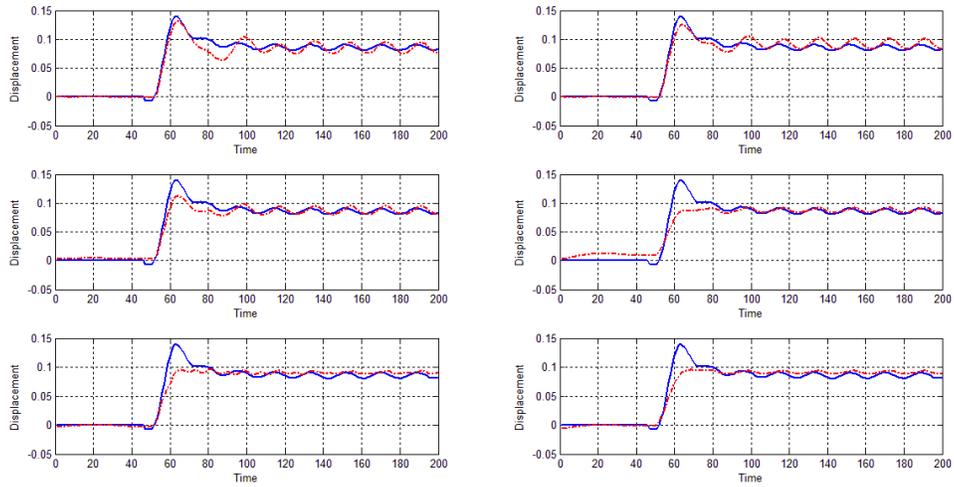


Fig.3.4.3-2(1) Decreasing of accuracy for data set 22, against increasing of combination models

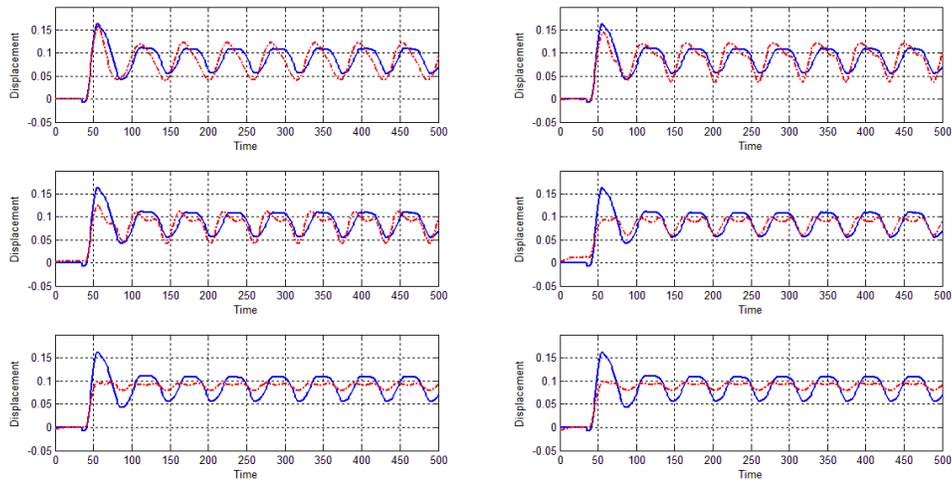


Fig.3.4.3-2(2) Decreasing of accuracy for data set 32, against increasing of combination models

From those data sets, the compromise between the high accuracy and low accuracy data sets comes to a general point, that the high accuracy data sets sacrifice its accuracy in order to cover more data sets, which in the result plot Fig.3.2.2-1(5) and Fig.3.4.1-2(4).

However, this compromise could only work with the data sets that have higher similarity. This means, the compromise could not cover the data sets which have lower similarity. The base of this hypothesis, are based on the phenomenon that different models cover different data sets (T.3.4.3-1, Fig.3.4.3-3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38		
6	0	0	0	0	0	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-9	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6-9-1	1	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
6-9-1-8	1	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0

T.3.4.3-1(2) Data sets covered by model 1 (1: below RMSE threshold, 0: over RMSE threshold)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	
20	1	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
20-16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
20-16-9	1	1	0	0	0	1	0	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
20-16-9-5	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	0	0	1	0	
20-16-9-5-35	1	1	0	0	1	0	0	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	0	1	0	
20-16-9-5-35-7	1	1	0	0	1	1	0	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	0	1	0	

T.3.4.3-1(3) Data sets covered by model 2 (1: below RMSE threshold, 0: over RMSE threshold)

From the table T.3.4.3-1(1), it could be easily seen that, the step data sets: 2, 3, 4, 5, 9, 10, and the sinusoidal data sets: 13, 14, 15, 17, 18, 19, 20, 21 and 23 are not easy to be covered with the model identified from the other single data sets. And the model identified based on single step data sets: 2, 3, 4, 5, 6, 9, 10, 11, 14, sinusoidal data sets: 7, 13, 14, 18, 23 could cover even less data sets. From the both ways, we could find out the data sets: 2, 3, 4, 5, 9, 10 of step, 13, 14, 18, 23 of sinusoidal data sets are quite independent data sets from the others. And this trend is even proved in the combination results table T3.4.3-1(2) and (3).

Based on the observation above, the hypothesis Point 3 could be supported and proved that, based on the data sets, it is not possible to find out a unique model which could cover all the data sets. Instead, several model with high accuracy to cover parts of the data sets could be found:

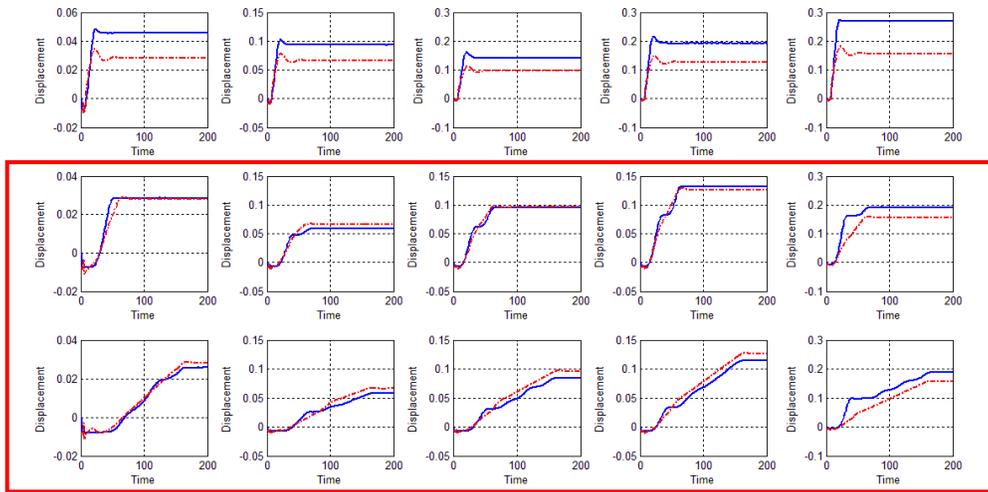


Fig.3.4.3-3(1) Combination Model 1 (6-9)

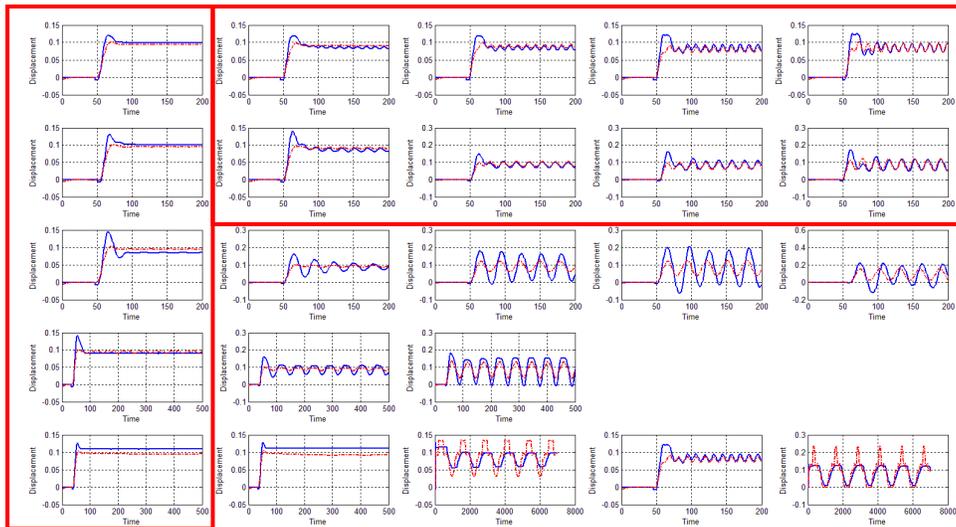


Fig.3.4.3-3(2) Combination Model 2

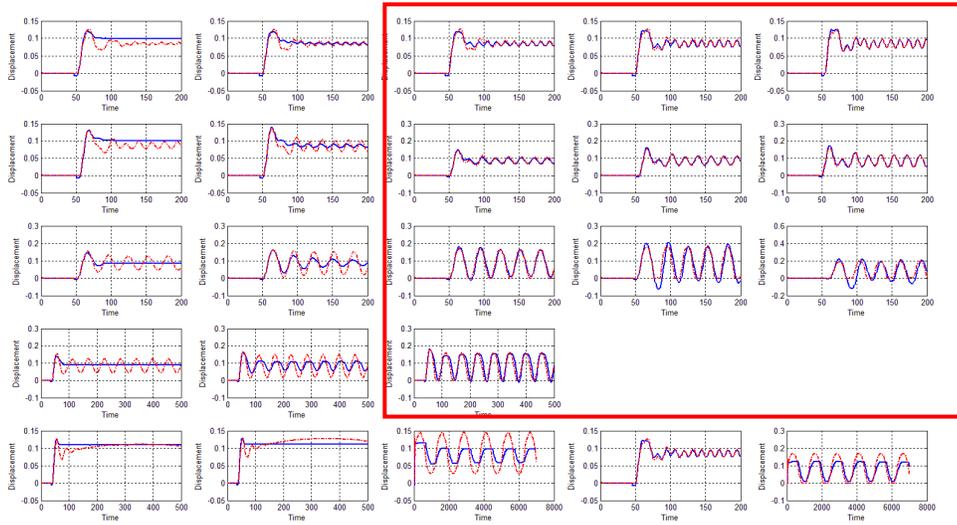


Fig.3.4.3-3(3) Combination Model 2 (20)

Further to have a well-covered data sets, the following table T3.4.3-2 of data sets to models are applied, and the figure could be check in Fig.3.4.3-4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
20-16-9-5-35-7		x			x										
6-9-1-8	x						x	x	x		x	x	x	x	x
6-9-1										x					
6-9				x		x									
20			x												

T3.4.3-2(1) Model to Step data sets

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
20-16-9-5-35-7	x					x					x					x			x				
20-16-9-5		x					x																
6-9-1																					x		x
20			x	x	x			x	x	x		x	x	x	x		x	x		x		x	

T3.4.3-2(2) Model to Sinusoidal data sets

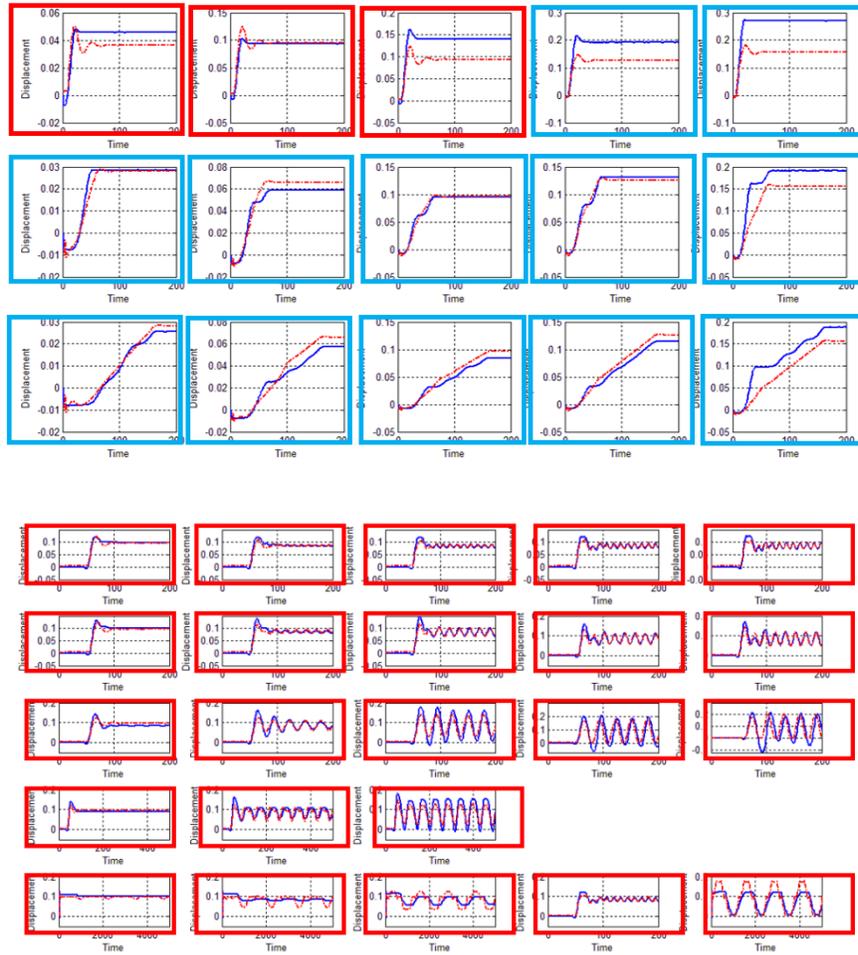


Fig.3.4.3-4(1) Sinusoidal and Step data sets with model (Marked)

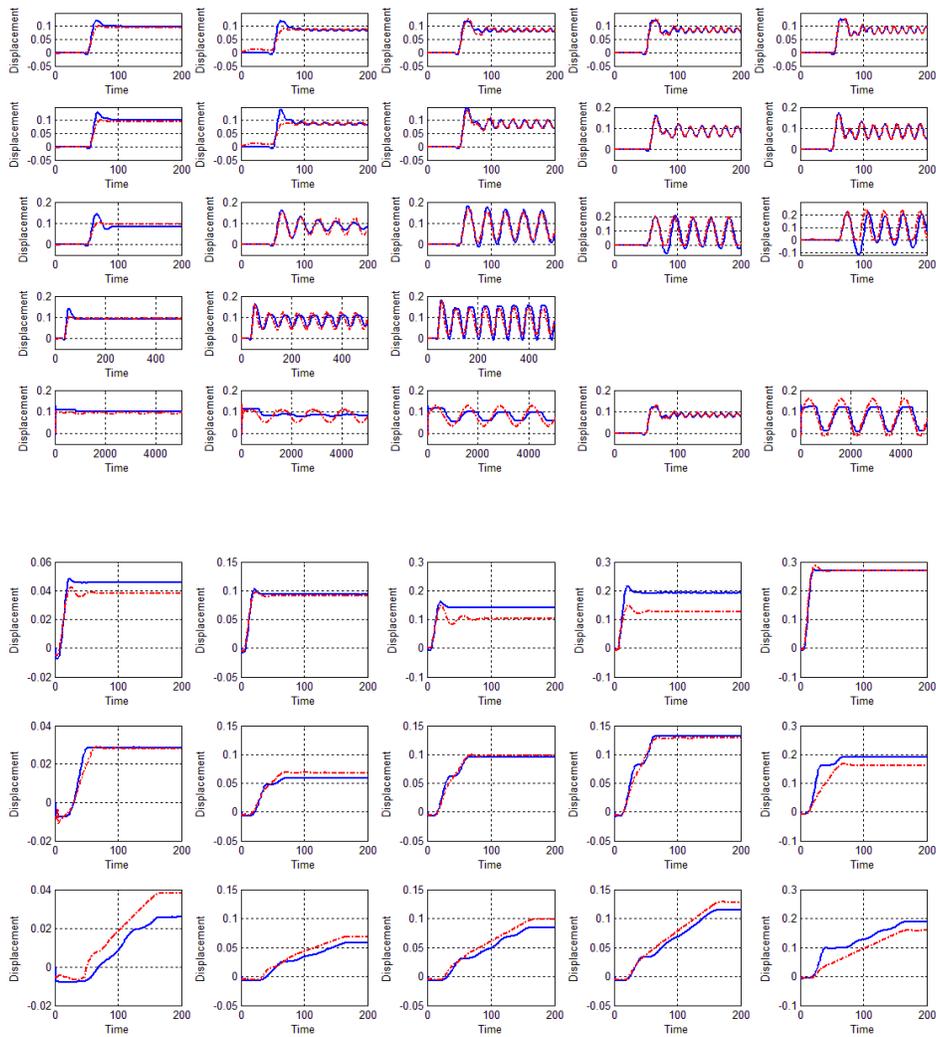


Fig.3.4.3-4(1) Sinusoidal and Step data sets with model

Finally, the reason of lower similarity could due to the following reason:

1. The welding gun system has some unknown parameters change for those data sets, which lead to a change in the system behavior.
2. The current is not the only determination factor for the displacement. There should be some missing additional inputs.

4 Conclusion

In this thesis, a method of the method of Non-linear Hammerstein-Weiner Model is applied to an servo-motor driven arc-robot, in order to identify a model for the further simulation works. By considering the input current and displacement as output, with limited number of measurements, the identification process used a measurement-combination method to improve the accuracy of the identified model, and use the RMSD and coverage number of test measurements to select the model. Finally, several accurate models with enough accuracy are identified to cover the different data sets, and the identified process could be considered as an applicable process for the limited measurement identification situation.

It could be concluded with the following points:

- The Non-linear Hammerstein-Weiner method could be used for identification of the servo-motor system
- The identification process with a well-covered data sets is necessary for improving the performance of the identified model
- With the number of data sets increasing, the accuracy of the identified model is increasing.
- With the limited number of data sets for identification, but careful designed process, could lead to an accurate enough model for the fast industry using.
- Several models are identified instead of a single model, due to the possible missing of the parameters or the system unknown changes.

5 Reference

- [1] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 546-550, 7 1966.
- [2] E.-W. Bai, "Identification of linear systems with hard input nonlinearities of known structure," *Automatica*, vol. 38, no. 5, pp. 853-860, 1 5 2002.
- [3] Y. Han, Block-oriented nonlinear system identification using semidenite programming, University of California, San Diego, 2012, p. 110.
- [4] P. Crama and J. Schoukens, "Initial estimates of Wiener and Hammerstein systems using multisine excitation," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 6, pp. 1791-1795, 2001.
- [5] D. Aryani, L. Wang and T. Patikirikorala, "Block-oriented nonlinear model based control for relative performance and resource management in virtualized software system," in *2016 Australian Control Conference (AuCC)*, 2016.
- [6] E.-W. B. Fouad Giri, Block-oriented Nonlinear System Identification, Springer, 2010.
- [7] Er-Wei Bai, "An optimal two stage identification algorithm for Hammerstein-Wiener nonlinear systems," in *Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No.98CH36207)*, 1998.
- [8] MATHWORKS, "Identifying Hammerstein-Wiener Models," MATHWORKS, [Online]. Available: <https://ww2.mathworks.cn/help/ident/ug/identifying-hammerstein-wiener-models.html#bsgkiwq>. [Accessed Feb 2019].

6 Appendix

6.1 MATLAB Code for Simulated servo-motor system

In the following part, the MATLAB code is attached. The result is packed and save as sperate 5 folders of 1A to 5A, and the identified model is named as [nb,nk,1] where nb and nk are of range 1 to 10, e.g. SYS_identi-1-3-1.mat. And file "Displacement_Identi.mat" represent the simulate result of the identified system.

6.1.1 Code for system identification

```
1 %% Data Set Generation
2 Input_Current_Simu = Input_current.data;
3 Output_Displacement_Simu = output_displacement.data;
4 Identi_DataSet =
  iddata(Output_Displacement_Simu,Input_Current_Simu);
5
6 %% System Identification with NLHW
7 for nb = 1:10      %combination of parameter nb and nk,
  identification of the NLHW model
8     for nk = 1:10
9         %% System identification
10        SYS_identi = nlhw(Identi_DataSet,[nb,nk,1]); % NLHW
  identification and model generation
11 %% Saving of the model
12    str1 = strcat('SYS_identi-',num2str(nb),'-',num2str(nk),'-
  1',' .mat');
13    savefile = str1;
14    save(savefile,'SYS_identi');
15    end
16 end
```

6.1.2 Simulate identified system

6.1.2.1 Code for simulate identified system

```
1 %% Load current input
2 Input_Current_Simu = Input_current.data;
3 Output_Displacement_Simu = output_displacement.data;
4 %% Load the identification model and simulate with current input
5 for nb = 1:10      %combination of parameter nb and nk of the NLHW
  model
6     for nk = 1:10
7         n = (nb-1)*10+nk; % number of identified model
8         str1 = strcat('SYS_identi-',num2str(nb),'-',num2str(nk),'-
  1',' .mat');
9         loadfile = str1;
10        load(loadfile); % load the identified model with parameter of
  [nb, nk, 1]
```

```

11             % the identified model is named as "SYS_identi"
12     %% Simulate the identified system with the current input
13     Output_Displacement_Identi =
sim(SYS_identi,Input_Current_Simu);    %simulate the displacement of
the identified model [nb,nk,1]
14     Displacement_Identi(n,1:length(Output_Displacement_Identi)) =
Output_Displacement_Identi';    %save the result to a
matrix
15     end
16 end
17 %% Saving of the result
18     str1 = strcat('Displacement_Identi.mat');
19     savefile = str1;
20     save(savefile, 'Displacement_Identi');

```

6.1.3 Code for RMSE calculation

```

1 %% Load simulated displacement
2 Output_Displacement_Simu = output_displacement.data;
3 %% Load identified displacement and calculation of RMSE result
4 for nb = 1:10    %combination of parameter nb and nk of the NLHW
model
5     for nk = 1:10
6         n = (nb-1)*10+nk; % number of identified model
7         Output_Displacement_Identi = Displacement_Identi(n,:);
8         RMSE = norm (Output_Displacement_Identi(:)-
Output_Displacement_Simu(:)) / sqrt (length
(Output_Displacement_Identi)); % RMSE Calculation
9         RMSE_Result(n,1) = RMSE';
10    end
11 end
12 %% Saving of the result
13     str1 = strcat('RMSE_Result.mat');
14     savefile = str1;
15     save(savefile, 'RMSE_Result');

```

Code for real measurement system

6.1.3.1 Code for identification of single measurement for step 1 to 3

```

1 for n = 1:1:15
2 %% Create the data set for identification {x, i}n
3 DspK_Output_Identi = x_Step_Aft_0_All(1:200,n);
4 i_Input_Identi = i_Step_Aft_0_All(1:200,n);
5 Id_nlhwh_Input=iddata(DspK_Output_Identi,i_Input_Identi);
6 %% System Identification based on data set {x, i}n, with parameter
nb and nk, [1,10]
7 for nb=1:10
8     for nk=1:10
9 %% Identification of the NLHW Model
10         NLHW_Step_model=nlhw(Id_nlhwh_Input, [nb,nk,1]);

```

```

11 %% Simulate displacement of the NLHW Model with current input
12
13     x_Step_identi=sim(NLHW_Step_model,i_Step_Aft_0_All(1:200,n));
14 %% RMSE calculation
15     nn = (nb-1)*10+nk
16     RMSE_Step_nlhw_Cross(nn,n)=norm (x_Step_Aft_0_All(1:200,n)-
17     x_Step_identi) / sqrt (length (x_Step_Aft_0_All(1:200,n)));
18 end
19 end
20 end

```

6.1.3.2 Code for identification of single measurement for step 1 to 5

```

1  for n = 1:1:15
2
3  DspK_Output_Identi = x_Step_Aft_0_All(1:200,n);
4  i_Input_Identi = i_Step_Aft_0_All(1:200,n);
5  Id_nlhw_Input=iddata(DspK_Output_Identi,i_Input_Identi);
6
7  for nb=1:10
8      for nk=1:10
9          NLHW_Step1=nlhw(Id_nlhw_Input,[nb,nk,1]);
10         x_Step0=sim(NLHW_Step1,Id_nlhw_Input);
11         x_Step_Tx=x_Step0.y;
12         x_Step1=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,1));
13         x_Step2=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,2));
14         x_Step3=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,3));
15         x_Step4=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,4));
16         x_Step5=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,5));
17         x_Step6=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,6));
18         x_Step7=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,7));
19         x_Step8=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,8));
20         x_Step9=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,9));
21         x_Step10=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,10));
22         x_Step11=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,11));
23         x_Step12=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,12));
24         x_Step13=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,13));
25         x_Step14=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,14));
26         x_Step15=sim(NLHW_Step1,i_Step_Aft_0_All(1:200,15));
27
28         nn = (n-1)*100+(nb-1)*10+nk
29
30
31         RMSE_Step_nlhw(nn,1)=norm (x_Step_Aft_0_All(1:200,1)-
32         x_Step1) / sqrt (length (x_Step_Aft_0_All(1:200,1)));
33         RMSE_Step_nlhw(nn,2)=norm (x_Step_Aft_0_All(1:200,2)-
34         x_Step2) / sqrt (length (x_Step_Aft_0_All(1:200,2)));
35         RMSE_Step_nlhw(nn,3)=norm (x_Step_Aft_0_All(1:200,3)-
36         x_Step3) / sqrt (length (x_Step_Aft_0_All(1:200,3)));
37         RMSE_Step_nlhw(nn,4)=norm (x_Step_Aft_0_All(1:200,4)-

```

```

    x_Step4) / sqrt (length (x_Step_Aft_0_All(1:200,4)));
35     RMSE_Step_nlhwh(nn,5)=norm (x_Step_Aft_0_All(1:200,5)-
    x_Step5) / sqrt (length (x_Step_Aft_0_All(1:200,5)));
36     RMSE_Step_nlhwh(nn,6)=norm (x_Step_Aft_0_All(1:200,6)-
    x_Step6) / sqrt (length (x_Step_Aft_0_All(1:200,6)));
37     RMSE_Step_nlhwh(nn,7)=norm (x_Step_Aft_0_All(1:200,7)-
    x_Step7) / sqrt (length (x_Step_Aft_0_All(1:200,7)));
38     RMSE_Step_nlhwh(nn,8)=norm (x_Step_Aft_0_All(1:200,8)-
    x_Step8) / sqrt (length (x_Step_Aft_0_All(1:200,8)));
39     RMSE_Step_nlhwh(nn,9)=norm (x_Step_Aft_0_All(1:200,9)-
    x_Step9) / sqrt (length (x_Step_Aft_0_All(1:200,9)));
40     RMSE_Step_nlhwh(nn,10)=norm (x_Step_Aft_0_All(1:200,10)-
    x_Step10) / sqrt (length (x_Step_Aft_0_All(1:200,10)));
41     RMSE_Step_nlhwh(nn,11)=norm (x_Step_Aft_0_All(1:200,11)-
    x_Step11) / sqrt (length (x_Step_Aft_0_All(1:200,11)));
42     RMSE_Step_nlhwh(nn,12)=norm (x_Step_Aft_0_All(1:200,12)-
    x_Step12) / sqrt (length (x_Step_Aft_0_All(1:200,12)));
43     RMSE_Step_nlhwh(nn,13)=norm (x_Step_Aft_0_All(1:200,13)-
    x_Step13) / sqrt (length (x_Step_Aft_0_All(1:200,13)));
44     RMSE_Step_nlhwh(nn,14)=norm (x_Step_Aft_0_All(1:200,14)-
    x_Step14) / sqrt (length (x_Step_Aft_0_All(1:200,14)));
45     RMSE_Step_nlhwh(nn,15)=norm (x_Step_Aft_0_All(1:200,15)-
    x_Step15) / sqrt (length (x_Step_Aft_0_All(1:200,15)));
46
47
48     RMSE_Step_nlhwh(nn,16) = mean(RMSE_Step_nlhwh(nn,1:15));
49     RMSE_Step_nlhwh(nn,17) = min(RMSE_Step_nlhwh(nn,1:15));
50     RMSE_Step_nlhwh(nn,18) = max(RMSE_Step_nlhwh(nn,1:15));
51
52 end
53 end
54 end

```