Supplementary Information

for

Systematic Design-of-Experiments, Factorial-Design Approaches for Tuning Simple Empirical Water Models

Diarmuid O'Connor and Niall J. English *

School of Chemical and Bioprocess Engineering, University College Dublin, Belfield, Dublin 4, Ireland.

RDF-Error Analysis for O-H and H-H cases

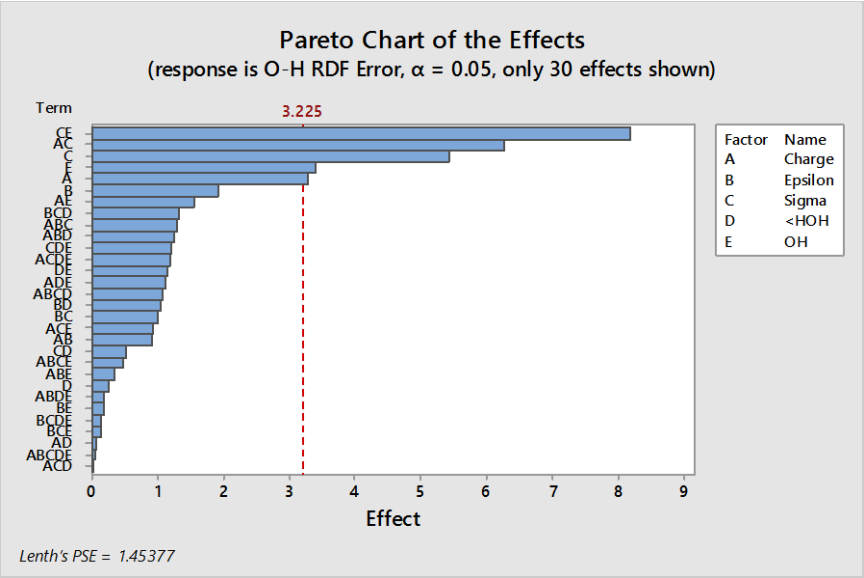


Figure S1: Pareto chart for effects on O-H radial distribution function

From the Pareto chart (Fig. S1), it can be seen that there are five significant effects:

- CE= Sigma.|OH| bond length
- AC= Charge.Sigma
- C = Sigma
- E = |OH| bond length
- A = Charge

* Corresponding author: niall.english@ucd.ie, Tel: +353-1-7161646, Fax: +353-1-7161177

It can be clearly seen that CE= Sigma.OH bond length represents the largest effect as it extends out the furthest.

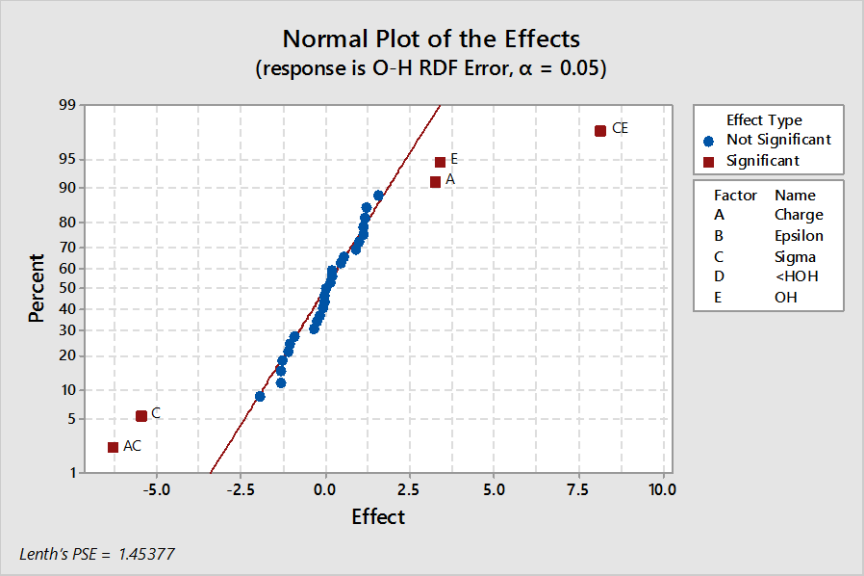


Figure S2: Normal-probability plot for O-H RDF

The normal-probability plot agrees with the Pareto plot when considering the effects which most affect the diffusivity of water. In addition to this the normal plot indicates the direction of the effect. Charge (A), OH bond length (E) and Sigma.OH bond length (CE) all have positive standardised effects. This indicates when the level is changed from a low level to a high level of the factor, the response will increase. Sigma (C) and Charge.Sigma (AC) both have a negative effect. This indicates when the level is changed from a low level to a high level of the factor, the response will decrease.

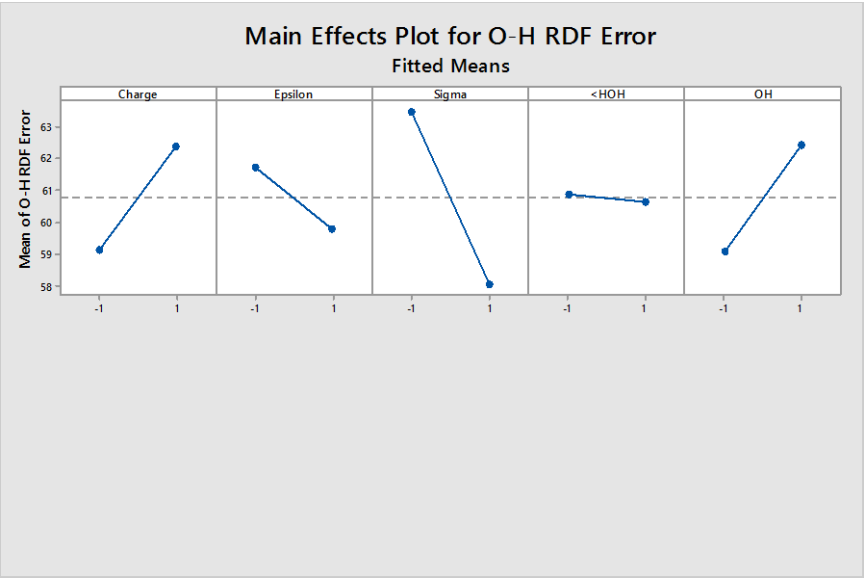


Figure S3: Main-effect plot for O-H RDF

Fig. S3 above shows the variable which have the largest effect on the response are charge, sigma and OH bond length. It also shows the epsilon and ∠HOH bond angle have a insignificant effect on the response. It can also be noted that:

- Charge: O-H RDF error is higher with charge at a higher level. This can also be seen in the normal plot of all main and interaction effects.
- Epsilon: The O-H RDF error will be higher at higher values for epsilon.
- Sigma: O-H RDF error will be lower at higher levels of charge. It can also be noted from the main effects chart that sigma appears to have the largest effect on the output response.
- ∠HOH bond angle: While the O-H RDF error will be lower at higher values for the bond angle a change will have little effect to the output response.
- OH bond length: O-H RDF error will be higher at higher levels of the OH bond length.

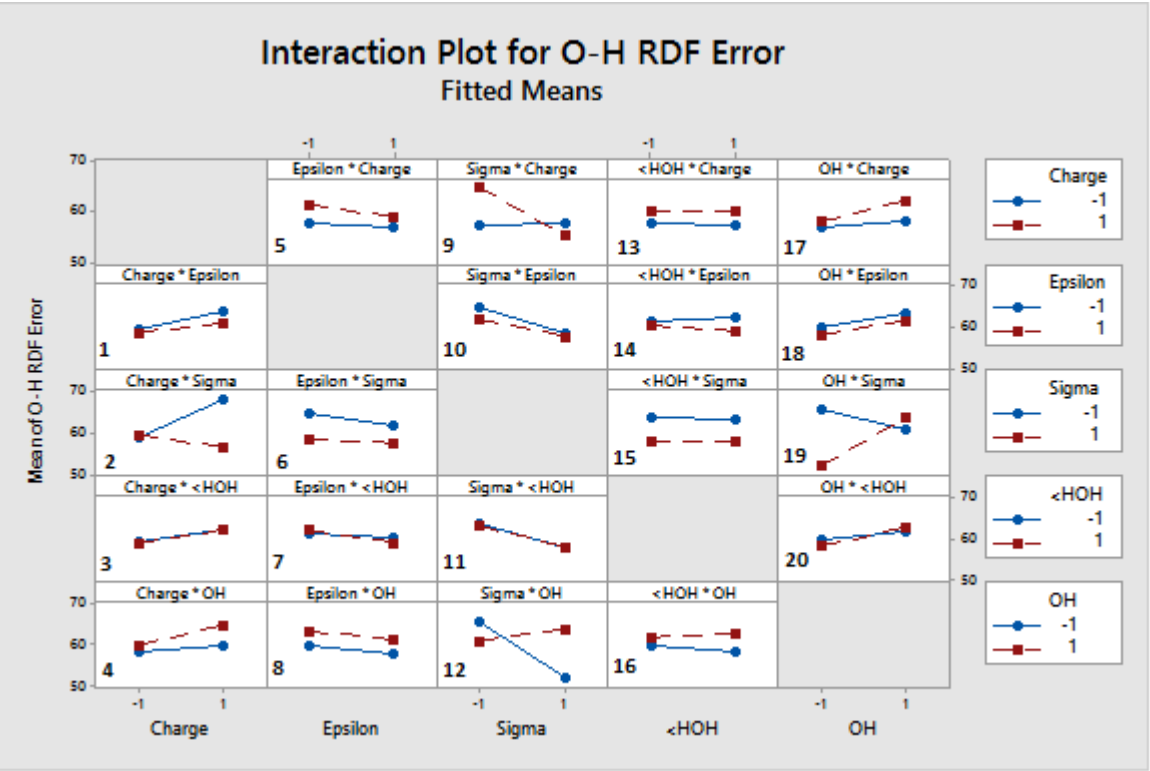


Figure S4: Interaction plots for two-way interactions affecting O-H RDF

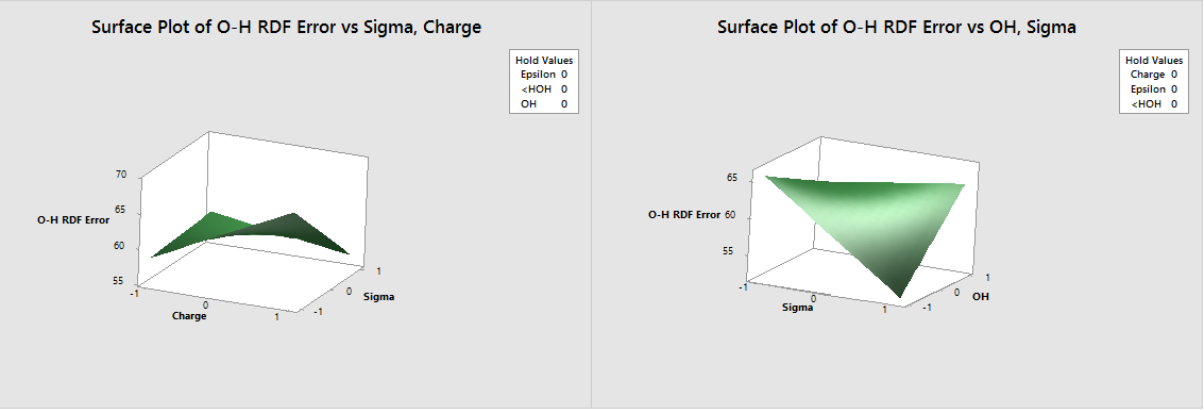
The table above displays all the two-way interactions within the system. As it can be seen many of the interactions are not significant. Some of the key interactions include:

- Section 2: The plot indicates that the increase and decrease in O-H radial distribution function error when charge changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is equal to when sigma is high. The response increases when sigma is low and decreases when sigma is high when charge increases.
- Section 9: The plot indicates that the decrease in O-H radial distribution function error when sigma changes from -1 to 1 depends on charge, only if charge is high. When charge is low, the change in response is less than when charge is high. The response only changes when charge is high.
- Section 12: The plot indicates that the increase and decrease in O-H radial distribution function error when sigma changes from -1 to 1 depends on the OH bond angle. When the OH bond angle is low, the change in response is greater than when the OH bond

angle is high. The response increases slightly when the OH bond angle is low and decreases when the OH bond angle is high when sigma increases.

- Section 19: The plot indicates that the increase and decrease in O-H radial distribution function error when the OH bond angle changes from -1 to 1 depends on the sigma. When the sigma is low, the change in response is less than when the sigma is high. The response increases slightly when the sigma is low and decreases when the sigma is high when the OH bond angle increases.

Sigma is heavily involved in most significant interaction effects which affect O-H radial distribution function error.



(1) (2)

Figure S5: Surface plots for the most significant effects on O-H RDF

The surface plots shown in Figure S5 above can be analysed as follows,

1. The interaction between charge and sigma is shown to output the highest response at a high level of charge and a low level of sigma. It is also shown to output the lowest response at a high level of charge and a high level of sigma.
2. The interaction between sigma and the OH bond length is shown to output the highest response at a low level of sigma and a low level of the OH bond length. It is also shown to output the lowest response at a high level of sigma and a low level of the OH bond length.

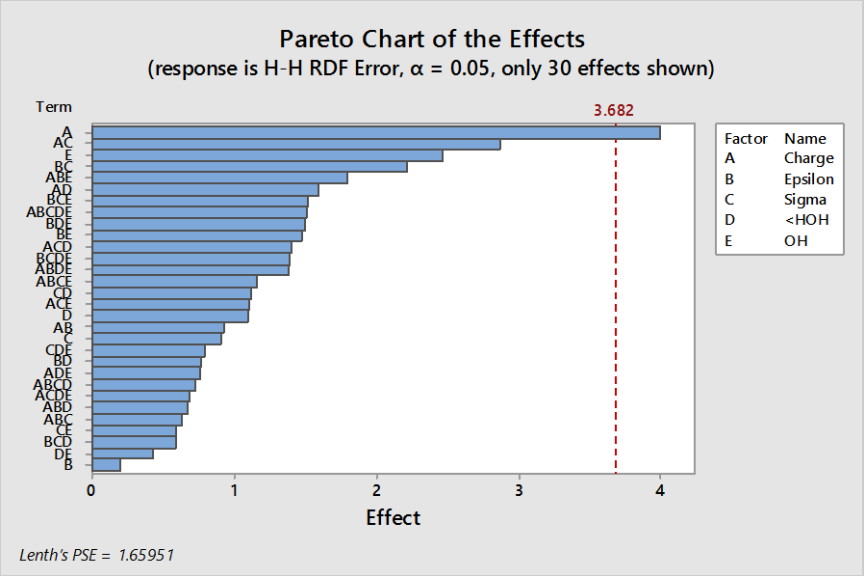


Figure S6: Pareto chart for effects on H-H radial distribution function

From the Pareto chart it can be seen that there is one significant effect, i.e., A = Charge

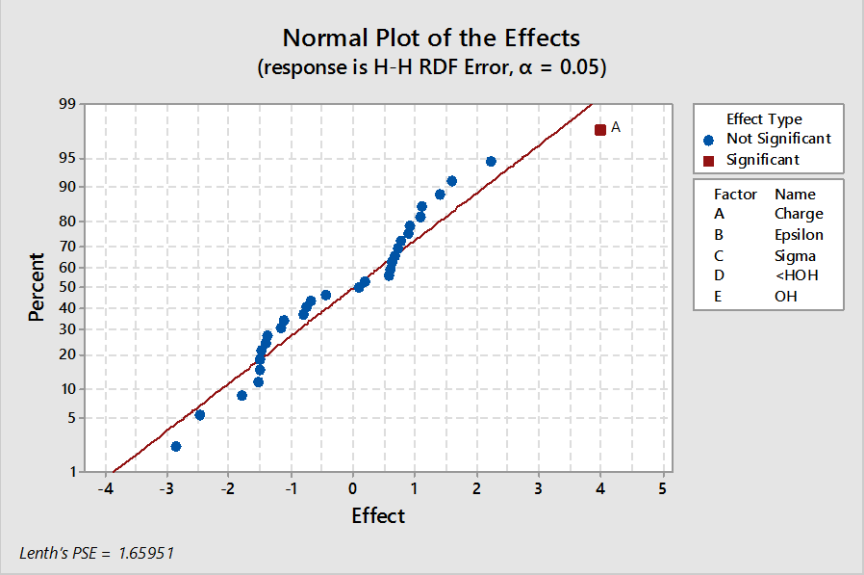


Figure S7: Normal-probability plot for H-H RDF

The normal plot agrees with the Pareto plot when considering the effects which most affect the diffusivity of water. Charge (A) has a positive standardised effect. This indicates when

the level is changed from a low level to a high level of the factor, the response will increase. There are no effects that have a negative effect.

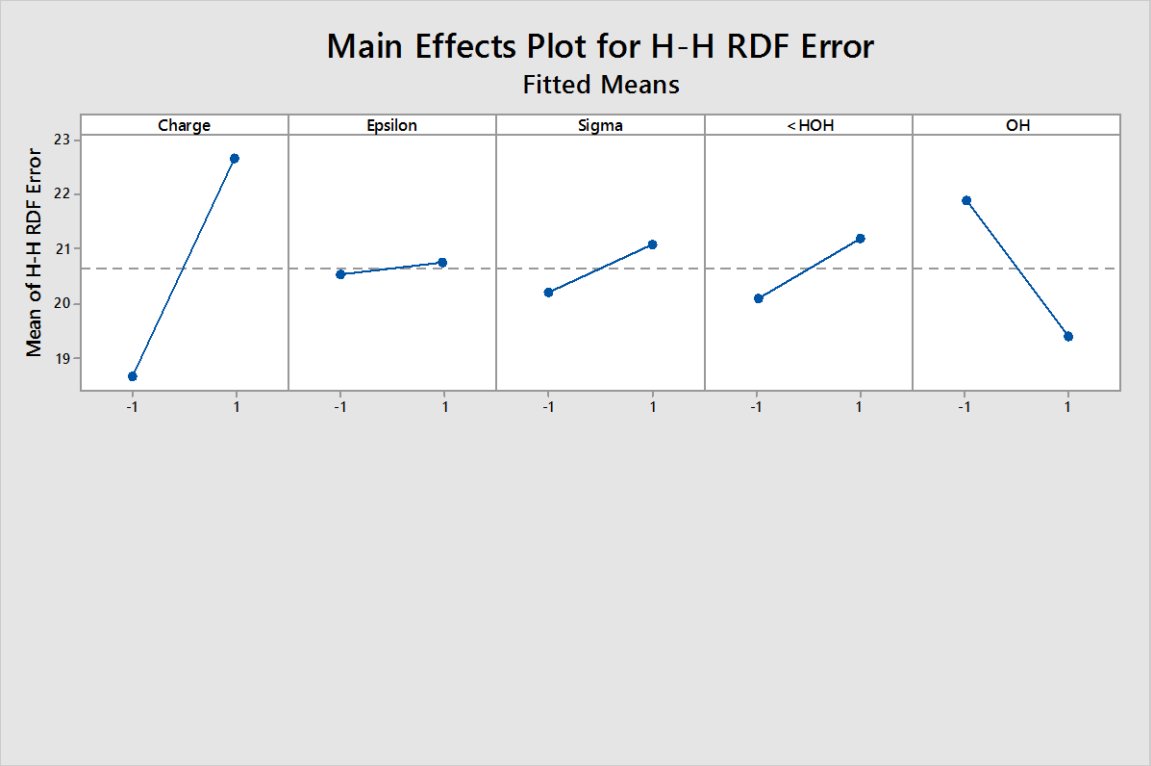


Figure S8: Main-effect plot for H-H radial Distribution function

Fig. S8 shows the variables which have the largest effect on the response are charge and OH bond length. It also shows the epsilon have a insignificant effect on the response. It can also be noted that:

- Charge: H-H RDF error is higher with charge at a higher level. This can also be seen in the normal plot of all main and interaction effects. It can also be noted from the main effects chart that charge appears to have the largest effect on the output response.
- Epsilon: The H-H RDF error will be higher at higher values for epsilon. It can also be noted that a change in epsilon will have little effect to the output response.
- Sigma: H-H RDF error will be higher at higher levels of charge.

- ∠HOH bond angle: The H-H RDF error will be higher at higher values for the bond angle,
- OH bond length: H-H RDF error will be lower at higher levels of the OH bond length.

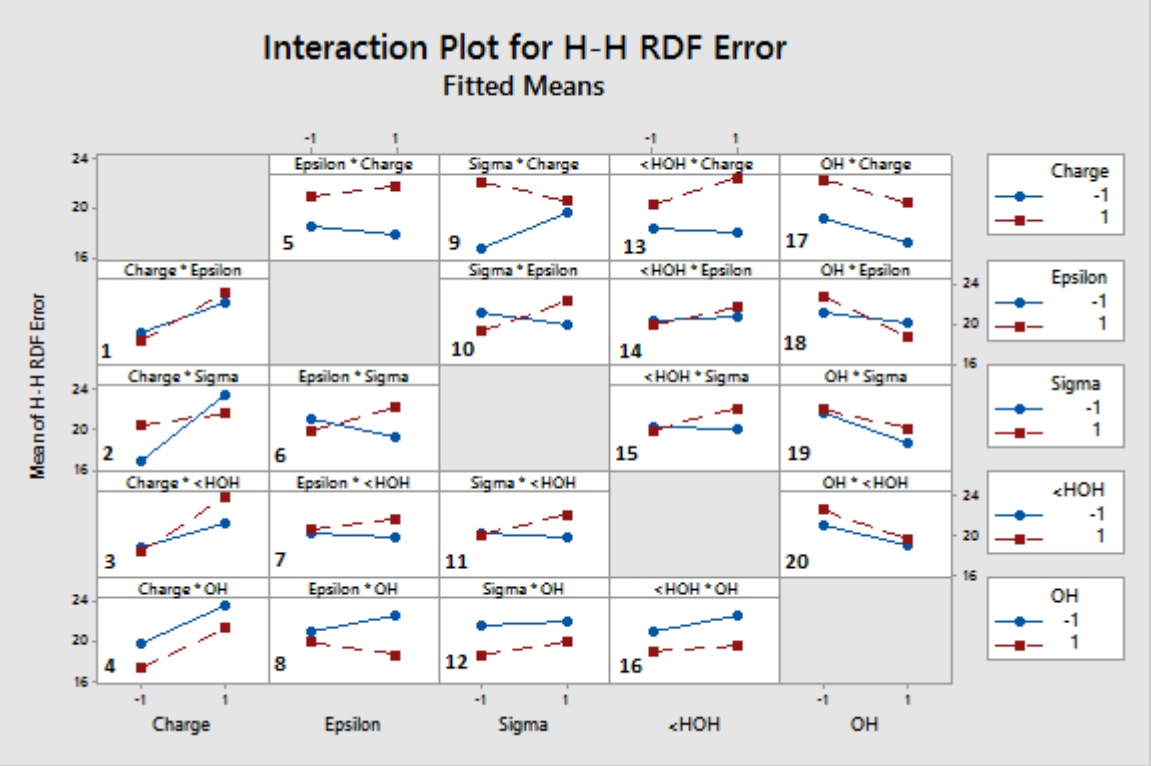


Figure S9: Two-way interaction plots on H-H radial distribution function

Fig. S9 displays all the two-way interactions within the system. As the Pareto plot shows, the majority of interactions are not significant; only the two most significant interactions will be discussed, which include:

- Section 2: The plot indicates that the increase in H-H RDF error when charge changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is greater than when sigma is high.

- Section 6: The plot indicates that the increase and decrease in H-H radial distribution function error when epsilon changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is equal to when sigma is high.

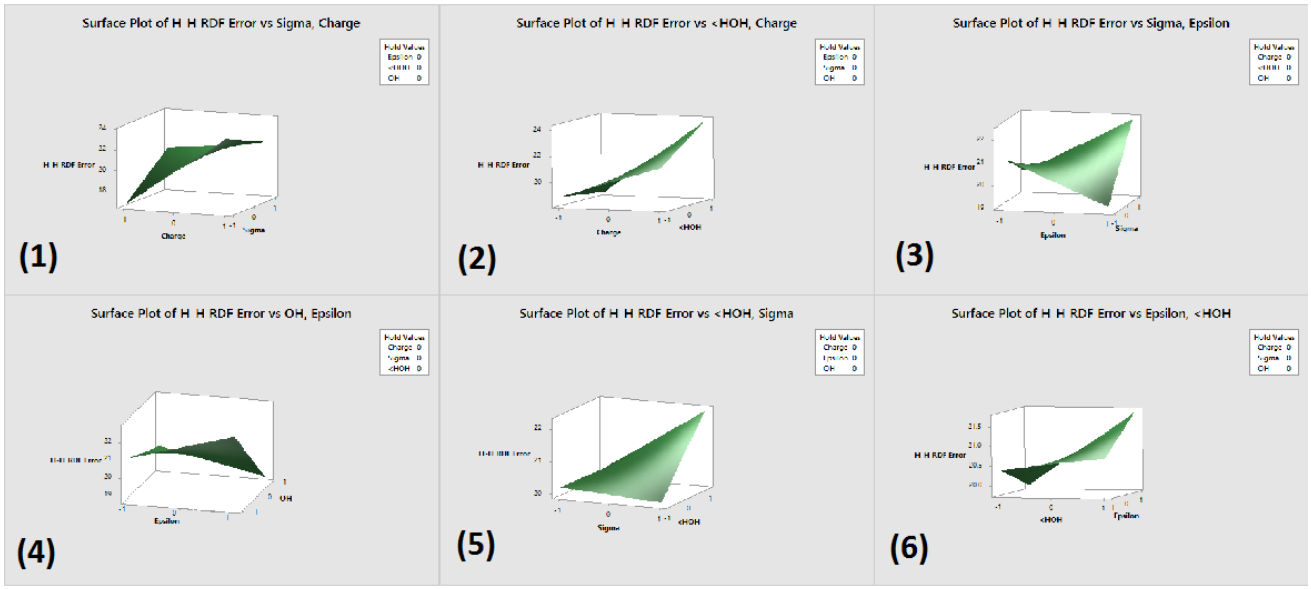


Figure S10: Surface plots for the most significant effects which affect H-H radial distribution function

The surface plots shown in Figure S10 above can be analysed as follows,

- The interaction between charge and sigma is shown to output the highest response at a high level of charge and a low level of sigma. It is also shown to output the lowest response at a low level of sigma and a low level of charge.
- The interaction between charge and HOH bond angle is shown to output the highest response at a high level of charge and a high level of HOH bond angle. It is also shown to output the lowest response at a low level of charge and a high level of HOH bond angle.

- The interaction between epsilon and sigma is shown to output the highest response at a high level of epsilon and a high level of sigma. It is also shown to output the lowest response at a high level of epsilon and a low level of sigma.
- The interaction between epsilon and OH bond length is shown to output the highest response at a high level of epsilon and a low level of OH bond length. It is also shown to output the lowest response at a high level of epsilon and a high level of OH bond length.
- The interaction between sigma and HOH bond angle is shown to output the highest response at a high level of sigma and a high level of HOH bond angle. It is also shown to output the lowest response at a high level of sigma and a low level of HOH bond angle.
- The interaction between HOH bond angle and epsilon is shown to output the highest response at a high level of HOH bond angle and a high level of epsilon. It is also shown to output the lowest response at a low level of HOH bond angle and a high level of epsilon.

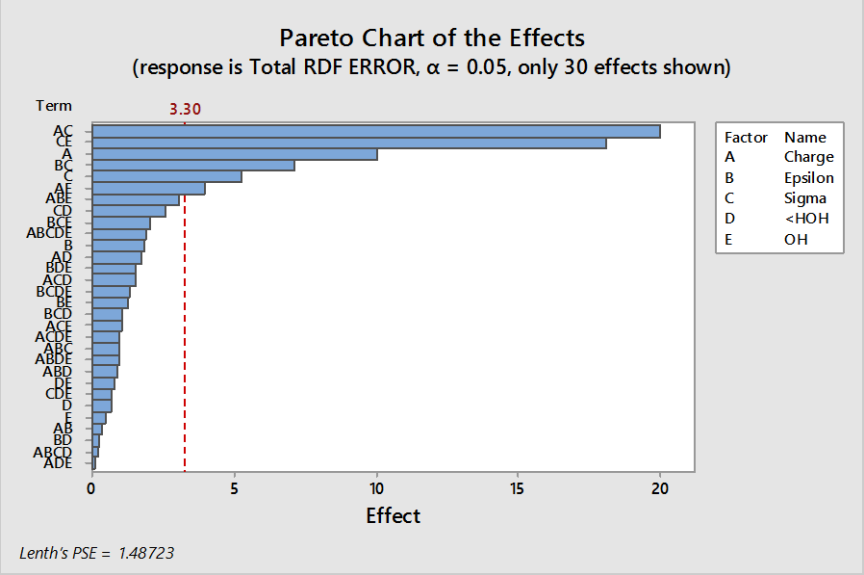


Figure S11: Pareto chart for effects on total radial distribution function

From the Pareto chart it can be seen that there are 6 significant effects. These significant effects include:

- AC= Charge.Sigma
- CE= Sigma.|OH| bond length
- A = Charge
- BC = Epsilon.Sigma
- C = Sigma
- AE= Charge.|OH| bond length

It can be clearly seen that AC (Charge.Sigma) and CE (Sigma.OH bond length) represents the largest effects as it extends out the furthest. It can also be seen that ADE has the smallest effect.

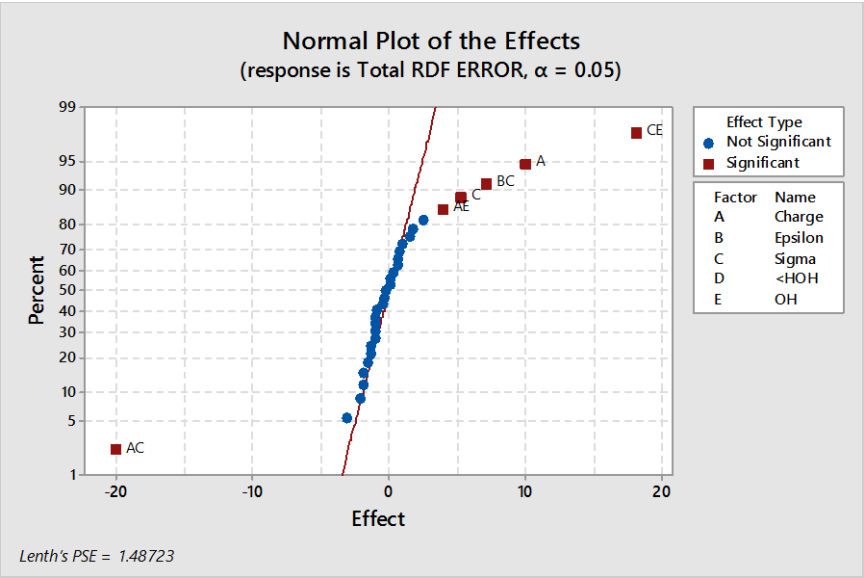


Figure S12: Normal-probability plot for Total radial Distribution function

The normal-probability plot agrees with the Pareto plot when considering the effects which most affect the total-RDF error. In addition to this the normal plot indicates the direction of the effect. Sigma.OH bond length (CE), Charge (A), Epsilon.Sigm (BC), Sigma (C) and Charge.OH bond lengh (AE) all have positive standardized effects. This indicates when the level is changed from a low level to a high level of the factor, the response will increase. Charge.Sigma (AC) has a negative effect. This indicates when the level is changed from a low level to a high level of the factor, the response will decrease.

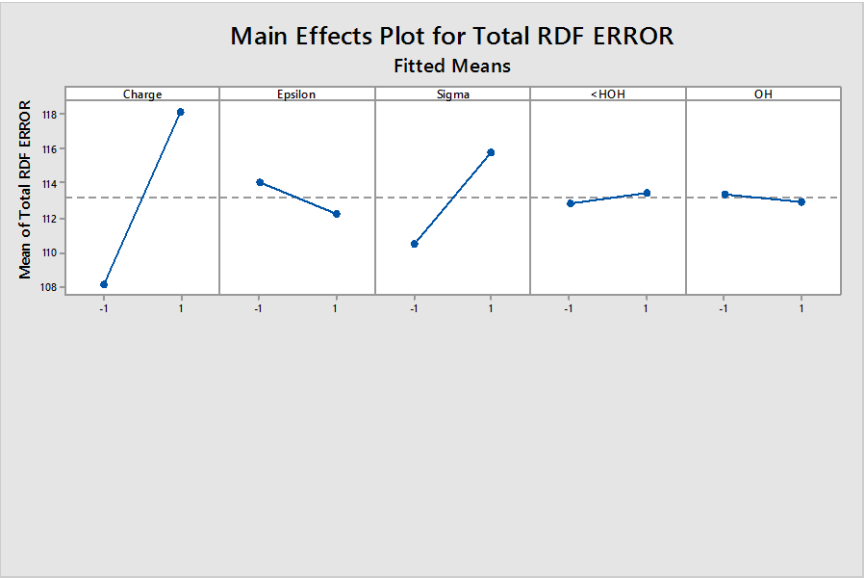


Figure S13: Main-effect plot for total-RDF error

For the plot above shows the variable which have the largest effect on the response are charge and OH bond length. It also shows the epsilon have a insignificant effect on the response. It can also be noted that:

- Charge: The total RDF error is higher with charge at a higher level. This can also be seen in the normal plot of all main and interaction effects. It can also be noted from the main effects chart that charge appears to have the largest effect on the output response.
- Epsilon: The total RDF error will be higher at lower values for epsilon.
- Sigma: The total RDF error will be higher at higher levels of charge.
- ∠HOH bond angle: The total RDF error will be higher at higher values for the bond angle. It can also be noted that a total RDF error in epsilon will have little effect to the output response.
- OH bond length: total RDF error will be lower at higher levels of the OH bond length. It can also be noted that a total RDF error in epsilon will have little effect to the output response.

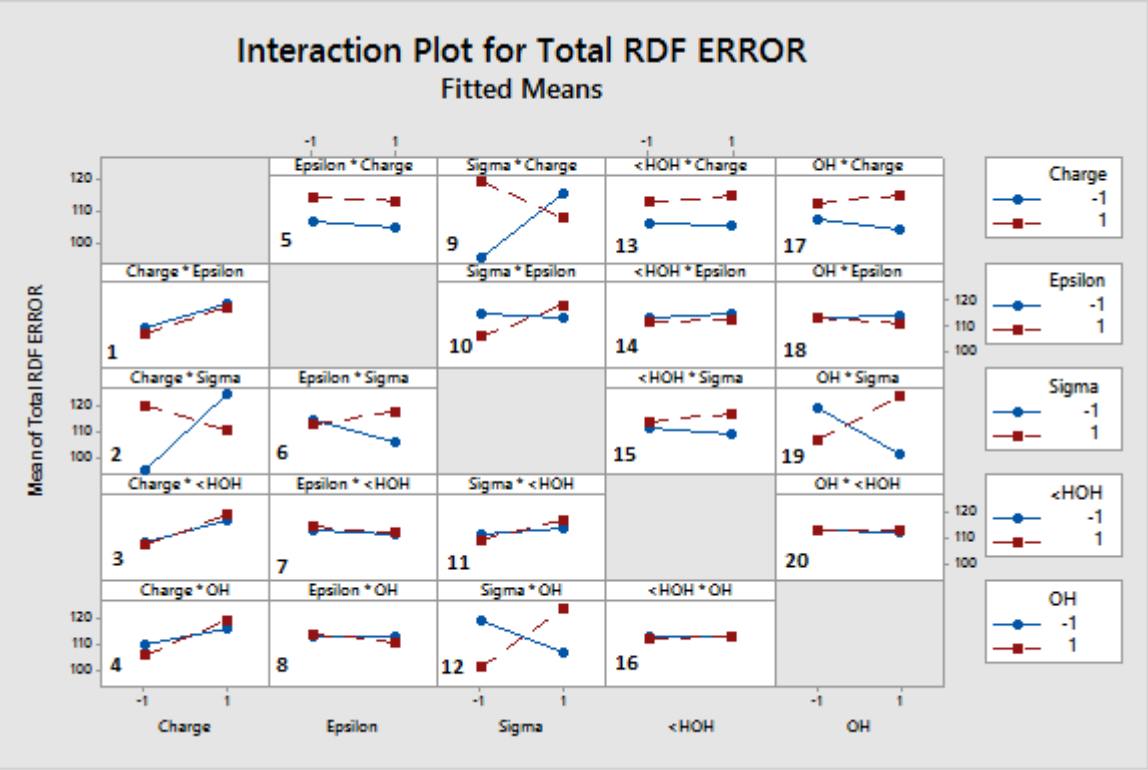


Figure S14: Interaction plots for two-way interaction which affect total radial distribution function error

The table above displays all the two-way interactions within the system. As it can be seen many of the interactions are not significant. Some of the key interactions include:

- Section 2: The plot indicates that the increase and decrease in total radial distribution function error when charge changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is less than when sigma is high. When charge changes from low to high, sigma at a low value increases the response while the response increases when sigma is high.
- Section 6: The plot indicates that the increase and decrease in total radial distribution function error when epsilon changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is equal to when sigma is high.

- Section 9: The plot indicates that the increase and decrease in total radial distribution function error when sigma changes from -1 to 1 depends on the charge. When charge is low, the change in response is less than when charge is high. When sigma changes from low to high, charge at a low value increases the response while the response increases when charge is high.
- Section 12: The plot indicates that the increase and decrease in total radial distribution function error when sigma changes from -1 to 1 depends on the OH bond length. When the OH bond length is low, the change in response is greater than when the OH bond length is high. When sigma changes from low to high, the OH bond length at a low value increases the response while the response increases when the OH bond length is high.
- Section 19 The plot indicates that the increase and decrease in total radial distribution function error when the OH bond length changes from -1 to 1 depends on the sigma. When sigma is low, the change in response is equal than when sigma is high.

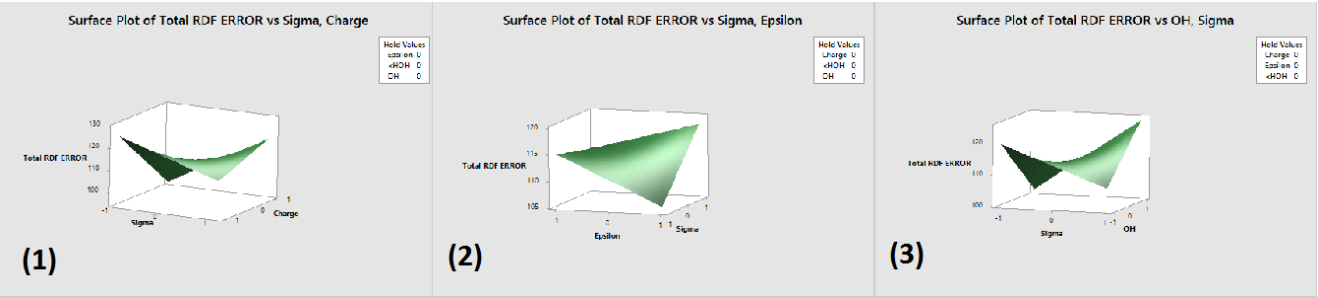


Figure S15: Surface plots for the most significant effects which affect total-RDF error

The surface plots shown in Figure S15 above can be analysed as follows,

- The interaction between charge and sigma is shown to output the highest response at a low level of charge and a high level of sigma. It is also shown to output the lowest response at a low level of charge and a low level of sigma.

- The interaction between epsilon and sigma is shown to output the highest response at a high level of epsilon and a high level of sigma. It is also shown to output the lowest response at a high level of epsilon and a high level of sigma.
- The interaction between sigma and the OH bond length is shown to output the highest response at a high level of sigma and a high level of the OH bond length. It is also shown to output the lowest response at a low level of sigma and a high level of the OH bond length.

Term	Effect	Term	Effect	Term	Effect	Term	Effect
O-O		O-H		H-H		Total RDF	
Sigma	9.826	Sigma*OH	8.175	Charge	4	Sigma*OH	18.1
Sigma*OH	9.339	OH	3.386	Epsilon*Sigma	2.22	Charge	10
Epsilon*Sigma	3.923	Charge	3.272	Charge*<HOH	1.6	Epsilon*Sigma	7.15
Charge	2.753	Charge*OH	1.561	Charge*Sigma*<HOH	1.41	Sigma	5.3
Charge*OH	2.277	Sigma*<HOH*OH	1.205	Sigma*<HOH	1.11	Charge*OH	3.94
Sigma*<HOH	0.924	Charge*Sigma*<HOH*OH	1.181	<HOH	1.1	Sigma*<HOH	2.57
Charge*Epsilon*Sigma*OH	0.79	<HOH*OH	1.14	Charge*Epsilon	0.93	Charge*<HOH	1.74
Charge*Sigma*<HOH*OH	0.452	Charge*<HOH*OH	1.118	Sigma	0.9	Charge*Sigma*<HOH	1.51
Charge*Epsilon	0.364	Epsilon*Sigma	1.008	Epsilon*<HOH	0.77	Charge*Sigma*<HOH*OH	0.95
Sigma*<HOH*OH	0.291	Charge*Sigma*OH	0.93	Charge*Epsilon*Sigma*<HOH	0.73	<HOH*OH	0.79
Charge*Epsilon*<HOH*OH	0.258	Sigma*<HOH	0.535	Charge*Epsilon*<HOH	0.67	Sigma*<HOH*OH	0.71
Charge*<HOH	0.204	Charge*Epsilon*Sigma*OH	0.486	Charge*Epsilon*Sigma	0.64	<HOH	0.69
Charge*Epsilon*Sigma*<HOH	0.141	Charge*Epsilon*<HOH*OH	0.187	Sigma*OH	0.59	Charge*Epsilon	0.38
Charge*Sigma*<HOH	0.123	Epsilon*OH	0.184	Epsilon*Sigma*<HOH	0.59	Charge*<HOH*OH	0.13
<HOH*OH	0.076	Epsilon*Sigma*<HOH*OH	0.146	Epsilon	0.2	Charge*Epsilon*Sigma*OH	0.12
Epsilon*OH	0.014	Epsilon*<HOH*OH	0.01	Charge*OH	0.1	Charge*Epsilon*Sigma*<HOH	-0.2
Epsilon*<HOH	-0.01	Charge*Sigma*<HOH	-0.02	<HOH*OH	-0.43	Epsilon*<HOH	-0.3
Epsilon*Sigma*<HOH*OH	-0.05	Charge*Epsilon*Sigma*<HOH*OH	-0.03	Charge*Sigma*<HOH*OH	-0.68	OH	-0.5
Epsilon*<HOH*OH	-0.05	Charge*<HOH	-0.06	Charge*<HOH*OH	-0.75	Charge*Epsilon*<HOH	-0.9
Epsilon	-0.14	Epsilon*Sigma*OH	-0.15	Sigma*<HOH*OH	-0.79	Charge*Epsilon*<HOH*OH	-0.9
<HOH	-0.15	<HOH	-0.26	Charge*Sigma*OH	-1.11	Charge*Epsilon*Sigma	-0.9
Charge*<HOH*OH	-0.23	Charge*Epsilon*OH	-0.34	Charge*Epsilon*Sigma*OH	-1.16	Charge*Sigma*OH	-1
Charge*Epsilon*Sigma	-0.29	Charge*Epsilon	-0.91	Charge*Epsilon*<HOH*OH	-1.38	Epsilon*Sigma*<HOH	-1
Epsilon*Sigma*<HOH	-0.3	Epsilon*<HOH	-1.04	Epsilon*Sigma*<HOH*OH	-1.4	Epsilon*OH	-1.3
Charge*Epsilon*<HOH	-0.32	Charge*Epsilon*Sigma*<HOH	-1.08	Epsilon*OH	-1.48	Epsilon*Sigma*<HOH*OH	-1.3
Charge*Epsilon*Sigma*<HOH*OH	-0.35	Charge*Epsilon*<HOH(=	-1.26	Epsilon*<HOH*OH	-1.49	Epsilon*<HOH*OH	-1.5
Epsilon*Sigma*OH	-0.41	Charge*Epsilon*Sigma	-1.29	Charge*Epsilon*Sigma*<HOH*OH	-1.51	Epsilon	-1.9
Charge*Sigma*OH	-0.86	Epsilon*Sigma*<HOH	-1.32	Epsilon*Sigma*OH	-1.52	Charge*Epsilon*Sigma*<HOH*OH	-1.9
Charge*Epsilon*OH	-0.92	Epsilon	-1.92	Charge*Epsilon*OH	-1.8	Epsilon*Sigma*OH	-2.1
OH	-1.38	Sigma	-5.43	OH	-2.47	Charge*Epsilon*OH	-3.1
Charge*Sigma	-10.9	Charge*Sigma	-6.27	Charge*Sigma	-2.87	Charge*Sigma	-20

Table S1: Summary of RDF-error effects in order of size